

KLIINISEN BIOSTATISTIIKAN PERUSKURSSI

SISÄLLYSLUETTELO:

Johdanto	5
Tilastollisten menetelmien käytöstä	5
Oppikirjoista	5
Ohjelmistoista	8
Lääketieteellisten artikkelien laatu tilastollisesti tarkasteltuna	10
Tutkimusprosessi	11
Tilastollinen päättely	12
Viitteet	13
Ohjelmia	15
Tutkimustyytit ja -asetelmat	17
Yleistä	17
Kokeelliset tutkimukset	18
Kokonaistutkimukset	20
Otantatutkimukset	20
Poikittaistutkimukset	20
Tapaus-verrokkitutkimukset	21
Kohorttitutkimukset, seurantatutkimukset	22
Meta-analyysi	24
Yleistä	24
Meta-analyysi vastaan suuri lääkekoe	25
Viitteet	26
Aineistokoko ja voima - analyysi ...	28
Johdanto	28
Aineistokoon määrittämiseen liittyvät avainkysymykset	31
Yhden binomijakaumaa noudattavan suhteellisen osuuden testaaminen	32
Kahden binomijakaumaa noudattavan suhteellisen osuuden vertailu	34
Kaksi toisistaan riippumatonta ryhmää	34
Yksi otos	37
Arviointi ristitulosuhteen avulla	38
Parittaiset aineistot	39
Kahden suhteellisen osuuden ekvivalenttisuuden testaaminen	41
Kvantitatiivinen lopputulos, kahden keskiarvon vertailu	42
Yksi otos, yksi tai kaksi ajankohtaa tai periodia	42

Kahden otoksen t-testi	44
Parametrittomat testit.....	46
Muita aineistokoon laskentatilanteita.....	47
Kommentteja	48
Viitteet	49
Taulukoita ja ohjelmia.....	49

Satunnaistaminen ja sokkouttaminen 51

Yleistä.....	51
Yksinkertainen satunnaistaminen	53
"Korttipakan sekoittamis" -menettely.....	54
Satunnaiset permutoidut lohkot	55
Harhaisen kolikon menetelmä.....	56
Ositettu satunnaistaminen.....	56
Minimointimenettely.....	57
Epäsymmetrisesti tasapainotetut menetelmät	59
Puutteet satunnaistamisessa	59
Ohjelmia	60
Viitteet	60

Aineiston kerääminen, tallentaminen ja käsittely 61

Otantamenetelmiä	61
Yksinkertainen satunnaisotanta.....	61
Systemaattinen otanta.....	61
Ositettu otanta	62
Monivaiheinen otanta	63
Aineiston kuvaaminen, tallentaminen ja käsittely.....	63
Muuttujaluettelo	65
Asteikot.....	67
Puuttuvat tiedot	68
Aineiston tallentaminen ja muokkaaminen	69
Poikkeavat havaintoarvot.....	70
Keskiluvut.....	71
Hajonnan mitat.....	74
Keskihajonta vs. keskivirhe	77
Jakauman vinous ja huipukkuus	78
Jakauman normaalisuus	78
Graafiset kuvaajat.....	79
Viitteet	82

Luottamusväliestimointi 83

Yleistä.....	83
Esimerkkejä luottamusvälien laskentatilanteista.....	87
Normaalijakaumaan perustuvat luottamusvälit.....	87
Keskiarvon luottamusväli	88
Keskiarvojen erotuksen luottamusväli	90
Mediaanien ja muiden prosenttipisteiden luottamusvälit	92
Mediaanieron luottamusväli	93
Suhteellisen osuuden luottamusväli.....	94
Suhteellisten osuuksien erotuksen luottamusväli.....	98
Riippumattomat otokset.....	98
Parittaiset otokset.....	100
Lukumäärän luottamusväli Poisson- jakauman tapauksessa	102
Lukumäärien suhteen luottamusväli Poisson-jakauman tapauksessa	104
Multinomijakaumaan perustuva luottamusväli.....	108
EkspONENTTijakaumaan perustuva luottamusväli.....	109
Altistus- ja riskisuhteen luottamusvälit	110
Viitteet	114

Tilastollisten testien valinta ja..... 116

toteuttaminen..... 116

Hypoteesien testaaminen.....	116
Monivertailut.....	117
Parametrinen ja parametriton testi	119
P-arvon käyttöön liittyviä rajoitteita ja ongelmia.....	119
Tilastollisten testien valintaperusteet	122
Normaalisuuden testaaminen	125
Varianssien homogeenisuustestit	129
Permutaatiotesti riippuville otoksille.....	130
Parittainen t-testi	131
Wilcoxonin parittainen testi	134
McNemarin testi.....	138
McNemar-Bowkerin testi	141
LR-testi siirtymätodennäköisyystaulukoille.....	144
Riippumattomien ryhmien t-testi	146
Mann–Whitneyn U-testi	150
Wilcoxonin järjestyslukujen summatesti.....	152
χ^2 -nelikenttätesti	156
Fisherin tarkka nelikenttätesti	160
Fisherin tarkka testi (r x c)-taulukoille.....	162
Binomiaalisten suhteellisten osuuksien trenditesti.....	163
Mantel-Haenszelin nelikenttätesti	166
Kruskal–Wallisin yksisuuntainen varianssianalyysi	170
Dunnin testi Kruskal–Wallisin testin yhteydessä.....	173
Jonckheere–Terpstra trenditesti.....	174
Friedmanin testi ja Pagen testi.....	175

Dunnin testi Friedmanin testin yhteydessä	178
Kaksisuuntainen parametriton varianssianalyysi	179
Pearsonin korrelaatiokerroin	182
Spearmanin järjestyskorrelaatiokerroin	186
Kendallin Tau ja Somerin D	187
Cohenin kapp ja painotettu kapp	187
Kendallin konkordanssikerroin (W)	192
Viitteet	193

Biostatistisia ohjeita artikkelien kirjoittamisessa 195

Johdanto	195
Aineisto- ja menetelmäosa	195
Yleistä	195
Tilastolliset menetelmät	196
Tulososa	197
Aineiston kuvailu	197
Tulosten esittäminen	197
Virheitä tulosten analysoinnissa	198
Kliiniset kokeet	199
Laboratoriotutkimukset	200
Meta-analyysi, systemaattinen katsaus	201
Havainnoivat tutkimukset	201
Viitteet	201

Johdanto

Tilastollisten menetelmien käytöstä

Tilastollisten menetelmien käyttö lääketieteellisissä julkaisuissa on erittäin yleistä. Jo parikymmentä vuotta sitten (Emerson & Colditz 1983) arvioitiin, että noin kolmessa neljästä julkaistusta tutkimuksesta oli käytetty merkittävässä määrin tilastollisia menetelmiä. Nykyisin tämä osuus lienee tuntuvasti suurempi. Tästä johtuen useimmissa arvostetuissa tieteellisissä lehdissä kiinnitetäänkin erityistä huomiota tilastollisten menetelmien riittävälle ja oikealle käytölle. Monet tieteelliset lehdet, esimerkiksi British Medical Journal ja Lancet, ovat luoneet erillisen tilastollisen arviointijärjestelmänsä tavanomaisen arviointijärjestelmän täydennykseksi. Näissä lehdissä on julkaistu myös hyviä ohjeistoja tilastollisten menetelmien asiallisesta käytöstä ja siitä mihin asioihin lääketieteellisiin lehtiin kirjoittavien tutkijoiden tulisi kiinnittää erityistä huomiota (Altman ym.1983, Gore ym.1992).

Kliinisissä kokeissa edellytetään, että tilastollista asiantuntemusta käytetään jo tutkimuksen suunnitteluvaiheesta alkaen. Myös muuntotyypisissä kliinisissä tutkimuksissa, joissa tilastollisten menetelmien käyttö on relevanttia, tulisi tilastollista asiantuntemusta hyödyntää mahdollisimman varhaisesta vaiheesta alkaen. Hyvin suunnitellun ja toteutetun, mutta huonosti tai puutteellisesti analysoidun tutkimuksen voi yleensä saada julkaisukuntoon uudelleenanalysoinnilla. Huonosti suunnitellun, hallinnoidun ja toteutetun tutkimuksen tuloksia ei hyvälläkään tilastollisella analyysillä saa tieteellisesti luotettaviksi.

Vain osassa lääketieteellisistä julkaisuista tilastotieteen asiantuntija on tutkimuksen tekijöiden joukossa. Tästä syystä lääkäritutkijoiden tulisikin itse perehtyä tilastotieteen peruskäsitteisiin ja -menetelmiin riittävässä määrin, jotta heidän tutkimuksiensa laatu olisi myös tilastollisesti pätevää. Valtaosa klinikoista ei itse suorita tutkimuksia, vaan he ovat muiden suorittamien tutkimusten hyödyntäjiä, tutkimustulosten siirtäjiä kliiniseen käyttöön. Myös tämän ryhmän kohdalla tilastollisten menetelmien tuntemisesta on hyötyä; on tärkeää kyetä erottelamaan jyvät akanoista ja ottamaan kliiniseen käyttöön vain tieteellisesti luotettavia tuloksia.

Oppikirjoista

Viime vuosina on ilmestynyt ilahduttavan paljon hyviä biostatistiikan kirjoja, jotka tarjoavat lääketieteellisiin lehtiin kirjoittaville ja kirjoituksia hyödyntäville hyvän perustietopaketin tutkimusasetelmista ja tavallisimmin käytetyistä tilastollisista menetelmistä sekä niiden oikeasta käytöstä. Monissa kirjoissa sisällön pääpaino on selkeästi asetettu menetelmävalintoihin, käyttöedellytyksiin ja analyysien tuottamien tulosten tulkintaan, eikä niinkään analyysien tekniseen suorittamiseen. Tilastollisten ratkaisumenetelmien tarkka sisältö matemaattisine kaavoineen ja erilaiset arviointitekniikat ovat asioita, joista ohjelmia hyödyntävän ei tarvitse kovin paljon tietää.

Mitä biostatistiikan peruskirjoja lääketieteen tutkijan ja klinikon kannattaa kirjahyllyynsä hankkia? Valintaan vaikuttaa aikaisempi perehtyneisyys biostatistiikkaan, kuinka paljon biostatistisia menetelmiä työssään tarvitsee ja ehkä jossain määrin kirjan hinta. Joitakin peruskirjoja tutkijan tai tutkijayhteisön kirjahyllyssä olisi kuitenkin hyvä olla. Seuraavassa esitellään joukko sellaisia alkeis- tai peruskirjoja, joista erityisesti kliinisten alojen tutkijat hyötyvät eniten.

Sofistikoiduimmatkaan tilastolliset menetelmät eivät voi pelastaa tutkijaa virheellisiltä tuloksilta ja vääriiltä johtopäätöksiltä ellei tutkimus ole oikein suunniteltu ja tutkimuksessa käytetyt tiedot

mitattu ja kerätty oikein ja luotettavasti. Lääketieteellisissä tutkimuksissa yleisimmin käytettyihin biostatistisiin menetelmiin keskittyviä kirjoja on runsaasti tarjolla, mutta tutkimusten suunnitteluun soveltuva kirjatarjonta on huomattavasti vähäisempää. Hyvä ja hyödyllinen kirja tähän tarkoitukseen on **Machinin** ja **Campbellin** ”Design of Studies for Medical Research”. Kirjassa annetaan hyviä vinkkejä ja ohjeita sekä kliinisten että epidemiologisten tutkimusten suunnitteluvaiheeseen. Hinta € 36.

<http://books.google.fi/books?id=UFGN93h482QC>

Langin ja **Secicin** kirja ”How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors and Reviewers” Kirjaan on koottu runsaasti asiantuntijavihjeitä ja suosituksia lääketieteellisiin lehtiin kirjoittaville. Kirja on tarkoitettu ensisijaisesti lääkäreille, mutta siitä on hyötyä myös heidän kanssaan yhteistyötä tekeville tai arvioitsijoina toimiville biostatistikoille. Toinen kirjan tekijöistä on lääkäri ja toinen biostatistikko. Kirja on tyypiltään manuaalimainen eikä pyri korvaamaan biostatistiikan perusoppikirjoja. Hinta \$64,95

<http://books.google.fi/books?id=kBUBRh1AWG4C>

<https://www.acponline.org/atpro/timssnet/catalog/books/statistics.htm>

Mikäli haluaa saada nopean käsityksen biostatistiikan perusasioista, niin kannattaa lukea kirja nimeltä ”Statistics at Square One” 10. painos. Tekijät **Swinscow** ja **Campbell** ovat saaneet mahdutettua 158 sivuun hämmästyttävän paljon asiaa biostatistiikan ja epidemiologian peruskäsitteistä. Jokaisen luvun lopussa on lisäksi kappale, joka on otsikoitu ”common questions”, missä tekijät tarkastelevat opiskelijoiden ja tutkijoiden tavallisimmin esittämiä kysymyksiä käsitellyistä asioista. Kunkin luvun lopussa on lisäksi kappale ”reading and reporting”, missä lukijalle annetaan tulkinta- menettelytapaoheja lääketieteellisten artikkelien lukemisessa tai raportoinnissa. . Hinta £15,19.

<http://books.google.fi/books?id=IRLgGwAACAAJ&dq=Swinscow+Campbell+Statistics+at+square+one>

Wassertheil–Smollerin ”Biostatistics and Epidemiology” on terveydenhuoltohenkilöstölle suunnattu alkeisoppikirja, jonka tavoitteena on ollut pyrkiä tekstiin, joka oli mahdollisimman vähän pelottava ja luotaantyöntävä henkilölle, joka ei ole matemaattisesti orientoitunut. Kirja kattaa olennaisimman osan biostatistiikan ja epidemiologian peruskäsitteistä ja uusimpaan painokseen on sisällytetty myös geneettisen ja molekyyliepidemiologian peruskäsitteitä. Kirjassa on käsitelty myös monikeskustutkimusten yleisperiaatteita, seulontatutkimuksia ja tutkimusetiikkaa. Kirja on helppolukuinen ja soveltuu esim. lääketieteen perusopetukseen. Kovin paljon lähtötietoja ei edellytetä. Hinta €29,95. <http://books.google.fi/books?id=Ez6ZJLmrNYgC>

Campbellin, Machinin ja Waltersin ”Medical Statistics: A Textbook for Health Sciences, 4th edition” on helppotajuinen peruskirja, joka on tarkoitettu lääketieteen ja sairaanhoitoalan opiskelijoille, mutta se on hyödyllinen myös potilaita hoitaville lääkäreille muistuttamaan tilastollisten tutkimusasetelmien ja tulosten tulkinnan olennaisimmista asioista. Kirja tarjoaa lukijalleen hyvän katsauksen tärkeimmistä analyysi- ja mittausmenetelmistä lääketieteellisessä tutkimuksessa. Monimuuttujamenetelmät on kuitenkin rajattu lähes täysin sisällön ulkopuolelle. Kirja sisältää paljon hyviä käytännön ohjeita myös lääketieteellisiin lehtiin kirjoittaville tutkijoille. Kirja on selkeästi kirjoitettu ja siinä on hyvät lähdeviitteet. Kirjan tekijöillä on laaja opetus ja konsultointikokemus. Se näkyy erityisesti esimerkkien valinnassa. Esimerkit ovat aitoja ja ne on otettu joko tekijöiden omista tutkimuksista tai lääketieteellisestä kirjallisuudesta. BMJ:n lukijoille monet esimerkeistä ovat hyvinkin tuttuja. Kirja sisältää runsaasti tehtäviä vastauksineen, joiden avulla lukijan on helppo evaluoida asioiden omaksumista. Kirja soveltuu siten hyvin sekä perusopetukseen että itseopiskeluun. Hinnaltaan se on edullinen, €24. <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470025190.html>

Altmanin kirja ”Practical Statistics for Medical Research” on erittäin hyvä biostatistiikan perusoppikirja ja vankka lähdeos. Kirja on tarkoitettu ennen kaikkea lääketieteen tutkijoille ja

kliinikoille, jotka haluavat saada hyvän yleiskäsityksen tilastollisista pääperiaatteista ja menetelmistä, mutta se on erinomainen kirja myös lääketieteen opiskelijoille ja statistikoille, jotka haluavat suuntautua kliiniseen biostatistiikkaan. Avainkäsitteet on määritelty selkeästi ja sisällön pääpaino on ennen kaikkea lääketieteellisten tutkimusten suunnittelun ja analysoinnin käytännön näkökohdissa. Kirjan sisältö on kattava sisältäen myös joidenkin keskeisimpien monimuuttujamenetelmien perusteet, tosin hyvin suppeasti. Kirjan rakennuselementteinä on käytetty Altmanin lukuisia yleistajuisia, mm. BMJ:ssä julkaistuja, biostatistisia artikkeleita vuosien varrelta. Kirja sisältää paljon sellaista hyödyllistä tietoa, mitä ei yleensä ole saatavilla yhdestä lähteestä. Siinä on paljon hyviä esimerkkejä ja harjoitustehtäviä todellisista kliinisistä tutkimuksista täydellisine tutkimusaineistoinen, joten kirjaa voidaan erinomaisesti hyödyntää sekä perus- että jatko-opetuksessa. Hinta \$93,95.

http://www.crcpress.com/shopping_cart/products/product_detail.asp?sku=C7630&isbn=9780412276309&parent_id=&pc

Biostatistiikan perusasiat jo osaavalle voi suositella Altmanin kirjan vaihtoehdoksi **Armitagen, Berryn** ja **Matthewsin** ”Statistical Methods in Medical Research”. Se on uusittu versio perinteikkäästä pelkästään Armitagen nimissä kulkeneesta monelle tutkijalle hyvinkin tutusta oppikirjasta. Olennaisimmat muutokset aiempiin painoksiin verrattuna on tapahtunut niissä kappaleissa, jotka käsittelevät eloonjäämismenetelmiä, sekventiaalisia ja epidemiologisia menetelmiä. Aihepiiriltään se kattaa suunnilleen samat asiat kuin Altmanin kirja, mutta se on esitykseltään jonkin verran syvällisempi ja edellyttää lukijalta enemmän perustietoja. Kirja on kuitenkin kirjoitettu ei-matematiikoille. Sisältö kattaa hyvin kaikki ne tilastolliset perustekniikat, joita useimmiten lääketieteellisissä tutkimuksissa käytetään. Hinta €96.

<http://books.google.fi/books?id=WZ2-ellsB0oC>

Suomenkielisiä terveydenhuoltohenkilöstölle suunnattuja biostatistiikan kirjoja on melko niukasti saatavilla. **Rannan, Koukin** ja **Ritan** kirjoittama pitkään markkinoilla ollut (8. painos) Biometria- niminen kirja on suunnattu ekologeille, mikä tarkoittaa ennen kaikkea esimerkkien aihepiirivalikoimaa. Sinänsä kirja kattaa hyvin myös ne tilastolliset perusmenetelmät, joita lääketieteilijätkin käyttävät. Kirja on selkeästi ja hyvin kirjoitettu. Jonkinasteisia vaikeuksia kliinikolle voi syntyä, kun hän yrittää muodostaa analogioita kirjan esimerkeistä omaan tutkimusongelmaansa. Hinta €49,30.

Uharin ja **Niemisen** ”Epidemiologia ja biostatistiikka” on helppolukuinen epidemiologian ja biostatistiikan peruskirja. Siinä tarkastellaan lääketieteellisen tutkimusaineiston keräystä, käsittelyä, analysointia, ja raportointia runsaan kuvitus- ja esimerkkiaineiston havainnollistamana. Tekijät ovat asettaneet tavoitteekseen täysipainoisen asiasisällön ohella mahdollisimman helposti omaksuttavan asioiden ilmaisemismuodon alentaakseen lukijakunnan kynnystä epidemiologian ja biostatistiikan opiskeluun. Tässä haasteellisessa tavoitteessaan tekijät ovat mielestäni onnistuneet melko hyvin. Kirjan hinnat: €54, Duodecim-seuran jäsenet €40,50 ja opiskelijat €32,40.

Metsämuurosen muhkea 1630-sivuinen teos ”Tutkimuksen tekemisen perusteet ihmistieteissä” on terveydenhuoltohenkilöstölle suunnattu hyvä hakuteos tutkimusentekoon liittyvistä asioista, biostatistiset menetelmät mukaan lukien. Kirja sisältää runsaasti tutkimuksen tekemiseen liittyvää hyödyllistä tietoa kattaen koko tutkimusentekoprosessin. Kirjassa opastetaan myös SPSS- tilastopakettin käyttöä sekä perusmenetelmissä että monimuuttuja-analyyseissä. Kirjan esitystapa on selkeä ja esimerkit ovat havainnollisia. Kirjaa voi tilata osoitteesta: <http://www.methelp.com>. Hinta €59.

Glantzin ”Primer of Biostatistics” on suosittu biostatistiikan alkeisoppikirja, joka on ollut jo vuosia tutkijoiden käytössä. Sen uusin kuudes painos on vuodelta 2005. Rakenteeltaan kirja on tavanomainen oppikirja, joka kattaa kliinisen tutkijan kannalta kaikki olennaiset perusasiat.

Se on selkeästi ja opiskelijaystävällisesti kirjoitettu. Jokaisen luvun lopussa on biolääketieteellisestä kirjallisuudesta poimittuja tehtäviä. Hinta \$49,95
<http://mhprofessional.com/product.php?isbn=0071435093&cat=&promocode=>
http://books.google.fi/books?id=SiJ2UB3dv9UC#reviews_anchor

Dunnin ja **Everittin** kirjan ”Clinical Biostatistics, An Introduction to Evidence-Based Medicine” alaotsikko paljastaa, että kyseessä ei ole tavanomainen oppikirja vaan nykysuuntauksen mukainen ongelmalähtöinen esitys. Kirja on mielenkiintoisesti kirjoitettu, mutta on sisällöltään se on melko suppea ja aihevalinnoiltaan tavallisista oppikirjoista poikkeava; mm. tilastollista testausta siinä ei käsitellä juuri ollenkaan, eikä monia muitakaan alkeisoppikirjojen käsittelemiä aiheita, kuten yksinkertaista regressioanalyysiä. Yksinään tämä kirja ei tutkijalle riitä. Runsas melko vaikeasti todettavien painovirheiden määrä häiritsee jonkin verran lukemista, silti suositeltavaa luettavaa. Kirjan hinta on noin \$35.

Satunnaistetut kliiniset kokeet näyttelevät tärkeää roolia uusien hoitojen ja lääkkeiden kehittämisessä. Tutkimustulosten oikeellisuus riippuu oleellisesti siitä kuinka hyvin tutkimus on organisoitu ja toteutettu. **Pocockin** kirja ”Clinical Trials, A Practical Approach” tarjoaa lyhyen mielenkiintoisen katsauksen kliinisten kokeiden historiaan ja tiiviin ja kattavan esityksen niiden yleisperiaatteista, suunnittelusta ja toteuttamisesta. Kirjoittajan laaja käytännön kokemus ja syvä perehtyneisyys aihepiiriin näkyy kirjan sisällössä ja tekstissä. Tilastollisia menetelmiä käsittelevä osa on suppea, mutta sisältää kuitenkin kaikkein olennaisimman. Hinta €48. <http://books.google.fi/books?id=Oj5rAAAAMAAJ&dq=Pocock+Clinical+Trials>
 Tekijä on kirjoittanut myös monia hyviä yleistajuisia artikkeleita tästä aiheesta (esim. Pocock 1985 ja 1987).

Lääketieteessä käytetään paljon **parametrittomia tilastollisia menetelmiä**. Niiden perusasiat on yleensä riittävästi kuvattu yllä luetelluissa perusoppikirjoissa. Mikäli joku kuitenkin haluaa yksityiskohtaisempaa tietoa näistä menetelmistä, niin kannattaa tutustua klassiseen **Siegelin** ja **Castellanin** kirjaan ”Non-parametric Statistics for the Behavioral Sciences” tai suomenkieliseen **Metsämuurosen** kirjaan pienten aineistojen analyysistä. Kirja on selkeästi ja havainnollisesti kirjoitettu ja sisältää runsaasti esimerkkejä SPSS- ohjelmiston käytöstä. Kirjan hinta on €22.

Ohjelmistoista

Mikrotietokoneiden helpot käyttöliittymät ja uudet ohjelmistot ovat viime vuosina mahdollistaneet tilastollisten menetelmien käytön kohtuullisella vaivannäöllä myös muillekin kuin ammattilaisille. Monissa ohjelmissa on lisäksi sisäänrakennettuna tilastollisten menetelmien käytöstä opastavia tutor-järjestelmiä.

Ohjelmistopaketeista **SPSS, STATA, NCSS, SAS, JMP, STATISTICA, ja S-PLUS** lienevät nykyisin tavallisimmin käytettyjä yleisohjelmia lääketieteilijöiden keskuudessa **PC-puolella**. Aiemmin runsaasti käytetty ja laadukas BMDP oli pitkään päivittämättä. Nykyisin ohjelmasta on kuitenkin saatavilla Windows-versio. Monista yllämainituista ohjelmista on tarjolla useita eri käyttöjärjestelmäversiota: Windows, UNIX, Linux. Mac-puolella tilanne on viime vuosina tilasto-ohjelmistojen tarjonnan ja laadun suhteen parantunut ja ainakin paketeista **SPSS, STATA ja STATISTICA** on myös **Mac-versiot**. Mac-puolella voidaan emulointiohjelmien avulla ajaa myös PC-ohjelmia. Päinvastaiseen ei oikeastaan ole ollut tarvetta PC-puolen paremman tarjonnan vuoksi. Osa ohjelmistoista toimii sekä valikko- että komentokielihojattuna. Aloitteleva käyttäjä tai ohjelmistoa harvemmin (ei työkseen) käyttävä

henkilö tulee aika pitkälle toimeen valikko-ohjauksella. Ajovirrat kannattaa kuitenkin silloinkin usein tallentaa, ja erityisesti silloin jos tekee aineistoon muunnoksia.

SPSS on helppokäyttöinen perusohjelmisto. Sen grafiikkaominaisuudet ovat kohtalaisen hyvät, menetelmätarjonta on monipuolinen ja tulosteiden ja kuvien siirto muihin ohjelmiin (esim. Wordiin) tai html-muotoon onnistuu helposti. Mikäli tulosteita halutaan siirtää muillekin kuin SPSS-käyttäjille, niin se kannattaa tehdä html-muodossa. Käytöltään **SPSS**, **NCSS** ja **STATISTICA** muistuttavat paljon toisiaan ja ohjelmistojen valikoista on löydettävissä suunnilleen samat analyysivaihtoehdot. Tulosteissa ja tulostettavissa suureissa on jonkin verran eroja. Pääasiassa komentokieliohjatut **SAS**, **STATA** ja **R** ovat lähinnä ammattilaisille tai työkseen tilastanalyysijä tekeville tutkijoille tarkoitettuja ohjelmistoja. Tosin molempien nykyversioita on kohtalaisen helppo myös muidenkin kuin ammattilaisten käyttää valikko-ohjattuna. Ohjelmiston valinnan ratkaisee paljolti ympärillä toimiva tiedeyhteisö ja erilaiset tottumiskysymykset.

Monet ohjelmapaketit kykenevät nykyisin aika hyvin lukemaan toinen toistensa tiedostoja. Tiedostokonversioiden tekemiseen on saatavilla kuitenkin myös erityisohjelmia, esimerkiksi **Stat/Transfer**. Aiemmin runsaasti käytetty **DBMS/Copy**-ohjelman kehitystyö on lopetettu. Sen viimeisin versio 8.0 lukee, muuntaa ja kirjoittaa yli 80:n eri tietokanta- taulukkolaskin-, tilasto- ja aikasarjaohjelmistopakettien tiedostoja.

Yleisohjelmien ohella tarjolla on myös hyviä erityisohjelmia. Esimerkiksi **eksakteihin testeihin** ja **pienen aineistojen tilastokäsittelyyn** erikoistunut **StatXact** ja **epidemiologisten aineistojen mallittamiseen** ja testaamiseen soveltuva **Egret** ja **LogXact** ovat hyviä ja paljon käytettyjä ohjelmistoja. Molemmat ohjelmat ovat etenkin tiivistetyn tiedon, kuten taulukkomuotoisten esitysten, käsittelyssä paljon helppokäyttöisempiä kuin yleispaketit. Ne ovat erittäin käteviä mm. sellaisissa tilanteissa missä alkuperäistä aineistoa ei ole käytettävissä vaan testaukset joudutaan suorittamaan jälkikäsitteilynä tiivistetystä aineistosta. Joitakin eksakteja ohjelmamoduuleita on tosin nykyisin sisällytetty myös lähes kaikkiin yleispaketteihin, mutta valikoima niissä on paljon suppeampi kuin erityispaketeissa. Etenkin laboratoriosovelluksia ajatellen kehitetty **MedCalc** on kätevä erityisohjelma mm. **diagnostisten- ja seulontatestien hyvyden arviointiin ja testaamiseen**. Monet yleispaketit, esim. SPSS ja SYSTAT tarjoavat myös mahdollisuuden analysoida ROC -käyriä, mutta vain yhden testin ollessa kyseessä. Sen sijaan esim. ohjelmilla **MedCalc** ja **GraphRoc** voi tehdä myös useampien testien välistä vertailua. Systemaattisten katsausten ja meta-analyysien tekoon on käytettävissä monia erityisohjelmia, mm. **Comprehensive Meta Analysis**.

Kliinisissä tutkimuksissa **voimalaskelmat** ovat tärkeitä. **nQuery** on tähän tarkoitukseen soveltuva helppokäyttöinen ja monipuolisia erityisohjelmia. **Egret[®] Siz** on tärkeä työväline epidemiologisten tutkimusten suunnittelussa. Se on helppokäyttöinen, tehokas ja sofistikoitu paketti aineistokoon arviointiin epälineaarisia malleja käytettäessä. Yleisohjelmilla aineistokoon arviointia voidaan joillakin paketeilla suorittaa, mutta erilaisten käyttötilanteiden ja vaihtoehtojen määrä on niukkaa. Myös verkosta löytyy sivuja voimalaskelmien tekoon, esim. <http://stat.uiowa.edu/~rlenth/Power/index.html>.

Varsinaisten tilastopakettien ohella perusstatistiikkaa voidaan laskea myös useimmilla **taulukkolaskinohjelmilla**; esim. **Excel**, joka on suunnilleen sama ohjelma sekä PC - että Mac -puolella. Excelin tilastolliset funktiot löytyvät valikosta: **“Insert “ ► ”Function”** valitsemalla kategoriaksi **“Statistical”**. Valikoima on melko suppea, mutta perehtymällä ohjelmaan vähän syvällisemmin, käyttäjä voi itse täydentää ohjelman tarjoamaa perusvalikoimaa tekemällä ns. makroja, sellaisista tilastollista suureista, joiden algoritmi (ratkaisumenetelmä) on verrattain helppo, esimerkiksi useimmat tässä monisteessa olevat laskentakaavat. Lisäksi netissä on runsaasti sivuja, joilla voi tehdä ilmaiseksi yksinkertaista

tilastollisia laskentatehtäviä. Esim. sivu <http://graphpad.com/quickcalcs> tai <http://home.clara.net/sisa/index.html>. Minkään varsinaisen tutkimusaineiston käsittelyyn näitä web-sivuja ei kuitenkaan voi suositella, sillä kukaan ei takaa ohjelmien luotettavuutta eikä syötettyjen aineistojen tietosuojaa.

Satunnaistamislistojen ("randomization list") tekeminen on keskeisellä osalla monien kliinisten tutkimusten alkuvaiheessa. On yllättävää, että useimmissa yleispaketissa ei ole kätevää tapaa tehdä edes tavallisimpiin käyttötilanteisiin sopivia satunnaistamislistoja. Komentokieliohjatuisissa ohjelmissa niiden teko ad hoc tietenkin onnistuu satunnaislukugeneraattorien avulla, mutta tutkijan pitäisi voida tehdä niitä paljon helpommin ja valikko-ohjattuna. Uudella nQuery-ohjelman versiolla 6.0 on mahdollista tehdä aineistokoon arvioinnin ohella myös satunnaistamislistoja. Verkossa niitä voi tehdä mm. osoitteissa: <http://www.randomization.com> tai <http://graphpad.com/quickcalcs>.

Nykyisin useimmat arvostetut lehdet korostavat **luottamusvälien** ("confidence interval") laskennan tärkeyttä, etenkin keskeisimmille tutkimustuloksille. Tilastollisten yleispakettien luottamusvälitarjonnassa on hämmästyttäviä puutteita tässä suhteessa. Läheskään kaikkia aivan tavallisia kliinisissä tutkimuksissa tarvittavia luottamusvälejä ei niistä löydy ja tutkija joutuu turvautumaan erityisohjelmiin tai ohjelmoimaan ne esimerkiksi Excelillä. Hyvä erityisohjelma PC -puolella luottamusvälilaskentaan on BMJ:n markkinoima **CIA-ohjelma**, johon sisältyy myös pieni kirja (Bryant TN, 2000). Ohjelma on erittäin helppokäyttöinen ja toimii Windows-käyttöjärjestelmissä <http://www.som.soton.ac.uk/cia/>.

Lääketieteellisten artikkelien laatu tilastollisesti tarkasteltuna

Lääketieteellisten artikkelien laatua biostatistiselta kannalta on tutkittu runsaasti. Vuonna 1992 ilmestyneessä tutkimuksessa (Gore, Jones ja Thompson 1992) aineistona oli 191 Lancet -lehden alustavan arvioinnin läpäissyttä artikkelia aikaväliltä marraskuu 1990 ja kesäkuu 1991. Artikkelit arvioitiin uudelleen tilastolliselta kannalta erityisen asiantuntijaraadin toimesta. Raati kiinnitti erityistä huomiota siihen olivatko tehdyt johtopäätökset perusteltuja tutkimusasetelman ja suoritettujen tilastollisten analyysien pohjalta. Tulos oli pääpiirteissään seuraava:

- Vain 54 % käsikirjoituksista katsottiin hyväksyttäväksi tai hyväksyttäväksi korjauksen jälkeen.
- Muita joko suositeltiin hylättäväksi (14 %) tai olennaista korjausta vaativiksi (32 %).
- Menetelmien kuvaukset olivat puutteellisia noin puolessa käsikirjoituksista.
- Noin joka neljännessä oli abstrakti tai johtopäätökset puutteellisia.
- Huomattavia virheellisyyksiä tilastollisessa päättelyssä ("statistical inference") esiintyi 48 artikkelissa. Näissä artikkeleissa esiintyi puutteellisuuksia myös tutkimuksen suunnittelun ("study design") suhteen.

Nykyisin monet lääkäritutkijat käyttävät tilastopaketteja sujuvasti ja asiantuntemuksella. Hyvistä tilastotieteen kirjoista ja ohjelmista on ollut selkeästi havaittavissa oleva hyöty. Ohjeita lääketieteellisiin lehtiin kirjoitettaville tutkijoille on julkaistu useita: esim. Altman, Gore, Gardner ja Pocock 1983 on suositeltavaa luettavaa. Tässä kirjoituksessa olevien ohjeiden suhteen lienee statistikkojen keskuudessa laaja konsensus. Lisää kirjallisuusviitteitä vastaavankaltaisiin artikkeleihin on mm. edellä mainituissa Langin ja Secicin sekä Byrnen kirjoissa.

Tutkimusprosessi

Tutkimusta voidaan luonnehtia prosessina, joka alkaa tutkimuksen suunnittelusta ja joka useimmiten johtaa tulosten julkaisemiseen ja niiden hyödyntämiseen. Tutkimusprosessin eri vaiheita ovat:

- Ongelman realisoituminen (tiedostamisen aste ja tärkeys)
- Tutkimuksen yleissuunnittelu (resurssien kartoitus, aikaisempien tutkimusten arviointi)
- Ongelman tieteellinen lähestymistapa (tutkimustyyppin, -asetelman ja -menetelmien valinta)
- Aineiston valinta (kohdeperusjoukon eli populaation määrittely, otantamenetelmien valinta)
- Tietojen kerääminen (mittarit ja mittaamismenetelmät, tietojen luotettavuus ja sen turvaaminen)
- Tietojen tallennus, muokkaaminen ja esikäsittely (virheiden tarkistus, poikkeavien arvojen etsintä, jakaumien muodot)
- Tulosten analysointi (menetelmävalinnat)
- Tulosten tiivistäminen ja esittäminen (tunnusluvut, taulukointitekniikka ja graafiset kuvaajat)
- Tulosten tulkinta ja johtopäätöstenteko (laajuus, yleistettävyyys ja kliininen merkitys)
- Tieteellinen julkaisu tai tutkimusraportti (julkaisufoorumien valinta)
- Tulosten käyttö (väärinkäyttö) (tiedon vastaanotto, tulkinta ja sovellukset)

Tutkimusprosessin eri vaiheisiin liittyy useita **virhe-** ja **harhamahdollisuuksia** ("bias") sekä salakuoppia, joihin kokenutkin tutkija saattaa langeta. Tilastolliseen tutkimukseen liittyvät harhat voidaan määritellä sellaisina tulkintoina ja toimenpiteinä missä tahansa tutkimusprosessin vaiheessa, jotka johtavat systemaattiseen poikkeamaan todellisuudesta. Harhasta voidaan myös käyttää nimitystä **systemaattinen virhe** ("systematic error"). Erilaisia harhatyyppejä on kirjallisuudessa kuvattu useita kymmeniä (Sackett 1979).

$$\text{(Mitattu arvo)} = \text{(Todellinen arvo)} + \text{(Harha)} + \text{(Satunnaisvirhe)}$$

Kriittisen tutkimuksen olennainen osa on harhamahdollisuuksien jatkuva arviointi tutkimuksen kaikissa vaiheissa. Harhamahdollisuuksien painottuminen vaihtelee kliinisen tutkimuksen eri osa - alueilla, tutkimustyyppistä riippuen.

Kliinisiä kokeita varten on laadittu monia normistoja. Laajimmin käytössä on **ICH-GCP** ("International Conference on Harmonization (ICH)/ WHO Good Clinical Practice standards") Tämä yhtenäinen standardi on käytössä mm. Euroopan Unionissa, Japanissa, Yhdysvalloissa, Kanadassa, Australiassa. Siinä määritellään yksityiskohtaisesti kuinka tulisi menetellä, jotta tutkimus täyttäisi kaikilta osin mahdollisimman hyvin tieteellisyyden kriteerit. Laboratoriopuolella on vastaavankaltaiset ohjeet, **GLP** ("Good Laboratory Practice"), joissa laboratoriotutkimusten ja -käytäntöjen normit. Monia näissä normistoissa määriteltyjä asioita tulisi noudattaa suunniteltaessa ja toteutettaessa myös muun tyyppisiä kliinisiä tutkimuksia. Erityisesti satunnaistettuja kokeita varten on olemassa useita ohjeistoja tutkimusprosessin eri vaiheista; suunnittelusta, protokollan laadinnasta, satunnaistamista, tilastokäsittelystä, tulosten raportointiin mm. "The **CONSORT** statement" (Begg C ym., JAMA 1996) ja monia muita on esitelty esim. Langin ja Secicin kirjassa.

Tilastollinen päättely

Tilastollinen päättely ”statistical inference” on tilastotieteen keskeisimpiä käyttötapoja. Se on päättelymekanismi, jolla kerätyn otoksen/otosten (tutkimusaineiston) perusteella tutkittavasta asiasta pyritään tekemään johtopäätöksiä (päätelmiä) laajempaa perusjoukkoa (populaatiota) koskeviksi. Tilastollisen päättelyn osa-alueet ovat arviointi (estimointi) ja hypoteesien testaaminen.

Arviointi, estimointi ”estimation” on osa tilastollista päättelyä. Sillä tarkoitetaan tutkittavan ilmiön kuvaamiseen käytetyn mallin tai jakauman sisältämien tuntemattomien suureiden, eli parametrien arviointia tutkimusaineiston (otoksen/otosten) perusteella. Estimointitapoja on kahta päätyyppiä: piste-estimointia (”point estimation”) ja luottamusväliestimointia (”confidence interval estimation”).

Piste-estimointi ”point estimation” tarkoittaa sitä, että tuntematonta suuretta, esim. jonkin taudin yleisyyttä, arvioidaan **yhdellä** kerätystä tutkimusaineistosta lasketulla lukuarvolla (estimaatilla). **Luottamusväliestimoinnissa** lasketaan nk. luottamusväli (”confidence interval”), joka sisältää tuntemattoman suureen tietyllä todennäköisyydellä, esim. 95 %:n varmuudella.

Perusjoukkoja on käsitteellisesti kahta tyyppiä: **kohdeperusjoukko** (”target population”) ja **otosperusjoukko** (”sample population”). Mikäli otanta on asianmukaisesti suoritettu siten, että tutkittavat on valittu toisistaan riippumatta ja satunnaisesti, niin otantatutkimuksen tulokset ovat yleistettävissä otosperusjoukkoa koskeviksi aineistokoon sallimalla varmuudella. Kohdeperusjoukkoa koskeviksi saatuja tuloksia ei kuitenkaan ilman muuta voida yleistää. Pitää ensin tutkia, onko otosperusjoukko mahdollisesti valikoitunut jollakin mekanismilla tavoiteperusjoukosta ja mitkä tekijät ovat olleet syynä tähän valikoitumiseen.

Esim.

Tutkitaan nivelreuman etenemiseen vaikuttavia tekijöitä ja niiden välisiä riippuvuussuhteita. Otosperusjoukon muodostaa esimerkiksi Reumasäätiön sairaalassa hoidettavat potilaat. Otos (tutkimusaineisto) kerätään jollain otantamenetelmistä näiden potilaiden joukosta. Saadut tulokset voidaan yleistää sairaalassa hoidettuihin potilaisiin, mutta ei välittömästi kaikkiin nivelreumatapauksiin. Valikoivana tekijänä on tässä tapauksessa mm. taudin vaikeusaste. Lievimpiä nivelreumatapauksia ei hoideta erikoissairaalassa ja riippuvuussuhteet saattavat lievemmissä tapauksissa olla aivan erilaiset kuin vaikeimmissa tapauksissa.

Kliinisissä tutkimuksissa, erityisesti hoitokokeissa tutkittavien joukko ei yleensä muodosta selkeää otosta mistään perusjoukosta, vaan se valikoituu sisäänotto - ja poissuljentakriteerien perusteella tutkimuksen tavoitteiden kannalta mahdollisimman tarkoituksenmukaiseksi. Tilastollisia päätelmiä tehtäessä tämä pitää ottaa huomioon, eikä yleistyksiä esim. hoitotuloksista tai hoitoeroista ei voi varauksettomasti tehdä jotakin muuta kuin tutkittavien edustamaa perusjoukkoa koskeviksi.

Esim.

SAVE -tutkimus (Moyé et al.1991): Satunnaistettu lumehoitokontroloitu kaksoissokkotutkimus ACE -estäjähoidon vaikutuksesta sydäninfarktin jälkeiseen kuolleisuuteen vasemman kammion vajaatoimintaa potevilla potilailla.

Kyseessä oli monikeskustutkimus, jonka potilasmateriaali seulottiin sepelvaltimotautihoitoyksiköistä USA:ssa kolmen vuoden ajalta. Seulottavien määrä oli kaikkiaan 95 856. Näistä 36 010 potilaalla todettiin kliininen sydäninfarkti, olivat selvinneet hengissä 72 tuntia infarktin jälkeen ja olivat iältään 21–79 vuotta (potilasjoukko A). Erilaisin poissuljennoin potilasjoukkoa rajattiin ja päädyttiin 2231 potilaaseen (potilasjoukko B), jotka satunnaistettiin SAVE -tutkimukseen. Merkittävin poissulku tapahtui, kun kaikki ne potilaat, joilla ejektiofraktio oli yli 40 %, suljettiin tutkimuksen ulkopuolelle. Potilasjoukko B oli siten 6,1 % potilasjoukosta A.

Tutkimuksessa todettiin selvä myönteinen vaikutus kuolleisuuteen ACE - estäjähoidolla. Tutkimuksen tuloksia hyödynnettiin välittömästi mm. lääkemainonnassa ja rohkeasti yleistettiin hoidon myönteiset vaikutukset lähes kaikkiin infarktitaapauksiin. Näin rajuihin yleistyksiin ei voida kuitenkaan yllä kuvatus tutkimuksen perusteella mennä, sillä todettu myönteinen tulos pätee ainoastaan sellaiseen perusjoukkoon, jota potilasjoukko B, satunnaistetut potilaat, edustaa. Myöhemmin on tosin suoritettu tutkimuksia, joiden perusteella ACE -estäjähoidon käyttöindikaatioita on voitu olennaisestikin laajentaa.

Viitteet

Altman DG, **Gore** SM, **Gardner** MJ, **Pocock** SJ. Statistical guidelines for contributors to medical journals. *BMJ* 1983; 286:1489-93.

Altman DG. *Practical Statistics for Medical Research*. Chapman & Hill, London 1991.
http://www.crcpress.com/shopping_cart/products/product_detail.asp?sku=C7630&isbn=9780412276309&parent_id=&pc

Armitage P, **Berry** G, **Matthews** JNS. *Statistical Methods in Medical Research*. 4th edition Blackwell Sci. Publ., Boston 2001. ISBN: 0632052570.
<http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0632052570.html>

Begg C, **Cho** M, **Eastwood** S, **Horton** R, **Moher** D, **Olkin** I **et al.** Improving the quality of reporting of randomized controlled trials: The CONSORT Statement. *JAMA* 1996; 276(8):637-9. <http://www.consort-statement.org/>

Byrne DW. *Publishing Your Medical Research Paper: What They Don't Teach In Medical School*. Lippincott Williams & Wilkins, Philadelphia PA 1998, ISBN: 0-683-30074-1.

Campbell MJ, **Machin** D, **Walters** SJ. *Medical Statistics, Textbook for the Health Sciences* 4th edition. John Wiley & Sons, 2007. ISBN 978-0-470-02519-2.
<http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470025190.html>

Dunn G, **Everitt** B. *Clinical Biostatistics. An introduction to Evidence-Based Medicine*. Edward Arnold, London 1999. ISBN 0-340-59531-0.

Emerson JD, **Colditz** GA. Use of statistical analysis in the *New England Journal of Medicine*. *N. Eng. J. Med.* 1983; 309:709-13.

Glantz SA. *Primer of Biostatistics*, 6th edition. McGraw-Hill Medical, 2005. ISBN: 0071447814. Hinta \$ 49,45 <http://mhprofessional.com/product.php?isbn=0071435093&cat=&promocode=>

Good Clinical Practice in FDA-regulated Clinical Trials. www.fda.gov/oc/gcp/default.htm

Gore S, Jones G, Thompson SG. Lancet's statistical review process: areas for improvements by authors. Lancet 1992; 340:100-2.

ICH-GCP Guidelines for Clinical Trials. www.vadscorner.com/internet29.html

Lang TA, Secic M. How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors and Reviewers. 2nd edition. ACP Press 2006. ISBN 1930513690.
<https://www.acponline.org/atpro/timssnet/catalog/books/statistics.htm>

Machin D and Campbell MJ. Design of Studies for Medical Research. John Wiley & Sons, 2005. ISBN 0-470-84495-7. <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470844957.html>

Moyé LA, Pfeffer MA, Braunwald E for the **SAVE** investigators. Rationale, Design and Baseline Characteristics of the Survival and Ventricular Enlargement Trial. Am J Cardiol 1991; 68: 70D-79D,

Metsämuuronen J. Tutkimuksen tekemisen perusteet ihmistieteissä. 4. painos, Gummerus Kirjapaino Oy Jyväskylä, 2009. ISBN: 987-952-5372-25-0. <http://www.methelp.com>

Metsämuuronen J. Pienten aineistojen analyysi, Parametrittomien menetelmien perusteet ihmistieteissä. Gummerus Kirjapaino Oy, Jyväskylä 2004. ISBN: 952-5372-16-2.
<http://www.methelp.com>

Pocock SJ. Clinical Trials, A practical approach. John Wiley & Sons, Chichester 1983. ISBN 0-471-90155-5. http://eu.wiley.com/WileyCDA/Section/id-WILEYEUROPE2_SEARCH_RESULT.html?query=Pocock

Pocock SJ. Current issues in the design and interpretation of clinical trials. BMJ 1985; 290: 39-42.

Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A Survey of Three Medical Journals. N Eng J Med 1987; 317: 426-32.

Ranta E, Kouki J, Rita H. Biometria -Tilastotiedettä ekologeille. Yliopistopaino 2005 9.p
<http://www.suomalainen.com/sk/servlets/ProductServlet?action=search&isbn=951570085X>

Sackett DL. Bias in analytic research. J Chron Dis 1979; 32: 51- 63.

Senn S. Cross-over trials in clinical research. John Wiley & Sons, Chichester 1993. ISBN: 0-471-93493-3. hinta € 85.80. <http://www.wileyurope.com/statistics>

Swinscow TDV & Campbell MJ. Statistics at Square One 10th edition. BMJ Books 2002. ISBN: 0-7279-1552-5. http://www.amazon.co.uk/Statistics-Square-One-Michael-Campbell/dp/0727915525/ref=sr_11_1?ie=UTF8&qid=1238770223&sr=11-1

Siegel S, Castellan NJ. Nonparametric statistics for the behavioral sciences. McGraw-Hill, Singapore 1988. ISBN 0-07-100326- 6.

Uhari M, Nieminen P. Epidemiologia & biostatistiikka. Kustannus Oy Duodecim 2001. ISBN: 951-656-040-7,Hinta: €54, D-jäsen €40,50, D-opiskelija 32,40
http://www.duodecim.fi/kotisivut/sivut.koti?p_sivusto=51

Wassertheil-Smoller S. Biostatistics and Epidemiology: A Primer for Health Professionals. 3rd edition. Springer-Verlag 2004. ISBN: 0-387-40292-6.
http://www.springer.com/life+sci/book/978-0-387-40292-5?cm_mmc=Google-_-Book%20Search-_-Springer-_-0

Ohjelmia

Analyse-it Method Evaluation Edition: on Excelille tehty ohjelmisto mm. laboratorio- ja diagnostiikkasovelluksiin, kuten ROC-käyrien vertailuun, menetelmien tarkkuuden, toistettavuuden ja lineaarisuuden arviointiin. http://www.analyse-it.com/products/method_evaluation/ Hinta € 499.

BMDP Statistical Software : Uusittu versio perinteikkäästä ohjelmasta.
<http://www.statsol.ie/index.php?pageID=9&productID=6&productContentID=316> Hinta €495.

CIA 2.1.1: (Confidence Interval Analysis) on hyvä ja helppokäyttöinen erityisohjelma luottamusvälien laskentaan ja satunnaistamistolistojen tekemiseen. Ohjelma on oheistuotteena kirjassa: Statistics with Confidence, Toim. Altman DG, Machin D, Bryant TN & Gardner MJ. BMJ books 2000. ISBN: 0-7279 -375 -1. Hinta £22.94.
http://www.bmjbookshop.com/shop/product_display.asp?SiteLanguage=ENG&ProductID=0727913751 Ohjelmapäivityksiä voi tilata osoitteesta: <http://www.som.soton.ac.uk/cia/>

FigSys 2.4: Biosoft, <http://www.figsys.com/features.html>
 Hyvä julkaisugrafiikkaohjelma. Hinta \$300.

Comprehensive Meta Analysis: Meta-analyysien tekoon tarkoitettu erityisohjelma. Hinta €995. <http://www.statsol.ie/index.php?pageID=25>

JMP 8.0.1 SAS Institute. <http://www.jmp.com/> Hyvät visualisointimahdollisuudet omaava ohjelmistopaketti. Helsingin yliopistolla on ohjelmaan kampuslisenssi.
<http://www.helsinki.fi/atk/myynti/index.shtml>

NCSS: Helppokäyttöinen ja tilasto- ja grafiikkaohjelmisto. Hinta €395.
<http://www.statsol.ie/index.php?pageID=24>

nQuery Advisor 7.0 Statistical Solutions 2008. Monipuolinen ja kätevä aineistokoon arviointiohjelma. Hinta €1295 <http://www.statsol.ie/index.php?pageID=2>

Origin 8: OriginLab Corporation. Laadukas julkaisugrafiikkaohjelma.
<http://www.originlab.com/index.aspx?s=8> Helsingin yliopistolla on ohjelmaan kampuslisenssi.
<http://www.helsinki.fi/atk/myynti/index.shtml>

MedCalc: on helppokäyttöinen biolääketieteen tutkijoille suunniteltu Windows-ohjelma, joka sisältää mm. Bland & Altman- plotit ja ROC-käyrien vertailut, jotka monista yleispaketeista puuttuvat. <http://www.medcalc.be> Hinta US \$395

SAS 9.2: SAS Institute, SAS on mm. lääketeollisuudessa yleisesti käytetty tilastopaketti.
<http://www.sas.com/software/sas9> Helsingin yliopistolla on ohjelmaan kampuslisenssi.
<http://www.helsinki.fi/atk/myynti/index.shtml>

SOLAS 3.0: Statistical Solutions, on monipuolinen puuttuvien tietojen käsittelyyn tarkoitettu erityisohjelma. <http://www.statsol.ie/index.php?pageID=5> Hinta: €1295.

SPSS 18: SPSS Inc. Chicago, IL, USA, on monipuolinen tilasto-ohjelmistopaketti opetus- ja tutkimuskäyttöön. <http://www.spss.fi/products/base/> Helsingin yliopistolla on ohjelmaan kampuslisenssi. <http://www.helsinki.fi/atk/myynti/index.shtml>
Opetusmateriaalia ja käyttöohjeita: <https://www.washington.edu/uware/spss/docs/>
<http://www.ats.ucla.edu/stat/spss/> <http://www.helsinki.fi/atk/tilasto/spssjatko.html>.

STATA 10: Stata Statistical Software. Monipuolinen tilastollinen perusohjelmisto, joka soveltuu vaativampaankin tutkimustyöhön. <http://www.stata.com/>

STATISTICA 8: StatSoft Scandinavia AB, perusohjelmisto, jonka etuja ovat mm. hyvä grafiikka. <http://www.statsoft.se/>

Stat/Transfer 9 Helppokäyttöinen tiedostojen konversio-ohjelma. <http://www.stattransfer.com/>
Hinta: Akateeminen yksittäislisenssi \$179, opiskeluhinta \$59.

StatXact 8: Cytel Inc. Cambridge, MA 02139, USA. Erityisohjelmisto pienten tai tiivistetyssä muodossa (taulukot jne.) olevien aineistojen käsittelyyn.
<http://www.cytel.com/Products/StatXact/default.asp>

Erilaisia **Ilmaisohjelmia ja -ohjelmistoja** on tarjolla netissä runsaasti. Esimerkkejä:

- **Epi Info** 3.4.3: <http://www.cdc.gov/epiinfo/> CDC:n ylläpitämä ohjelmisto epidemiologiseen tilastokäsittelyyn
- **R:** Ohjelman S-PLUS ei-kaupallinen versio, jonka voi imuroida osoitteesta: <http://www.r-project.org/>
- **SISA:** "Sample Interactive Statistical Analysis" <http://home.clara.net/sisa/>
- **StaTable:** Cytel Software Corporation, MA, USA 1998, Windows ohjelma, jonka avulla voidaan laskea 25:een eri tilastolliseen todennäköisyysjakaumaan liittyviä todennäköisyyksiä. Tämä ilmaisohjelma on imuroitavissa osoitteesta: <http://www.cytel.com/Products/StaTable/>
- **Quickcalcs:** <http://graphpad.com/quickcalcs> Hyvä yksinkertainen tilastolaskin mm. p-arvojen, luottamusvälien laskentaan.

Lisää vinkkejä tilastollisesta kirjallisuudesta ja ohjelmista löytyy mm. sivuilta: <http://statpages.net> , <http://www.stat.sc.edu/rsrch/gasp/> ja <http://www.lehanathabane.com/personal/metalinks.htm> (meta-analyysi)

Helsingin yliopiston opiskelijat ja työntekijät voivat hankkia ohjelmia Tietotekniikkaosastolta: <http://www.helsinki.fi/atk/myynti/index.shtml>

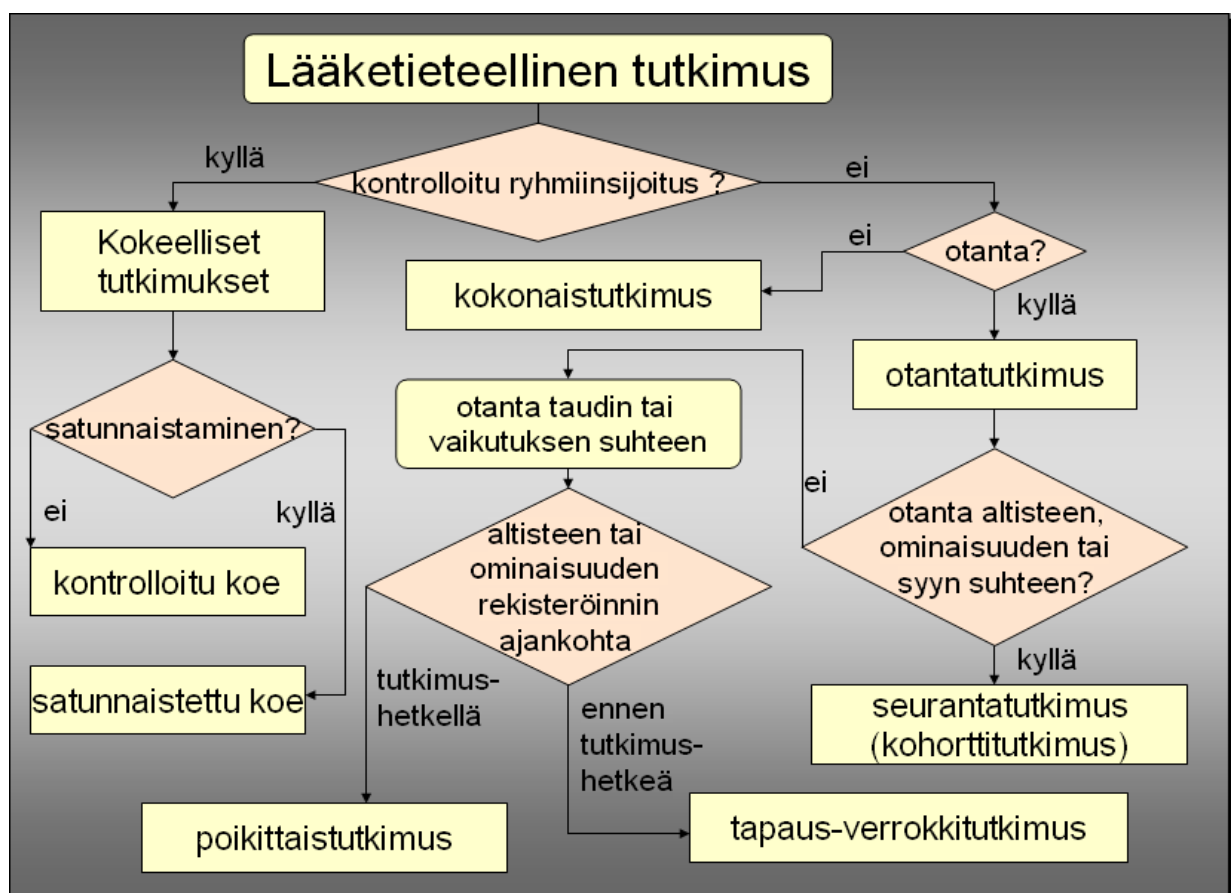
Tutkimustyytit ja -asetelmat

Yleistä

Lääketieteessä samaa ongelmaa voidaan lähestyä monin eri tutkimustyyteihin ja -asetelmin. Tilastollista käsittelyä vaativat lääketieteelliset tutkimukset voidaan jakaa kahteen päätyyppiin: **1) kokeelliset** ("experimental", "intervention") tutkimukset ja **2) havainnoivat** ("survey") tutkimukset. Kokeellisille tutkimuksille on tyypillistä tutkijan aktiivinen rooli havaintoyksiköiden (esim. potilaiden) jakamisessa tutkittaviin ryhmiin ja koejärjestelyitä ja -olosuhteita luonnehtivien parametriasetusten määrittelyssä (esim. annosten määrittely annos- vaste - tutkimuksissa). Kokeellisissa tutkimuksissa tutkittavien (koe - eläimet, ihmiset jne.) jako ryhmiin tapahtuu koesuunnitelman mukaisesti.

Havainnoivissa tutkimuksissa tutkijalla on passiivisempi rooli rekisteröidessään tutkittavaan ilmiöön liittyviä tapahtumia tutkimussuunnitelman mukaisissa ryhmissä. Näissä tutkimuksissa tutkijalla ei ole mahdollisuutta suoranaisesti vaikuttaa tutkittavien jakautumiseen tutkittaviin ryhmiin, esimerkiksi jollekin riskitekijälle **altistuneisiin** ("exposed group") ja **altistumattomiin** ("non-exposed group").

altistuneiden ryhmä "exposed group" on niiden henkilöiden ryhmä, jotka ovat tai joiden oletetaan olevan altistunut jollekin agenssille, jonka tuottaa sellaisen kemiallisen, fyysisen tai biologisen vaikutuksen, jolla on joko kausaalinen tai muu yhteys tautiriskiin; esimerkiksi ravinnon runsaan suolapitoisuuden vaikutus verenpainetaudin kehittymiseen. Vastakohta on **altistumattomien ryhmä "non-exposed group"**.



Havainnoivia tutkimuksia on kahta tyyppiä:

- **kokonaistutkimukset** ("population studies") ja
- **otantatutkimukset** ("sample studies").

Otantatutkimukset voidaan edelleen jakaa kahteen tyyppiin riippuen siitä suoritetaanko otanta:

- taudin tai vaikutuksen vai
- altistustekijän, ominaisuuden tai oletetun syytekijän suhteen.

Tutkimuksen tietojenkeruun aikasuuntauksen mukaan tutkimukset voidaan edelleen jakaa:

- **poikittaistutkimuksiin** ("cross-sectional study"),
- **takautuviin** ("retrospective study"),
- **eteneviin** ("prospective study") ja
- **molempisuuntaisiin** ("ambispective study") tutkimuksiin.

Kokeelliset tutkimukset

Kokeellinen tutkimusasetelma "experimental study design" on ihmisillä, koe - eläimillä, koemaljoilla tai muilla koeyksiköillä suoritettu tarkkaan tutkijan suunnittelemaan koejärjestelyyn perustuva asetelma. Kokeellisessa asetelmassa tutkija määrää koeasetelmillaan kuka altistuu ja kuka ei tutkittavalle tekijälle (esim. tietylle hoidolle).

Käyttö: Lääke - ja hoitokokeet, interventiotutkimukset (väestökokeet)

Havaintoyksiköt: koe - eläimet, ihmiset, mallit

Kliininen koe ("Clinical trial") on mikä tahansa ihmisillä suoritettu kokeellinen tutkimus, joka perustuu tarkkaan etukäteissuunnitelmaan ("protocol") ja jonka tavoitteena on löytää mahdollisimman hyvä hoitomenetelmä uusille tuleville potilaille. Lääkekokeet luokitellaan tavallisesti koesuunnittelultaan neljään eri vaiheeseen (faasiin, "phase"). Faasi I:n tutkimukset suoritetaan tavallisesti terveillä vapaaehtoisilla ja ne tähtäävät lähinnä lääkkeen imeytymisen ja metaboloitumisen selvittelyyn sekä turvallisten lääkeannosten löytämiseen. Faasi II:n pienimuotoisilla, aineistokooltaan yleensä korkeintaan 100 -200 potilaan tutkimuksilla haetaan alustavaa selvyyttä lääkkeiden tehoon ja sivuvaikutuksiin. Faasi III:n tutkimukset ovat täyslaajuisia lääkkeen tai hoidon tehon luotettavaan selvittelyyn tähtäviä tutkimuksia. Lääkkeen markkinoille tulon tai hoidon laajamittaisen käyttöönoton jälkeisiä seuranta - ja monitorointitutkimuksia kutsutaan faasi IV:n tutkimuksiksi.

Huomattava osa faasi III:n tutkimuksista on sellaisia, että niissä vertaillaan kokeiltavana olevaa hoitoa tai interventioimenpidettä johonkin kontrollina olevaan hoitoon tai toimenpiteeseen. Vertailtavia ryhmiä voi olla myös useita ja niiden tulee olla mahdollisimman samankaltaisia muiden paitsi tutkittavan tekijän (esim. lääkehoidon) suhteen. Vertailtavien ryhmien samankaltaisuudella taustatekijöiden (esim. ikä, sukupuoli, taudin vaikeusaste) suhteen pyritään varmistamaan se, että mikäli tutkittavien ryhmien välillä todetaan ero, niin se todella johtuu tutkittavasta tekijästä eikä taustatekijöistä. Samankaltaisuus pyritään takaamaan siten, että tutkimukseen soveltuvien henkilöiden sijoittelussa vertailtaviin ryhmiin käytetään **satunnaistamismenettelyä** ("randomization"). Valtaosa faasi III:n tutkimuksista on nykyisin satunnaistettuja.

Yksi satunnaistamismenettelyn syistä on estää subjektiivisuuden vaikutus tutkimuksen tuloksiin. Satunnaistamismenettelyllä vertailtavista ryhmistä tulisi lähtötilanteessa täysin samankaltaisia, mikäli satunnaistamismenettelyä voitaisiin jatkaa rajatta. Käytännössä kuitenkin joudutaan tyytymään melko pieniin ryhmäkokoisiin ja huomattaviakin eroja taustatekijöiden suhteen saattaa ryhmien välillä esiintyä. Jokaisessa satunnaistetussa kliinisessä kokeessa on syytä suorittaa ryhmien vertailu kaikkien relevanttien taustatekijöiden suhteen, jotta voitaisiin varmistua siitä, kuinka hyvin ryhmät ovat lähtötilanteessa tasapainossa.

Mikäli käytetään yksinkertaista satunnaistamista ja ryhmäkoot ovat pieniä, niin ryhmistä voi tulla huomattavastikin erisuuret. Samansuuruisten ryhmien aikaansaamiseksi täytyy käyttää nk. rajoitettuja tai **lohkosatunnaistamismenetelmiä** ("block randomization").

Lohkosatunnaistaminen "block randomization" on satunnaistamismenetelmä, missä yhtä suurien ryhmäkokojen turvaamiseksi, tutkittavat satunnaistetaan useamman koehenkilön muodostamissa ryhmissä, lohkoissa ("block"). Lohkokoko riippuu vertailtavien ryhmien määrästä, esim. jos hoitoja on 2, voidaan käyttää lohkokokoja 2 tai 4, jne. ja kolmen hoidon tapauksessa lohkokokoja 3 tai 6, jne. Tutkittavien määrä on tasapainossa aina lohkon päätyttyä.

Mikäli satunnaistaminen on suoritettu siten, että tutkittavat henkilöt eivät tiedä mihin ryhmään he kuuluvat, mutta tutkija ja/tai muu hoitohenkilökunta tietää, kyseessä on **yksöissokkoke** ("single blind"), mutta jos kumpikaan osapuoli ei tiedä ryhmiin sijoittelua ennen tutkimuksen päättymistä (mieluiten ennen tutkimusten analysointia), niin kyseessä on **kaksoissokkoke** ("double blind"). Näistä jälkimmäinen sisältää vähemmän harhamahdollisuuksia.

Yksinkertaisen satunnaistamismenettelyn asemesta voidaan käyttää monimutkaisempiakin menetelmiä takaamaan vertailtavien ryhmien vertailukelpoisuutta lähtötilanteessa. Eräs tällainen menettely on **ositettu satunnaistaminen** ("stratified randomization"). Tällöin potilaat jaetaan taustatekijöiden mukaan ositteisiin ja suoritetaan satunnaistaminen kussakin ositteessa erikseen. Kliinisessä kokeessa päästään harvoin täysin satunnaiseen valintaan. Varsin tavallinen on esim. tilanne, että tutkija ottaa hoitoryhmään ne potilaat, jotka tulevat hänen klinikalleen ja naapuriklinikalla hoidetut potilaat muodostavat verrokkiryhmän. Tällainen asetelma on kuitenkin varsin altis kritiikille ja harhamahdollisuuksille.

Toisinaan on epäeettistä tai mahdotonta suorittaa puhtas satunnaistaminen (esim. fataalit sairaudet, potilaan saama hyöty jne.). Tällöin tutkimustuloksia analysoitaessa tulisi tutkia huolella ryhmien vertailukelpoisuutta ja käyttää tarvittaessa erilaisia korjaus- eli adjustointitoimenpiteitä analyysivaiheessa, esim. erilaisten tilastollisten mallien avulla.

Monissa lääkekokeissa voidaan säästää tutkittavien määrässä suorittamalla nk. **ristikkäistutkimus** ("cross-over trial"). Tässä tutkimussuunnitelmassa potilaat jaetaan satunnaistamalla tutkimuksen alkaessa esim. ryhmiin A ja B, jotka saavat tiettyä hoitoa määrätyn pituisen ajan, jonka jälkeen seuraa **puhdistumisajanjakso** ("wash-out-period"), jolloin koehenkilöt eivät saa mitään tutkimustuloksiin vaikuttavaa hoitoa. Tämän jälkeen seuraa kolmas tutkimusajanjakso, jolloin ne potilaat, jotka aloittivat hoidolla A saavat ajanjaksona hoitoa B ja päinvastoin. Ristikkäistutkimusasetelmassa ei siten ole erillistä kontrolliryhmää, vaan potilaat toimivat itsensä kontrolleina. Lueteltujen kolmen tutkimusvaiheen lisäksi tutkimukseen liitetään usein lumehoitojaksot eri tutkimusvaiheiden alkuun ja loppuun.

Ristikkäistutkimusasetelma soveltuu sellaisiin tutkimuksiin, joissa hoidon tai lääkkeen vaikutus on lyhytkestoinen, sillä muuten ei voida yhdistää tutkimustuloksia tutkimusjaksoista 1 ja 3 ja

siten menetetään asetelman käytön peruste, säästövaikutus potilasmäärissä. Mikäli edellisestä tutkimusvaiheesta lääkkeen tai hoidon efekti jää päälle (**”carry-over effect”**), niin tutkimusperiodi pitää ottaa huomioon tutkimustuloksia analysoitaessa. Tällaisissa tilanteissa tutkimustulosten tulkinta on usein vaikeaa.

Kliinisissä tutkimuksissa suoritetaan usein samoista potilaista toistomittauksia, käytetään **toistomittausasetelmia** (”repeated measures designs”). Niiden käytön perusideana on vähentää yksilöiden välisen vaihtelun vaikutusta tutkittavaan asiaan, esimerkiksi hoitojen vaikutukseen. Samoin kuin ristikkäisasetelmassa, toistoasetelmassakin aikaansaadaan säästöjä koe -eläin tai henkilömäärissä. Sama **voima** (”power”) saavutetaan pienemmällä tutkittavien määrällä kuin jos jokaisesta tutkittavasta olisi vain yksi mittausarvo. Toistoasetelmat antavat monia mahdollisuuksia tutkia mm. ajan ja erilaisten olosuhteiden sekä niiden **yhdysvaikutusten** (”interaction”) vaikutusta koe - tai hoitotuloksiin.

Valtaosa kliinisistä kokeista ovat luonteeltaan vertailevia, ja todettu ero ryhmien (esim. hoito - ja vertailuryhmän) välillä voi johtua seuraavista syistä:

- otosvaihtelusta tai sattumasta
- sisäisistä eroavuuksista ryhmien välillä (esim. prognoosiin vaikuttavat tekijät)
- eroista ryhmien käsittelyssä ja seurannassa (esim. hoitoryhmän intensiivisempi seuranta)
- todellinen ero ryhmien välillä tutkittavan asian suhteen

Kokonaistutkimukset

Kokonaistutkimusasetelma ”population study” on asetelma, missä tutkitaan kaikki tutkimuksen kohteena olevaan perusjoukkoon kuuluvat henkilöt. Kokonaistutkimuksia käytetään silloin, kun halutaan saada ehdottoman tarkka tieto tutkittavista asioista. Asetelmaan on harvoin tarvetta kliinisissä tutkimuksissa, yleensä otantatutkimus riittää.

Käyttö: hallinnolliset kartoitukset, väestötilastot

Havaintoyksiköt: henkilöt, kuolintodistukset, sairaalakäynnit, väestö - ja tautiryhmät

Kokonaistutkimus luo usein hypoteeseja ja antaa vihjeitä mahdollisista riippuvuuksista sekä tarjoaa tutkimusaineistoja muun tyyppisille tutkimuksille. Kokonaistutkimuksen haittoja ovat kalleus, hitaus ja usein laadunvalvonnan vaikeus.

Otantatutkimukset

Otantatutkimus ”sample study” on otantaan perustuva epäkokeellinen tutkimusasetelma, jonka tavoitteena on tehdä luotettavia perusjoukkoa koskevia johtopäätöksiä tutkimalla huolella valittu ja sopivan kokoinen osa perusjoukkoa. Otantatutkimuksen tulokset eivät koskaan ole tarkkoja vaan niihin sisältyy aina otannasta aiheutuvaa vaihtelua.

Poikittaistutkimukset

Poikittaistutkimukset (”cross-sectional studies”) perustuvat tutkimusasetelmaan, jossa ei ole aikasuuntausta, kaikki mittaukset suoritetaan tutkimushetkellä. Asetelmaa käytetään

pääasiassa silloin, kun halutaan tutkia eri asioiden ja ominaisuuksien yleisyyttä ja niiden välisiä riippuvuussuhteita.

Käyttö:

- Yleiskartoitukset
- Vallitsevuudet ("prevalence")
- Riippuvuudet tautien tai riskitekijöiden välillä
- Seurantatutkimusten osana

Koska poikittaistutkimukset eivät sisällä aikatekijää, ei todettujen riippuvuuksien perusteella voida päätellä mitään syy-seuraus-suhteesta. **Kohorttitutkimusten** (seurantatutkimusten) tiedonkeruujärjestelmä pohjautuu usein poikittaistutkimuksiin.

Tapaus-verrokkitutkimukset

Tapaus-verrokkitutkimusasetelma "case-control" tai "case referent -study design" on sairauslähtöinen tutkimus asetelma, joka suoritetaan siten, että valitaan joukko tutkimuksen kohteena olevaa tautia sairastavia henkilöitä ("tapaukset") ja niille mahdollisimman samankaltainen vertailuryhmä ("verrokki"); samankaltainen muiden paitsi tutkittavan taudin suhteen.

Tapaus- verrokki -tutkimusta nimitetään myös sairauslähtöiseksi tutkimukseksi. Tutkimuksessa kerätään altistus - ja riskitekijätietoja kummastakin ryhmästä (yleensä takautuvasti, retrospektiivisesti) ja tutkitaan, esiintyykö jokin tietty tekijä tai tekijäyhdistelmä useammin tai voimakkaampana tapauksilla kuin verrokeilla.

Asetelmasta on olemassa kaksi varianttia:

- **kaltaistettu** ("matched") ja
- **kaltaistamaton** asetelma ("unmatched")

Kaltaistetussa asetelmassa tapaukset ja verrokki on valittu siten, että ne ovat samankaltaisia tiettyjen tekijöiden (ikä, sukupuoli jne. suhteen). Kaltaistus on eräs vakiointikeinoista, jolla pyritään poistamaan kaltaistustekijöiden vaikutus tutkittavasta ilmiöstä, eli riski - tai altistustekijän ja taudin välisestä yhteydestä. Yleensä tekijöitä, joiden suhteen kaltaistus suoritetaan, ei saisi olla kovin monta, sillä muuten vaanii ylikaltaistuksen vaara; kaltaistuksella poistetaan osa tutkittavaa vaikutusta. Kaltaistamattomassa asetelmassa tutkittavien tulisi olla otoksia vastaavista perusjoukoista.

Tapaus -verrokkitutkimusta edeltää tavallisesti **sairastavuus tai kuolleisuustutkimus**, jonka pohjalta todettu (tai todetut) yhteydet tiettyjen altistus- tai riskitekijöiden ja taudin (tai tekijöiden aiheuttaman vaikutuksen) välillä halutaan varmentaa. Pääasiallinen ero poikittaistutkimukseen verrattuna on aikatekijän olemassaolo. Tapaus -verrokkitutkimuksesta saatavien tulosten oikeellisuuden kannalta keskeisiä seikkoja ovat:

- tapausten ja verrokkien määrittelykriteerin (tai kriteerien) oikeellisuus
- sekä tapaus - että verrokkiryhmän edustavuus
- tapaus- ja verrokkiryhmän vertailukelpoisuus
- altistus - ja riskitekijätietojen tarkkuus ja oikeellisuus

Kaikkiin näihin kohtiin liittyy lukuisia **harhamahdollisuuksia**. Mikäli ryhmien määrittelykriteeri (tai kriteerit) ovat puutteellisia, saattaa osa tapauksista olla sellaisia, ettei heillä olekaan tutkimuksen kohteena olevaa tautia (väävät positiiviset), vastaavasti osa verrokkiryhmään kuuluvista saattaa kuulua sinne virheellisesti (väävät negatiiviset). Nämä

luokitusvirheet voivat peittää alleen todellisen riippuvuuden joko täydellisesti tai osittain tai ne saattavat synnyttää harhaisen riippuvuuden altistus - tai riskitekijöiden ja taudin välillä.

Mikäli **edustavuus** on puutteellinen, putoaa tutkimukselta pohja pois joko täysin tai osittain, sillä tilastollisessa tutkimuksessa pyritään siihen, että todetut tulokset voidaan yleistää perusjoukkoa koskeviksi.

Verrokkiryhmän valinta tapaus -verrokkitutkimuksissa on usein vaikea tehtävä. Usein verrokkiryhmä joudutaan keräämään eri lähteistä kuin tapaukset ja tällöin vertailukelpoisuus saattaa olla puutteellinen. Riippuvuudet voivat tällöin tulla joko liian voimakkaiksi tai liian heikoiksi riippuen siitä onko verrokkiryhmä liian ”huono” tai ”hyvä”. Tapaus-verrokkiasetelma edellyttää, että verrokeilla on ollut sama mahdollisuus sairastua kuin tapauksilla ja siten verrokkiryhmä tulisi valita sellaisten henkilöiden joukosta, jotka vastaavat taustoiltaan, esim. demograafisilta tekijöiltään, tapauksia.

Tietojen keruu tapaus -verrokkitutkimuksissa joudutaan usein suorittamaan takautuvasti ja tällaisten historiatietojen luotettavuus ei aina ole paras mahdollinen. Myös tästä saattaa syntyä harhaisia tutkimustuloksia. Esimerkiksi tapauksilta altistus - ja riskitekijätiedot saattavat olla huolellisemmin ja/tai useammin tallennettuja kuin verrokeilta ja haastattelu - ja kyselytutkimuksissa tapaukset saattavat muistaa verrokkeja paremmin altistus - ja riskitekijähistoriansa. Tällöin seurauksena on harhainen tai liian voimakas riippuvuus altistus - tai riskitekijän ja taudin välillä.

Myös usein ne henkilöt, joilla on sekä tutkimuksen kohteena tauti että jokin siihen liittyvä riskitekijä, saattavat tulla suuremmalla todennäköisyydellä valituksi tutkimusaineistoon (otokseen) kuin ne henkilöt, joilla on pelkästään ko. tauti tai riskitekijä. Tuloksena on harhaisen korkea riskitekijän esiintyvyys tapausten joukossa.

Tapaus -verrokkitutkimuksessa ei ole mahdollista suoraan arvioida altistus - tai riskitekijän vaikutuksen suuruutta, lisääntynyttä riskiä tapauksilla. Epäsuorasti arviointi on kuitenkin mahdollista. Arviointi perustuu olettamuksiin:

- taudin **ilmaantuvuus** on alhainen perusjoukossa
- sekä tapaukset että verrokkit muodostavat kumpikin erikseen **satunnaisotoksen** vastaavista perusjoukoista

Vain harvoissa tapaus -verrokkitutkimuksissa tämä jälkimmäinen oletamus on tarkalleen voimassa, mutta käytännössä poikkeamat siitä eivät yleensä ole ratkaisevia tulosten merkitsevyyden kannalta.

Tapaus -verrokki -asetelman **etuja** ovat:

- taloudellisuus
- nopea toteuttaminen
- riittävät määrät tapauksia helposti saatavissa
- aineisto helposti kerättävissä.

Kohorttitutkimukset, seurantatutkimukset

Seurantatutkimusasetelma ”follow-up study design” on epäkokeellinen tutkimusasetelma, missä henkilöt valitaan heidän altistusstatuksensa perusteella. Henkilöitä seurataan tietty ajanjakso, jonka riittävä pituus arvioidaan tutkimuksen voimalaskelmien yhteydessä. Seurannan kuluessa rekisteröidään tutkittaville ilmaantuvat tautitapahtumat.

Sana **kohortti** merkitsee henkilökuntaa, jolla on jokin yhteinen ominaisuus ja jota seurataan tietyn ajan tutkimuksen kohteena olevien tapahtumien rekisteröimiseksi. Tämä yhteinen ominaisuus voi olla esimerkiksi tietyn altistus - tai riskitekijän olemassaolo tai puuttuminen tai niiden voimakkuus. Pääasiallinen ero tapaus-verrokkitutkimukseen verrattuna on, että vertailtavat ryhmät muodostetaan altistus - tai riskitekijätietojen perusteella ja seurannan tarkoituksena on rekisteröidä kuinka paljon tutkimuksen kohteena olevaa tautia ilmaantuu eri ryhmissä ja tehdä tämän perusteella johtopäätökset altistus - tai riskitekijän ja taudin ilmaantuvuuden välisestä yhteydestä.

Kohorttitutkimusasetelmassa on mahdollista suoraan arvioida johonkin altistus - tai riskitekijään (olemassaoloon tai voimakkuuteen) liittyvä lisääntynyt tai vähentynyt sairastumisriski ja siten riskierot eri ryhmien välillä.

Tietojen keruu kohorttitutkimuksessa voi tapahtua ajallisesti myös takautuvasti tai molempisuuntaisesti. Kohorttitutkimus aloitetaan lähtötutkimuksella (poikittaistutkimus), jonka yhteydessä

- varmistetaan, että seurattavista ryhmistä suljetaan pois ne henkilöt, jotka tutkimuksen lähtötilanteessa sairastavat seurannan kohteena olevaa tautia tai eivät voi siihen jostain teknisestä syystä johtuen sairastua
- määritellään altistus - tai riskitekijätietoihin perustuvat kriteerit, joiden perusteella seurattavat ryhmät muodostetaan.

Tässä tutkimusasetelmassa on lukuisia **etuja** verrattuna tapaus-verrokkiasetelmaan:

- voidaan tutkia ryhmien vertailukelpoisuutta jo tutkimuksen lähtöhetkellä ja tarkastella jo ennakkoon, minkä muuttujien suhteen erilaiset korjaukset (adjustoinnit) analyysivaiheessa ovat tarpeen
- lisääntyneen riskin suora arviointi on mahdollista
- on helpompi välttää harhaisuutta kerättävässä tiedossa (esim. muistamattomuusharha)
- on helpompi kartoittaa altistus - ja riskitekijöissä seurannan aikana tapahtuneita muutoksia
- voidaan tutkia samanaikaisesti myös muita tauteja ja niiden yhteyksiä varsinaiseen tutkimuksen kohteena olevaan tautiin
- toisin kuin useimmissa tapaus-verrokkitutkimuksissa tutkimusaineistosta ei valikoidu pois henkilöitä, jotka ovat esimerkiksi eläneet hyvin lyhyen ajan altituksesta.

Asetelman **haittoja** ovat:

- tarvitaan tapaus-verrokki -tutkimuksiin verrattuna huomattavasti suurempi tutkimusaineisto ryhmien välisten riskierojen tai riskisuhteen merkitsevyyden toteennäyttämiseksi. Tästä on seurauksena tutkimustulosten viivästyminen ja tutkimuksen kalleus
- tutkimukseen kuuluvien henkilöiden tieto tutkimukseen kuulumisestaan saattaa aiheuttaa sen, että henkilöt muuttavat elintapojaan ym., jolla saattaa olla vaikutusta sairastumisriskiin
- on mahdollista, että otantasuhde tulee eri ryhmissä erilaiseksi valikoitumismekanismeista johtuen. Esimerkiksi lähtötutkimuksessa, missä seurattavat ryhmät muodostetaan, terveet tupakoitsijat saattavat jättää vastaamatta kyselyyn useammin kuin terveet tupakoitsemattomat (esimerkiksi johtuen kyselylomakkeen komplisoidusta rakenteesta tupakkamuuttujien osalta). Tästä aiheutuu keinotekoisesti terveiden liian suuri edustus tupakoitsemattomien ryhmässä.

Meta-analyysi

Meta-analyysi ”meta-analysis” on analyysi, joka tehdään kahden tai useamman samaa ongelmaa tarkastelevan tutkimuksen tulosten pohjalta. Tavoitteena on tehdä johtopäätöksiä, jotka olisivat luotettavampia kuin mihin yksittäisissä tutkimuksissa on päädytty ja analysoida lisäksi yksittäisten tutkimustulosten välistä vaihtelua. Meta-analyysissä on osa laajempaa kokonaisuutta, josta käytetään nimitystä systemaattinen katsaus, ts. meta-analyysissä käytetty aineisto on kerätty systemaattisen katsauksen käyttämin menettelytavoit. Meta-analyysi pyrkii tuottamaan yhdistetyn arvion lopputulosmuuttujasta, esim. hoidon tehokkuudesta, käyttäen mm. Mantel–Haenszelin tai Peton menetelmiä.

Yleistä

Kun tarkastellaan useiden samoja hoitoja vertailevien kliinisten tutkimusten tuloksia, niin voidaan joskus todeta yllättävänkin suurta vaihtelua todetuissa hoitoeroissa. Osa tästä vaihtelusta selittyy luonnollisesti satunnaisvaihtelulla, mutta osa tutkimuksissa käytettyjen potilasaineistojen eroilla. Potilasaineistoissa voi olla huomattaviakin eroja vaikka tutkimusprotokollat olisivat lähes samanlaisia. Viimeisten kymmenen vuoden aikana on ollut voimakas suuntaus kehittää tilastollisia analyysimenetelmiä, joiden avulla voitaisiin formaalisesti analysoida kaikki tieto sekä julkaistuista että julkaisemattomista samoja hoitoja vertailevista tutkimuksista. Analyysistä käytetään nimitystä **meta-analyysi** (”meta-analysis”). Se tarkoittaa analyysiä, joka suoritetaan kahden tai useamman samaa ongelmaa tarkastelevan tutkimuksen tulosten pohjalta. Tavoitteena on tehdä johtopäätöksiä, jotka olisivat luotettavampia kuin mihin yksittäisissä tutkimuksissa on päädytty ja analysoida lisäksi yksittäisten tutkimustulosten välistä vaihtelua. On tärkeää selvittää vaihtelevatko tulokset esim. tutkimusaineiston koon tai jonkin muun tutkimuksen erityispiirteiden suhteen.

Perussyöt meta-analyysin suorittamiselle ovat siten:

- lisätä voimaa (vähentää väärin negatiivisten johtopäätösten todennäköisyyttä)
- lisätä tarkkuutta (parantaa estimaattien luotettavuutta)
- vähentää harhan mahdollisuutta (vähentää väärin positiivisten johtopäätösten todennäköisyyttä)

Meta-analyysin tilastollisen perustan on esittänyt esimerkiksi **Fleiss** (1993).

Yleensä tutkimuksia suunniteltaessa ei erityisesti oteta huomioon, että tuloksia tulisi mahdollisesti myöhemmin käyttämään meta-analyysin aineistona (ad hoc meta-analysit). Vastakohtana ovat sellaiset meta-analyysit, jotka perustuvat kahteen tai useampaan tutkimukseen, jotka on erityisesti suunniteltu myöhempää meta-analyysikäyttöä varten (”designed meta-analysis”)

Käytettyjen tietolähteiden perusteella meta-analyysit luokitellaan A) kirjallisuuteen perustuviin, B) tutkijoilta saatuihin yksittäisiin potilastietoihin perustuviin, ja C) tutkijoilta tai tutkimusrekistereistä saatuihin yhteenvetotietoihin perustuviin.

Tyyppin A tutkimuksiin sisältyy mm. seuraavia vaaroja ja ongelmia:

- Positiivisia tuloksia julkaistaan lääketieteellisissä lehdissä herkemmin kuin negatiivisia (Begg ja Berlin 1988.). Näin ollen on helpompi saada julkaistuksi sellaisen tutkimuksen tulos, jossa todetaan merkitsevä hoitoero kuin sellaisen, missä ero ei saavuta maagisena pidettyä **tilastollisen merkitsevyyden** rajaa $p < 0,05$. Näitä jälkimmäisiä tutkijat myös jättävät helpommin pöytälaatikkoonsa.

- Tutkimustulokset on julkaistu epäyhtenäisesti tai huonosti hyödynnettävässä muodossa: esim. elossaolokäyrät ("survival curves").
- Julkaistut tulokset eivät ole käyttökelpoisia esim. perusteettomasta poissuljennoista tai ristikkäisasetelman purkamisesta johtuen.

Tyyppin B tutkimuksissa on monia etuja verrattuna A:han ja C:hen

- Erilaiset tarkistukset ovat mahdollisia (tärkeää harhojen välttämiseksi).
- Ei menetetä informaatiota (saavutetaan maksimaalinen voima).
- Enemmän analyysimahdollisuuksia (esim. voidaan tehdä prognostisten tekijöiden suhteen adjustointeja tai osaryhmäanalyyssejä, jotka eivät tosin ole kovin suositeltavia).
- Paremmat mahdollisuudet antaa deskriptiivisiä tietoja aineistona olevista tutkimuksista.
- Eri tutkimusten tutkijoilla on mahdollisuus kontribuutioon.

Tyyppiä C olevia tutkimuksia käytetään yleensä seuraavissa tilanteissa:

- Voimanmenetyksellä B -tyypin tutkimuksiin verrattuna ei ole ratkaisevaa merkitystä. Yleensä asia on näin. Poikkeuksena eloonjäämistutkimukset, joissa on korkea kuolleisuus (**Buyse ja Ryan** 1987)
- Tulokset ovat helposti tiivistettävissä; yksinkertaiset lopputulosmittarit (esim. reduktio kokonaiskuolleisuudessa)
- Paljon pieniä tutkimuksia
- Tutkimukset ovat liian vanhoja, potilaskohtaisia tietoja ei ole enää käytettävissä tai niiden uudelleen kerääminen on liian kallista.

On arvioitu (**Cook et al** 1993), että vain **30 %**:ssa meta-analyysejä on hyödynnetty myös julkaisematonta tietoa. Julkaisemattoman tiedon poisjättäminen aiheuttaa meta-analyysien tuloksiin helposti ns. julkaisuharhan ("publication bias"), jota on käytetty voimakkaimpana argumenttina kritisoidessa meta-analyysejä (esim. **Oakes** 1993).

Lancetissa ja BMJ:ssä on viime aikoina julkaistu runsaasti meta-analyysiin kriittisesti suhtautuvia artikkeleita (**Eggert** 1997a-d, **Smith et al.** 1997), osassa näistä on ehdotettu testejä julkaisuharhan toteamiseksi ja menetelmiä sen korjaamiseksi. Julkaisemattomien tutkimusten tuloksia on viime vuosina pyritty enenevässä määrin tallentamaan myös erityisiin tutkimusrekistereihin, mm. lääketeollisuus noudattaa yleisesti tätä käytäntöä. Vaikkakin suositeltavaa, niin on erittäin vaikeaa hyödyntää julkaisemattomien tutkimuksien tuloksia meta-analyysissä (esim. **Yusuf** ym. 1985)

Meta-analyysi vastaan suuri lääkekoe

Meta-analyysien antamien tulosten oikeellisuutta on viime vuosina, etenkin USA:ssa, pyritty tarkistamaan suorittamalla lääkekokeita ("mega trials") kymmenien tuhansien suuruisilla tutkimusaineistoilla. Ristiriitaisiakin tuloksia on todettu. Esimerkiksi magnesiumhoito akuutissa sydäninfarktissa osoittautui lumehoitoa keskimäärin 55 % tehokkaammaksi meta-analyysissä (**Teo** ym. 1991), mutta ei suuressa ISIS-4 tutkimuksessa (**ISIS-4 Collaborative group** 1995). Mahdolliseksi selityksiksi tälle ristiriidalle on ehdotettu mm. potilaan saamia muita lääkehoitoja ja eroja magnesiumhoidon ajoituksessa (**Borzak ja Ridker** 1995).

Meta-analyysien tulokset saattavat vaihdella huomattavastikin riippuen käytetystä tilastollisesta analysointitavasta, jollaisia ovat mm. **satunnaisten vaikutusten** ("random effects") ja **kiinteiden vaikutusten** ("fixed effects") mallit. Muun muassa Thompson (1993) on tarkastellut esimerkein havainnollistettuna näitä ristiriitaisuuksia. Satunnaisten vaikutusten mallissa meta-analyyseissä käytetyt tutkimukset ajatellaan otokseksi kaikkien vastaavanlaisten tutkimusten perusjoukosta tuoden siten otantavaihtelun mukaan kuvaan.

Meta-analysejä voidaan suorittaa myös käyttäen ns. Bayesiläistä lähestymistapaa ("Bayesian approach"), joka on eräällä tavalla yleistys satunnaisten vaikutusten mallille. Bayesilaiset menetelmät saattavat olla hyödyllisiä etenkin tutkimuskohtaisia kovariaatteja käsiteltäessä. Ongelmana Bayesilaisessa lähestymistavassa on aina se, että tarvitaan tietoa a priori jakaumista, ne ovat laskennallisesti hankalia ja ohjelmia on niukasta saatavilla.

Viitteet

Begg CB, and Berlin JA. Publication bias: a problem in interpreting medical data J Roy Statist Soc A 1988; 151: 419-63.

Borzac, and Ridker. Ann Intern Med 1995; 123: 873-7.

Buyse M & Ryan LM. Issues of efficiency in combining proportions of deaths from several clinical trials. Stat in Med 1987; 6: 565-576.

Cook DJ, Gyatt GH, Ryan G, Clifton J, Buckingham L, Willan A et al. Should unpublished data be included in meta-analyses? Current convictions and controversies. JAMA 1993; 269: 2749-52.

Egger M, Smith GD.

- Meta-analysis: bias in location and selection of studies. BMJ 1997; 315: 61-66.
- Meta-analysis: potentials and promise. BMJ 1997; 315: 1371-4.

Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. BMJ 1997; 315:1533-7.

Egger M, Smith GD, Schneider M, and Minder CE. Bias in meta-analysis detected by simple graphical test. BMJ 1997; 315: 629-34.

Egger M, Schneider M, and Smith GD. Spurious precision? Meta-analysis of observational studies. BMJ 1997; 316: 140-144.

Fleiss JL. The statistical basis of meta-analysis. Statistical Methods in Medical Research 1993; 2: 121-45.

ISIS-4 (Fourth International Study of Infarct Survival Collaborative Group) ISIS-4, A randomized factorial trial assessing early oral Captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. Lancet 1995; 345: 669-85.

Oakes M. The logic and role of meta-analysis in clinical research. Statistical Methods in Medical Research 1993; 2: 147-60.

Smith GD, Egger M, and Phillips AN. Meta-analysis: beyond the grand mean? BMJ 1997; 315:1610-4.

Teo KK, Yusuf S, Collins R, Held PH, Peto R. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomized trials. BMJ 1991; 303:1499-503.

Thompson SG. Controversies in meta-analysis: the case of the trial of serum cholesterol reduction. *Statistical Methods in Medical Research* 1993; 2: 173-192.

Yusuf S, Collins R Peto R et al. Intravenous and intra-coronary fibrinolytic therapy in acute myocardial infarction. Overview of results on mortality, re-infarction and side-effects from 33 randomized controlled trials. *Eur. Heart J* 1985; 6: 556- 585.

Aineistokoko ja voima - analyysi

Johdanto

Aineisto - eli otoskoon arviointi ("sample size determination") ja tutkimuksen **voima - analyysi** ("power analysis") ovat tilastollisen tutkimuksen suunnittelussa keskeisimpiä kysymyksiä. Ne riippuvat monista eri tekijöistä. Näitä ovat mm. vertailutaso, ryhmien välillä oletettu kliininen ero, so. **vaikutuksen suuruus (Δ)** ("effect size") lopputulosmuuttujassa ("outcome measure") ja sen mittaamistapa, tutkijan sopivaksi katsoma **tilastollinen merkitsevyytaso ja testin voima**.

Tilastollinen merkitsevyytaso "statistical significance level" on tilastolliseen päättelyyn liittyvä todennäköisyys (α - virhe, I -lajin virhe), joka ilmoittaa kuinka suuri erehtymisriski nollahypoteesin virheelliseen hylkäämiseen halutaan sallia. Tilastollinen merkitsevyys on välttämätön edellytys kliiniselle merkittävyydelle ("importance"), mutta se ei kerro mitään todellisesta vaikutuksen suuruudesta.

Voima (1- β), "power" Tutkimuksen (testin) voima ilmoittaa millä todennäköisyydellä käytetyllä aineistokoolla ja tilastollisella merkitsevyytasolla voidaan todeta tietyn suuruinen vaikutus (efekti) käytetyssä lopputulosmuuttujassa, mikäli vaikutus on olemassa, ts. testin voima ilmoittaa todennäköisyyden hylätä nollahypoteesi silloin kun vaihtoehtoinen hypoteesi on tosi.

Suurilla aineistoilla kyetään näyttämään toteen pieniäkin eroja tilastollisesti merkitseviksi, vaikka niillä ei ehkä olisi olennaista kliinistä merkitystä, koska aineistokoon kasvaessa testattavien suureiden keskivirheet (ks. Luku "Aineiston kerääminen, tallentaminen ja käsittely") pienenevät, eli tarkkuus paranee. Paljon suurempi ongelma on kuitenkin liian pienten tutkimusaineistojen käyttö. Tästä syystä huolelliset aineistokokolaskelmat tulisikin sisällyttää jokaiseen hyvin tehtyyn tutkimussuunnitelmaan.

Erityisesti ns. "negatiivissa" satunnaistetuissa kliinisissä kokeissa β -virheen merkitys korostuu. Kun todetaan, että hoitojen välille ei saatu tilastollisesti merkitsevää eroa, herää automaattisesti kysymys: kuinka suuri ero tärkeimmässä lopputulosmuuttujassa oli mahdollista todeta käytetyllä aineistokoolla, eli mikä oli tutkimuksella voima.

Kirjallisuudessa on runsaasti heikkovoimaisia tutkimuksia, joissa eroa vertailtavien hoitojen välille ei ole saatu ja pöytälaatikkoonkin näitä tutkimuksia on takavuosina jäänyt aiheuttaen mm. meta-analyysein julkaisuharhaa. Nykyisin tilanne on toisin; kaikki oleelliset tiedot tutkimuksista, joissa lääketeollisuus on mukana, tallennetaan isoihin maailmanlaajuisiin tietokantoihin (esim. Cochrane-tietokanta)

Freiman et al. (1978) on suorittanut arvioinnin **71 julkaistusta** kahta eri hoitoa vertailevista satunnaistetuista kontrolloiduista tutkimuksista, joissa kaikissa oli päädytty toteamukseen: "Kahden hoidon vertailu ei saavuttanut tilastollista merkitsevyyttä ($P > 0,05$). Kaikissa päälopputulokset oli binomiaalinen, kaksiarvoinen. Tutkimukset arvioitiin uudelleen siten, että olivatko tutkijat käyttäneet riittävän suurta aineistokokoa, jotta he olisivat voineet todeta todennäköisyydellä ($P > 0,90$) 25 % ja 50 % hoitovaikutuksen. Todettiin, että **67:llä (94 %)** tutkimuksella oli **yli 10 %:n** riski olla toteamatta **25 %:n** hoitoveroa ja **50:llä (70 %)** tutkimusta ei olisi todennut edes **50 %:n** hoitoveroa, jota yleensä pidetään jo suurena erona satunnaistetuissa tutkimuksissa. Kun todellista hoitoveroa arvioitiin 90 %:n luottamusvälien (tarkempi käsittely myöhemmin tässä

monisteessa) avulla, niin todettiin, että 57:ssä (80 %) tutkimuksessa 25 %:n hoitoero olisi ollut mahdollinen ja 34:ssä (48 %) tutkimuksessa jopa 50 % olisi voinut jäädä toteamatta.

Esimerkkinä heikkovoimaisesta negatiivisesta tutkimuksesta voidaan mainita BMJ:ssä julkaistu verenpaine tutkimus (Cappuccio et al.1985), missä dietaarista magnesiumhoitoa verrattiin lumehoittoon lievää tai keskivaikeaa essentiellää hypertensiota potevilla henkilöillä. Tutkimus oli satunnaistettu kahden periodin ristikkäistutkimus. Tutkimuksessa päädyttiin erittäin varmaan johtopäätökseen: **"Tulokset eivät anna mitään evidenssiä siitä, että oraalilla magnesiumhoidolla olisi merkitystä verenpaineen säätelyssä"**. Ottamatta kantaa tuloksen oikeellisuuteen voidaan todeta, että tutkimuksessa oli kuitenkin vain 17 potilasta. Tutkijat ilmoittivat, että aineistokoko oli arvioitu tavoitellen systolisen/diastolisen verenpaineen muutosta 8/6 mm Hg aktiivihoidon ja lumehoidon välillä, mutta he eivät ilmoittaneet standardoitua hoitoeroa, joka on erotusmuuttujan "verenpaine aktiivihoidolla -verenpaine lumehoidolla" keskiarvo jaettuna sen keskihajonnalla. Käyttäen tutkijoiden ilmoittamaa 90 %:n voimaa ja kaksisuuntaisen testin merkitsevyytensä 5 % voidaan helposti laskea, esim. ohjelmalla nQuery, että tutkijat tavoittelivat standardoitua hoitoeroa 0,84. Voidaan kysyä perustellusti: "Eikö näin suuri ero ole kyseiseen tutkimukseen epärealistisen suuri?" Tutkimus on siten heikkovoimainen kliinisesti kohtuullisille 0,5 – 0,6 suuruusluokkaa oleville hoitoeroille. Vastaavanlaisia liian heikkovoimaisia tutkimuksia on tieteellisessä kirjallisuudessa runsaasti.

Aineistokoko-ongelmaan ei yleensä ole olemassa yksinkertaista ratkaisua. Aineistokoko kasvattamalla saadaan tunnuslukujen arvioita tarkennettua, mutta kasvattaminen lisää kuitenkin tutkimuskustannuksia, jotka monissa kliinisissä tutkimuksissa per tutkittava ovat suuria. Tämän takia tutkija joutuu yleensä tasapainoilemaan toisaalta arvioitavien suureiden tarkkuuden, toisaalta määrityskustannusten välillä. Myös tutkimukseen käytettävissä oleva aika rajaa käytännössä tutkimusaineiston keräämismahdollisuuksia. Lisäksi tilannetta vielä yleensä komplisoi se, että yhden tutkittavan suureen asemesta ollaan kiinnostuneita samanaikaisesti monista eri suureista. Tällöin aineistokoko, joka on riittävä jonkin tietyn suureen suhteen, ei välttämättä ole riittävä enää jonkin toisen suureen suhteen. Tällaisissa tilanteissa voidaan menetellä siten, että lasketaan kunkin tärkeimmän lopputulosta mittaavan suureen osalta erikseen aineistokokovaatimukset ja valitaan suurin näin saatavista vaatimuksista. Aineistokokolaskelmat toimivat käytännössä kuitenkin vain yleisenä ohjenuorana tehtäessä päätöstä aineistokoosta. Muita asiaan vaikuttavia tekijöitä ovat potilaiden saatavuus sekä eettiset ja taloudelliset kysymykset.

Ellei ilmoiteta tutkimuksen voimaa, niin jää epäselväksi:

- a) sitä ei ole ennalta määritetty
- b) ei ole saavutettu aiottua aineistokoko
- c) aineistokoko on kasvatettu alun perin suunnitellusta paremman tilastollisen voiman saavuttamiseksi
- d) tulokset on raportoitu ennen kuin suunniteltu aineistokoko on saavutettu, esim. hyvien välianalyysien ("interim analysis") tulosten takia.

Mikäli näistä vaihtoehdoista a) tai b) on tosi ja tutkimuksessa on saatu negatiivinen tulos, herää epäily tutkimuksen voiman puutteesta. Luottamusväli auttaisi!

Mikäli a) tai c) tai d) ovat tosia ja on saatu positiivinen tulos, herää epäily että on suoritettu toistettuja **välialalyseja** ("interim analysis") ja tulokset raportoitu, kun toivottu merkitsevä ero on saavutettu. Tästä ilmiöstä käytetään nimitystä raportoinnin ajoittamisesta johtuva harha, ja sen vaikutuksesta α -virhe kasvaa.

Välialalyysi "interim analysis" on mikä tahansa etukäteen suunniteltu kliinisen tutkimuksen kuluessa suoritettu analyysi. Syyt suorittamiseen ovat joko eettisiä tai taloudellisia. Välialalyysien suorittaminen vaikuttaa tilastolliseen merkitsevyytasoon (α - virhe kasvaa).

Kliinisten tutkimusten aineistokoon arviointiin on olemassa hyviä taulukkokokoelmia, esim. Machinin ja Campbellin "Statistical Tables for the Design of Clinical Trials". Se soveltuu mm. mukana kannettavaksi esim. kliinisten tutkimusten suunnittelukokouksiin. Nykyisin on saatavissa myös hyviä ja helppokäyttöisiä Windows -ympäristössä toimivia tilastollisia erityisohjelmistoja, esim. **nQyery** ja **Egret Siz**. Näistä jälkimmäinen on suunnattu erityisesti epidemiologiisiin tutkimusasetelmiin. Myös joissakin tilastollisissa yleispaketeissa, kuten esim. NCSS:ssä on aineistokoon arviointimoduli.

Aineistokoon arvioinnista kannattaa kuitenkin pitää mielessä seuraava tosiseikka: eksakteilla hienoilla laskentakaavoilla ja ohjelmilla synnytetään helposti väärä illuusio aineistokoon arvioinnin tarkkuudesta, vaikka arviointi perustuu usein huomattavan karkeisiin lähtöoletuksiin esim. vaikutuksen suuruudesta.

Tilastokäsittelyn kannalta on oleellista minkälaisilla suureilla ja muuttujilla havaintoja tehdään ja mitataan. Muuttujat voidaan kahteen päätyyppiin:

- **epäjatkuvat** ("discrete") ja
- **jatkuvat** ("continuous")

Epäjatkuvat muuttujat voivat saada vain tiettyjä, määrättyjä arvoja, ja jatkuvat muuttujat voivat saada arvoalueensa sisältä mitä tahansa arvoja (mittaustarkkuuden puitteissa), esimerkiksi henkilön ikä.

Epäjatkuvat muuttujat voidaan edelleen jakaa:

- **luokkamuuttujiin** ("categorical") ja
- **numeerisiin muuttujiin** ("numerical"),

joista edellisistä esimerkkinä ABO - veriryhmäjärjestelmä, ja jälkimmäisestä lukumäärät ("counts").

Tyypillinen ongelma: Suunnitellaan satunnaistettua kliinistä hoitotutkimusta, missä vertaillaan perinteistä ja uutta hoitoa keskenään. Hoitovastetta mitataan joko **a) binomiaalisella asteikoilla**, esimerkiksi parantunut/ei parantunut tai **b) jatkuvalla asteikolla**, esimerkiksi verenpaineen lasku tai elossaoloajan pidentyminen.

Tavallisimpia laskentatilanteita ovat:

- kahden suhteellisen osuuden vertaaminen
- kahden keskiarvon vertaaminen
- kahden eloonjäämiskäyrän vertaaminen
- riskisuhteen testaaminen
 - tapaus-verrokkitutkimus
 - kohorttitutkimus.
- suhteellisten osuuksienjen vertailu parittaisissa asetelmissa
- kahden suhteellisen osuuden yhtä suuruuden testaaminen

Aineistokoon määrittämiseen liittyvät avainkäsitteet

Arvioitaessa aineistokoko tutkijan täytyy määrittellä seuraavat asiat:

- Tutkimuksen keskeisin **lopputulosuuttuja** ("main outcome measure") ja sen mittaluku.
- **Vaikutuksen suuruus** (Δ = "effect size").
- **Vertailutaso** (π_c = "reference level") π_c on arvioitavan suhteellisen osuuden oletettu suuruus tai vaikutus vertailuryhmässä tai altistuksen määrä tapaus-verrokki ja kohorttitutkimusten verrokkiryhmässä jne.)
- **Tilastollinen merkitsevyystaso** ("statistical significance level") (α = "väärän positiivisen tuloksen riski"). Suositus: $\alpha \leq 0,05$
- **Testin voima $1 - \beta$** , missä β = "väärän negatiivisen tuloksen riski". Suositus: $0,80 \leq 1 - \beta \leq 0,95$. Testin voima on varmuus todeta Δ :n suuruinen ero, jos se on olemassa.
- **Analysointimenetelmä** eron toteamiseksi

Aineistokoon arviointikaavan yleinen muoto on: $n = f(\alpha, 1 - \beta, \Delta, \pi_c)$, missä n on kuhunkin ryhmään tarvittava aineistokoko ja f on **parametreista** α , $1 - \beta$, Δ ja π_c riippuva funktio. Parametrit α , β , Δ ja n riippuvat toinen toisistaan siten, että kaikkien välillä on **käänteinen** riippuvuus; jos jokin niistä kasvaa, täytyy toisen pienentyä ja päinvastoin olettaen, että π_c pidetään kiinteänä.

Parametrien α , β , Δ ja n riippuvuus toinen toisistaan

	α	β	Δ	n
α		Käänteinen	Käänteinen	Käänteinen
β	Käänteinen		Käänteinen	Käänteinen
Δ	Käänteinen	Käänteinen		Käänteinen
n	Käänteinen	Käänteinen	Käänteinen	

Käänteinen tarkoittaa, että kun rivillä oleva parametri kasvaa niin sarakkeella oleva parametri pienenee ja päinvastoin

20

Esimerkiksi aineistokoko (n) kasvaa, kun α -tasoa tiukennetaan (pienennetään) pitäen kaksi muuta suuretta β ja Δ kiinteinä. Laskentakaavan perusteella voidaan kääntäen ratkaista myös mikä tahansa parametreista α , $1 - \beta$ tai Δ . Yksinkertaisimmissa tilanteissa ongelmille on olemassa algebrallinen ratkaisu. Usein kuitenkin joudutaan käyttämään iteratiivista numeerista ratkaisumenetelmää.

Iteraatio "Iteration" on toistettu matemaattinen ratkaisumenettely, missä edellisen vaiheen tulosta käytetään lähtökohtana seuraavan vaiheen tuloksen laskemiselle, esim. paljon käytetty Newton-Raphsonin numeerinen ratkaisumenetelmä.

Useimmissa aineistokoon arviointiohjelmissa (esim. nQuery) voidaan mikä tahansa parametreista n , α , $1 - \beta$ ja Δ ratkaista joustavasti toistensa funktioina, ja lisäksi kullekin parametrille voidaan vuorollaan antaa vaihteluväli, jolloin saadaan taulukko arviointipäätöksen tekemisen helpottamiseksi. Ohjelmassa nQuery on lisäksi mukana esitysgrafiikka ja mahdollisuus tuottaa protokollaa varten aineistokoon arviointilaskelma.

Yhden binomijakaumaa noudattavan suhteellisen osuuden testaaminen

Käyttötilanne:

Suunnitellaan tutkimusta, jonka tavoitteena on arvioida esim. jonkin taudin vallitsevuutta (π) tietyssä väestöryhmässä (perusjoukossa). Asetetaan kysymys: Kuinka suuri otos (n) (tutkittavien määrä) tarvitaan, jotta arvioitavan suureen virhe olisi korkeintaan $\pm e$.

Tämän ongelman ratkaiseminen edellyttää suhteellisen osuuden π luottamusvälin määrittelyä. Luottamusvälien tarkastelu tulee myöhemmin, mutta tässä yhteydessä riittää määritelmä:

Luottamusväli ”confidence interval” on otoksen perusteella laskettu väli, johon arvioitavana oleva tuntematon perusjoukon suhteellinen osuus sijoittuu luottamustason (tavallisesti 95 %) ilmoittamalla varmuudella.

Luottamusvälin pituus riippuu otoksen suuruudesta ja arvioinnin varmuudesta.

Binomijakauma ”Binomial distribution” kuvaa toistettujen satunnaisten kaksiarvoisten tapahtumien eri lopputulosten esiintymistodennäköisyyksiä. Se on siten epäjatkua jakauma.

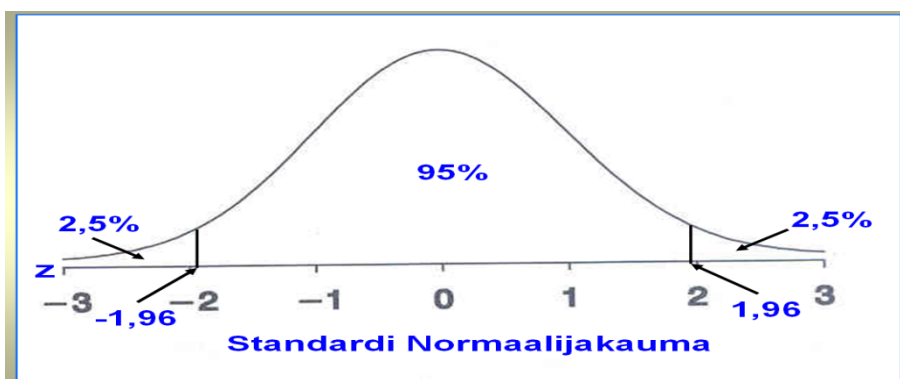
Normaalijakaumaan perustuva arvio binomiaaliselle suurelle saadaan kaavalla:

$$\pi \pm z_{1-\alpha/2} \cdot (\pi(1-\pi)/n)^{1/2}$$

missä $z_{1-\alpha/2}$ on Normaalijakauman prosenttipiste ja $100 \cdot (1-\alpha)$ ilmaisee luottamustason.

Tavallisimmin käytettyjä luottamustasoja ovat 90 %, 95 % tai 99 %, jolloin vastaavat z - arvot ovat **1,64**, **1,96** ja **2,58**.

Normaalijakauma ”Normal distribution”, ”Gaussian distribution”, ”bell curve” on eräs yleisimmistä tilastollisissa testeissä ja menetelmissä käytetyistä jatkuvista jakaumista. Monet parametrisista testeistä ja menetelmistä perustuvat normaalisuusoletukseen. Standardoidun normaalijakauman odotusarvo on 0 ja hajonta on 1. Jakaumakäyrän alle jäävästä pinta-alasta 95 % jää arvojen $z=-1,96$ ja $z=1,96$ väliin.



Standardoidun Normaalijakauman (keskiarvo=0, SD=1) prosenttipisteet saadaan esim. **Excelillä** valikosta ”insert” ► ’function’ valitsemalla kategoriaksi ’statistical’ ja sieltä funktioksi **NORMINV**, eli Normaalijakauman kertymäfunktion käänteisfunktion. Tasoa $\alpha=0,10$ vastaava z -arvo **1,64** saadaan sijoittamalla lokeroon ”Probability” arvo $1-\alpha/2 = 0,95$.

Function Arguments

NORMINV

Probability 0,95 = 0,95

Mean 0 = 0

Standard_dev 1 = 1

= 1,644853476

Returns the inverse of the normal cumulative distribution for the specified mean and standard deviation.

Probability is a probability corresponding to the normal distribution, a number between 0 and 1 inclusive.

Formula result = 1,644853476

[Help on this function](#) OK Cancel

Luottamusväli saadaan asettamalla $z_{1-\alpha/2} \cdot (\pi (1 - \pi) / n)^{1/2} = e$ ja ratkaisemalla yhtälö n :n suhteen, jolloin saadaan kaava:

$$n = \text{int} \left[(\pi (1 - \pi) \cdot (z_{1-\alpha/2} / e)^2 \right] \quad (1)$$

Funktio $\text{int}[\cdot]$ tarkoittaa pyöristystä ylöspäin lähimpään kokonaislukuun. Tämä kaava on likimääräinen ja pätee ainoastaan suurilla otoksilla ($n > 30$). Kaavan 1 mukainen arvio voidaan laskea mm. osoitteessa:

www.mv.helsinki.fi/home/sarna/stats/SuhtOsuusLuottamusväli.xls olevalla laskimella.

Esim.

Oletetaan, että halutaan arvioida aikuistyyppin diabeteksen vallitsevuutta jossain väestössä. Arvioinnin virheeksi (e) sallitaan $\pm 3\%$ kaksipuolista 95 % luottamusväliä käytettäessä. Oletetaan, että vallitsevuus (π) on korkeintaan 30 %.

Suhteellisen osuuden arviointi		nQuery Advisor - [POC0-1]	
n:n arviointi $1-\alpha$:n, π :n ja e :n perusteella		Confidence interval for proportion using normal a	
Luottamustaso $1-\alpha$	0,950		1
1 tai 2 suuntainen väli	2	Confidence level, $1-\alpha$	0,950
Ennakoarvio suhteellisesta osuudesta π	0,30	1 or 2 sided interval?	2
Sallittu virhe e	0,03	Expected proportion, π	0,300
Aineistokoko (n)	897	Distance from proportion to limit, ω	0,030
		n	897

Mikäli arvioitava suhteellinen osuus π on pieni / suuri (alle 0,05 / yli 0,95), niin binomiaalinen väli toimii huonosti. Tällöin kannattaa käyttää **Poisson- jakauman** perustuvaa väliä (ks. Luku Luottamusväliestimointi).

Poisson - jakauma "Poisson distribution" on harvinaisten kaksiarvoisten (kyllä /ei) tapahtumien tutkimisessa ja testaamisessa käytetty jakauma.

Huom.

Useimmissa tutkimuksissa halutaan arvioida monia suhteellisia osuuksia. Tällöin on vaikea valita kaavassa tarvittavaa yksittäistä lukua π . Koska kaava antaa maksimaalisen arvon n :lle silloin kun $\pi = 0,5$, sen antama arvo kannattaa tällöin ottaa aineistokooksi. Yllä olevassa esimerkissä saataisiin $n = 1068$.

Kahden binomijakaumaa noudattavan suhteellisen osuuden vertailu

Oletetaan, että kyseessä on kaksi vertailtavaa ryhmää ja hoidon tehoa arvioidaan **binomiaalisella asteikolla**, esimerkiksi parantunut / ei parantunut.

Kaksi toisistaan riippumatonta ryhmää

Käyttötilanne:

Suunnitellaan tutkimusta, jonka tavoitteena on vertailla esimerkiksi kahta eri lääkehoitoa saavaa, eri potilaista koostuvaa ryhmää toisiinsa. Lopputulosta arvioidaan suhteellisilla osuuksilla π_1 (esim. traditionaalinen hoito) ja π_2 (esim. uusi hoito).

Aineistokoon (n) arviointiin tässä tilanteessa on olemassa useita vaihtoehtoisia kaavoja, jotka eroavat toisistaan ainoastaan suhteellisten osuuksien varianssien erilaisten arviointitapojen osalta. Useimmissa käytännön tilanteissa n voidaan laskea kaavalla:

$$n = \left[z_{1-\alpha} \sqrt{2\bar{\pi}(1-\bar{\pi})} + z_{1-\beta} \sqrt{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)} \right]^2 / \Delta^2 \quad (2)$$

missä n = aineistokoko per ryhmä, $\Delta = \pi_1 - \pi_2$ on suhteellisten osuuksien erotus, $\bar{\pi} = (\pi_1 + \pi_2)/2$ niiden keskiarvo, α on yksi - tai kaksipuolinen ja β on aina yksipuolinen todennäköisyys. Kokonaisaineistokooksi tulee siten $2 \cdot n$ koska kaava edellyttää, että halutaan yhtä paljon potilaita kumpaankin ryhmään.

Esim.

Polviartroosiin liittyvien kipujen hoitokokeen suunnittelu (Smith et al., 1983).
 Hoitona käytetään ihon läpi annettavaa sähköstimulaatiota (TNS -hoito).
 Aikaisemman kokemuksen perusteella tiedetään, että lumehoidolla saadaan noin **25 %**:n vaste neljän viikon seurannassa. Tutkija arvioi, että TNS -hoidolla voitaisiin saavuttaa **65 %**:n vaste vastaavana ajanjaksona, eli $\Delta = 0,40$. Kuinka monta potilasta kumpaankin hoitoryhmään tarvitaan, jos $\alpha = 0,05$ (yksipuolinen) ja $1 - \beta = 0,90$? Lopputulosmuuttujana oli hoitovaste neljän viikon seurannassa siten, että lopputulosvaihtoehtoina oli: parantui tai ei parantunut.

Osoitteessa: www.mv.helsinki.fi/home/sarna/stats/KahdenSuhtOsuudenVertailu.xls olevalla laskimella voidaan laskea kaavaan 2 perustuva arvio.

Kahden Suht. osuuden vertailu	
$n_1:n$ ja $n_2:n$ arviointi $\alpha:n$, $\pi_1:n$, $\pi_2:n$, $(1-\beta):n$ ja $r:n$ perusteella	
Merkitsevyystaso α	0,050
1 tai 2 suuntainen testi	1
Ryhmän 1 suht.osuus π_1	0,25
Ryhmän 2 suht.osuus π_2	0,65
Efektin suuruus $\Delta = \pi_2 - \pi_1$	0,40
Ristitulosuhde(OR) $\pi_2(1-\pi_1)/[\pi_1(1-\pi_2)]$	5,57
Voima (%) $100(1 - \beta)$	90 %
Allokaatiosuhde $r = n_1 / n_2$	1
Aineistokoko (n_1)	25
Aineistokoko (n_2)	25

Ohjelmalla nQuery saadaan sama tulos:

nQuery Advisor - [PTT0-2]	
Two group χ^2 test of equal proportions (odds ratio = 1) (equal r	
	1
Test significance level, α	0,050
1 or 2 sided test?	1
Group 1 proportion, π_1	0,250
Group 2 proportion, π_2	0,650
Odds ratio, $\psi = \pi_2 (1 - \pi_1) / [\pi_1 (1 - \pi_2)]$	5,571
Power (%)	90
n per group	25

Huom. Kaava (2) perustuu olettamukseen, että lopputuloksen arvioinnissa käytetään χ^2 -testiä ilman jatkuvuuskorjausta. Jatkuvuuskorjauksen käyttö antaa yleensä liian konservatiivisen arvion.

χ^2 -testi on yksi yleisimmin käytetyistä ristiintaulukkotesteistä. Sen tavallisimmat käyttöindikaatiot ovat frekvenssilukujen heterogeenisuuden tai suhteellisten osuuksien trendin (monotonisuuden) testaaminen. Sitä käytetään myös yhteensopivuustestinä, kun verrataan jotain havaittua jakaumaa johonkin teoreettiseen jakaumaan.

Jos halutaan eri määrä potilaita vertailtaviin ryhmiin allokaatiosuhteessa $r = n_1 / n_2$, niin kaava (2) modifioituu muotoon:

$$n = \left[z_{1-\alpha} \sqrt{(r+1)\bar{\pi}(1-\bar{\pi})} + z_{1-\beta} \sqrt{r \cdot \pi_1(1-\pi_1) + \pi_2(1-\pi_2)} \right]^2 / (r \cdot \Delta^2)$$

Laskimella voidaan laskea myös käänteisesti voima $(1-\beta)$ suureiden α , Δ ja n funktiona

Esim.

Kahden lääkkeen vertailu mahahaavan hoidossa (Familiari et al 1981).

Hoidon kesto oli neljäkymmentä vuorokautta. Vertailtavat lääkehoidot: trithioziini (hoito1), pirenzepiini (hoito 2). Lopputuloksena todettiin, että hoidolla 1 18 / 31 (58,1 %) ja hoidolla 2 23 / 30 (76,7 %) parani. Ero ei ole tilastollisesti merkitsevä riskitasolla $\alpha = 0,05$ (kaksisuuntainen testi).

Jos edellä olevan tuloksen perusteella tehdään johtopäätös "lääkkeiden välillä ei ole tilastollisesti merkitsevää eroa", niin kuinka suuri on todennäköisyys, että tutkimuksessa saatu 18,6 %:n todellinen ero jäisi käytetyllä aineistokoolla ja 5 % merkitsevyystasolla toteamatta kaksisuuntaisella testillä?

Laskimella www.mv.helsinki.fi/home/sarna/stats/KahdenSuhtOsuudenVertailu.xls voimaksi saadaan 34 %.

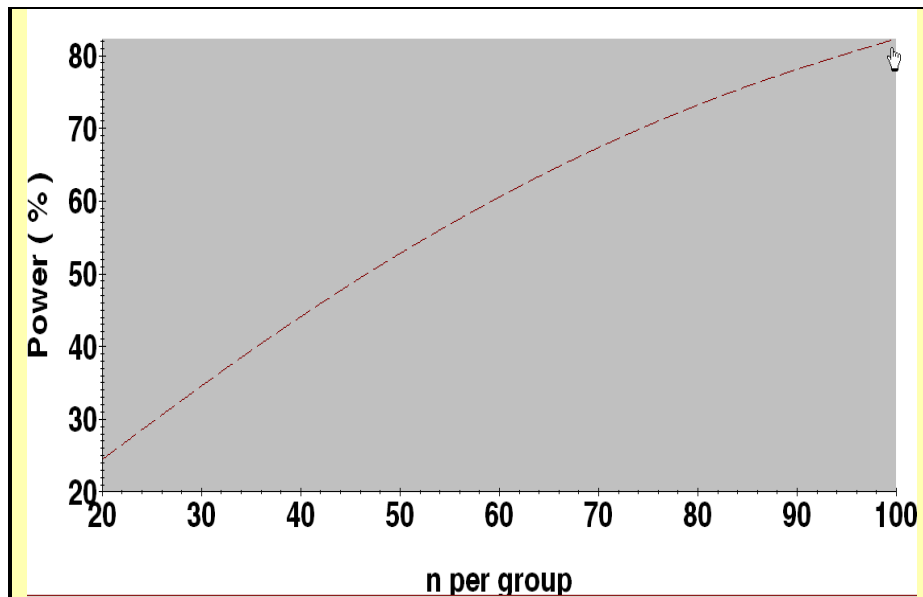
Kahden Suht. osuuden vertailu (1-β):n arviointi $n_1:n, n_2:n, \alpha:n, \pi_1:n$ ja $\pi_2:n$ perusteella	
Merkitsevyystaso α	0,050
1 tai 2 suuntainen testi	2
Ryhmän 1 suht.osuus π_1	0,581
Ryhmän 2 suht.osuus π_2	0,767
Efektin suuruus $\Delta = \pi_2 - \pi_1$	0,19
Ristitulosuhde(OR) $\pi_2(1-\pi_1)/[\pi_1(1-\pi_2)]$	2,37
Voima (%) $100(1 - \beta)$	33,9 %
Allokaatiosuhde $r = n_1/n_2$	1,0
Aineistokoko (n_1)	31
Aineistokoko (n_2)	30

Likimain sama tulos saadaan ohjelmalla **nQuery**. Ero johtuneen selittyneen pyöristysvirheillä.

nQuery Advisor - [PTTOU-1]		
Two group χ^2 test of equal proportions (odds ratio = 1) (unequal n's)		
	1	2
Test significance level, α	0,050	
1 or 2 sided test?	2	
Group 1 proportion, π_1	0,581	
Group 2 proportion, π_2	0,767	
Odds ratio, $\psi = \pi_2 (1 - \pi_1) / [\pi_1 (1 - \pi_2)]$	2,378	
Power (%)	33	
n_1	31	
n_2	30	
Ratio: n_2 / n_1	0,968	
$N = n_1 + n_2$	61	

Tulkinta: Koska voima on vain 34 % , niin aineistokoko ei ole riittävä nollahypoteesin (lääkkeiden välillä ei ole eroa) hyväksymiseen. On 66 %:n

mahdollisuus, ettei kyseistä eroa todeta. Tarvittaisiin noin 100 henkilöä per ryhmä 80 %:n voiman saavuttamiseksi.



Kaavasta (2) ei voi algebrallisesti laskea Δ :aa suureiden $(1-\beta)$, α , n_1 ja n_2 funktiona, mutta kokeilemalla voidaan helposti todeta, että käytetyllä aineistokoolla voitaisiin 80 % voimalla ja 5 % merkitsevyystasolla todeta 31 % ero lopputulosmuuttujassa ryhmien välillä, mikäli ero on olemassa. Kun ohjelmassa määritellään pienemmäksi suhteelliseksi osuudeksi 0,581, niin suurempi suhteellinen osuus 0,881 tuottaisi 80 %:n voiman

Kahden Suht. osuuden vertailu (1-β):n arviointi $n_1:n$, $n_2:n$, $\alpha:n$, $\pi_1:n$ ja $\pi_2:n$ perusteella	
Merkitsevyystaso α	0,050
1 tai 2 suuntainen testi	2
Ryhmän 1 suht.osuus π_1	0,581
Ryhmän 2 suht.osuus π_2	0,891
Efektin suuruus $\Delta = \pi_2 - \pi_1$	0,31
Ristitulosuhde(OR) $\pi_2(1-\pi_1)/[\pi_1(1-\pi_2)]$	5,90
Voima (%) $100(1 - \beta)$	80,0 %
Allokaatiosuhde $r = n_1/n_2$	1,0
Aineistokoko (n_1)	31
Aineistokoko (n_2)	30

Yksi otos

Mikäli π_1 tunnetaan, esimerkiksi aikaisempien vastaavien tutkimusten perusteella, niin kaavan (2) antama aineistokoko n voidaan tällöin karkeasti ottaen puolittaa edellyttäen, että π_1 ja π_2 ovat suunnilleen yhtä suuria. Tarvittavaksi kokonaisaineistokooksi tulee siten vain 25 % ($100 \cdot (n/2) / 2n = 25\%$) verrattuna tilanteeseen, missä molempien ryhmien varianssit pitää estimoida. Esimerkiksi jos TNS -hoitoesimerkissä on vain yksi ryhmä ja vertailu suoritetaan hypoteettiseen arvoon 0,25., niin aineistokooksi tulee 11.

nQuery Advisor - [POT0-3]

File Edit View Options Assistants

One group χ^2 test that proportion equals user sp

	1
Test significance level, α	0,050
1 or 2 sided test?	1
Null hypothesis proportion, π_0	0,250
Alternative proportion, π_A	0,650
Power (%)	90
n	11

Arviointi ristitulosuhteen avulla

Etenkin tapaus-verrokkiasetelmissa tarkastellaan sairastumisriskiä kuvaavien todennäköisyyksien π asemesta suhdetta $\pi/(1-\pi)$. Tätä suhdetta nimitetään tapahtumismahdollisuudeksi ("odds").

Mahdollisuus "odds" on suhdeluku, joka kuvastaa kuinka monikertainen jonkin tapahtuman toteutumismahdollisuus on verrattuna siihen, ettei tapahtumaa tule. Esimerkiksi odds=6 (6:1) voisi tarkoittaa, että 6 henkilöä sairastuu ja yksi ei. Suuretta käytetään tutkimuksissa usein riskiä kuvaavana. Sen tulkinta on kuitenkin paljon ongelmallisempaa kuin riskin. Yli yhden olevat luvut on helppo hahmottaa, mutta alle yhden olevien tulkinnaissa on ongelmia. Esim. odds = 0,20 = 1:5, riski = 0,17 (yksi kuudesta), odds=0,67, riski = 0,40 tai odds = 4, riski = 0,8. Harvinaisilla tapahtumilla riski ja mahdollisuus ovat lähellä toisiaan.

Näiden "odds":ien suhdetta nimitetään ristitulosuhteeksi tai altistesuhteeksi ("odds ratio") tilanteesta riippuen ja sillä mitataan eri ryhmien välistä suhteellista yhteyttä esim. taudin ja altisteen välillä.

Ristitulosuhde (OR), Altistesuhde "odds ratio" on kahden kaksiarvoisen muuttujan välinen riippuvuuden mitta. OR=1 merkitsee, ettei ole riippuvuutta. Mittaa käytetään etenkin tapaus – verrokki - tutkimusasetelmassa taudin ja altisteen välistä yhteyttä arvioitaessa. OR = ("odds" altistuneille) / ("odds" altistumattomille).

Joissakin tilanteissa tutkijan on helpompi formuloida aineistokoko - ongelma suhteellisen osuuden π_1 ja **ristitulosuhteen** $OR = (\pi_2 \cdot (1 - \pi_1)) / (\pi_1 \cdot (1 - \pi_2))$ avulla Δ :n asemesta. Tällöin π_2 voidaan laskea yhtälöstä: $\pi_2 = (OR \cdot \pi_1) / [OR \cdot \pi_1 + (1 - \pi_1)]$ ja sijoittaa kaavaan (2).

Polviartroosia käsittelevässä esimerkissä saadaan OR $(0,65 \cdot 0,75) / (0,25 \cdot 0,35) = 5,57$ ja siten voitaisiin kysyä aineistokokoja, joka antaisi vertailtavien ryhmien välille 5,57-kertaisen ristitulosuhteen hoitoeron 0,40 sijasta.

Aineistokoon (n), vertailuryhmän suhteellisen osuuden (π_1) ja ristitulosuhteen (OR) perusteella voidaan arvioida OR:n luotettavuutta sen luottamusvälin perusteella (luottamusvälien laskentakaavat esitetään myöhemmin).

Esim.

Ennenaikaisten synnytysten riski raskaana olevilla äideillä (nQuery 6.0 manuaali, s.15 -21.) Tavoitteena oli ehkäistä ennenaikaisia synnytysten määrää tietyn koulutusohjelman avulla. Tutkittavia ryhmiä oli kaksi: ryhmä 2 sai koulutusta ja ryhmä 1 ei. Tutkijat olivat valmiit ottamaan kumpaankin ryhmään 500 äitiä ja halusivat pudottaa koulutuksen avulla ennenaikaisten synnytysten $OR_{2/1:n}$ 0,5:een (50 %:n suojaava vaikutus). Aikaisempien tutkimusten perusteella he arvioivat, että ryhmässä 1 ennenaikaisia synnytyksiä voisi olla 8 %. Tutkijat kysyivät: kuinka suuri on $OR:n$ 95 %:n luottamusväli? Ohjelmalla nQuery saadaan väliksi **(0,290, 0,862)**. Todetaan, että luku 1 ei sisälly väliin, mutta väli on melko laava.

Confidence interval for ln(odds ratio) (large equal n's)	
	1
Confidence level, $1-\alpha$	0,950
1 or 2 sided interval?	2
Group 1 proportion, π_1	0,080
Group 2 proportion, π_2	0,042
Odds ratio, $\psi = \pi_2 (1 - \pi_1) / [\pi_1 (1 - \pi_2)]$	0,500
$B = \ln(\psi)$	-0,693
Distance from B to limit, ω	0,545
n per group	500
$\psi_L = \exp(B - \omega)$	0,290
$\psi_U = \exp(B + \omega)$	0,862

Parittaiset aineistot

Kaavaa (2) voidaan soveltaa myös parittaisten (**kaltaistettujen**) aineistojen tapauksessa seuraavasti. Oletetaan, että tutkimusaineisto muodostuu havaintopareista: tapaus - verrokki tai sama henkilö kahtena eri hoitajaksona (ristikkäistutkimus) ja lopputulos on kaksiarvoinen (kyllä / ei). Oletetaan, että m:stä havaintoparista on saatu lopputulokseksi seuraava frekvenssitaulukko:

a) Tapaus-verrokkiasetus				b) Ristikkäistutkimusasetus			
Altistus		Verrokki (C)		Positiivinen vaste		Hoitajakso 1 (C)	
		Kyllä	Ei			Kyllä	Ei
Tapaus (E)	Kyllä	t	r	Hoitajakso 2 (E)	Kyllä	t	r
	Ei	s	u		Ei	s	u

Lopputuloksen "kyllä" suhteelliset osuudet ryhmässä C ja E: $\pi_C = (t + s) / m$, $\pi_E = (t + r) / m$

Parittainen testaus perustuu diskordantteihin (lopputulos erilainen) havaintopareihin, eli frekvensseihin r ja s. Testisuure (**McNemarin testi**) on:

$$\chi^2 = |r - s|^2 / (r + s)$$

Nollahypoteesin mukaisesti diskordantteja pareja on yhtä paljon ($r = s$), eli niiden suhteellinen osuus $\Omega_0 = r / (r + s) = s / (r + s) = 0,5$. Vaihtoehdoisen hypoteesin mukaista

suhteellista osuutta Ω_1 on käytännössä vaikea ennalta määrittellä. Jos voidaan karkeasti olettaa, että kaltaistustekijät eivät vaikuta "kyllä" - vastauksen todennäköisyyteen (esim. myönteinen hoitovaste), niin Ω_1 saadaan seuraavasta kaavasta:

$$\Omega_1 = \pi_C \cdot (1 - \pi_E) / (\pi_C \cdot (1 - \pi_E) + \pi_E \cdot (1 - \pi_C)) \quad (4)$$

Sijoittamalla kaavaan (2) π_C :n ja π_E :n paikalle Ω_0 ja Ω_1 ja jakamalla näin saatu luku n kahdella saadaan diskordanttien parien odotettu määrä k . Yllä esitetyin oletuksin tarvittava aineistokoko m (= parien määrä) saadaan kaavasta:

$$m = k / (\pi_C \cdot (1 - \pi_E) + \pi_E \cdot (1 - \pi_C)) \quad (5)$$

Esim.

Estrogeenialtistus ja kohtusyöpä (Brown 1976). Tutkimuksessa oli mukana 63 "tapaus -verrokki" -paria ja lopputulokseksi saatiin nelikenttä:

	<i>Altistunut</i>	<i>Verrokki (C)</i>		<i>Yhteensä</i>
		<i>Kyllä</i>	<i>Ei</i>	
<i>Tapaus (E)</i>	<i>Kyllä</i>	27	29	56
	<i>Ei</i>	3	4	7
<i>Yhteensä</i>		30	33	63

Jos oletetaan, että vasta suunniteltaisiin tätä tutkimusta ja kysyttäisiin: Kuinka paljon pareja tarvitaan, jos $\alpha = 0,01$, $1 - \beta = 0,90$, $\pi_C = 30 / 63 = 0,4762$ ja $\pi_E = 56 / 63 = 0,8889$, $\pi_E - \pi_C = 0,4127$.

Ratkaisu: Kaavasta (4) saadaan ensin $\Omega_1 = 0,10$. Sijoittamalla kaavaan (2) arvot: $f(\alpha, \beta) = 14,9$, $\pi_C = \Omega_0 = 0,50$, $\pi_E = \Omega_1 = 0,10$ ja $\Delta = 0,40$ saadaan $n = 31,7$. Diskordanttien parien odotettu määrä on siten puolet tästä määrästä, eli $k = 16$. Sijoittamalla kaavaan (5) saadaan tutkimukseen tarvittavien parien määräksi: $m = 16 / (0,4762 \cdot 0,1111 + 0,8889 \cdot 0,5238) = 31$. (nQuery: $m = 35$, ero johtuu jatkuvuuskorjauksesta)

McNemar's test (χ^2) of equality of paired proportions	
	1
Test significance level, α	0,010
1 or 2 sided test?	2
Difference in proportions, $\delta = \pi_1 - \pi_2 $	0,413
Proportion of discordant pairs, $\eta = \pi_{10} + \pi_{01}$	0,508
Power (%)	90
n (number of pairs)	35

Huom.

Jos edellä olevasta nelikentästä lasketaan **McNemarin testi**, niin saadaan $\chi^2 = 21,1$ ja $P < 0,0001$.

Kahden suhteellisen osuuden ekvivalenttisuuden testaaminen

Sellaisten kliinisten tutkimusten määrä, joissa pyritään ns. bioekvivalenttisuuden toteennäyttämiseen, esimerkiksi kahden eri hoitomuodon välillä, on viime vuosina lisääntynyt voimakkaasti. Tavallisin tilanne on sellainen, että on olemassa jokin perinteinen hoitomuoto C, jonka teho π_C tiedetään hyväksi, mutta jolla on paljon sivuvaikutuksia. Tutkimuksessa kokeillaan uutta hoitomuotoa E, jolla on vähemmän sivuvaikutuksia, mutta tehon π_E ei oleteta aivan yltävän klassisen hoidon tasolle. Hoidot katsotaan **bioekvivalenteiksi**, mikäli tehojen ero ei ylitä tutkijan ennalta määrittelemää pientä määrää ϵ (> 0); $\pi_C - \pi_E < \epsilon$. Tällaisessa tilanteessa tarvittava aineistokoko n yhtä hoitomuotoa kohden saadaan kaavasta:

$$n = (z_{1-\alpha/2} + z_{1-\beta})^2 \cdot (\pi_C \cdot (1 - \pi_C) + \pi_E \cdot (1 - \pi_E)) / (\epsilon - (\pi_C - \pi_E))^2 \quad (6)$$

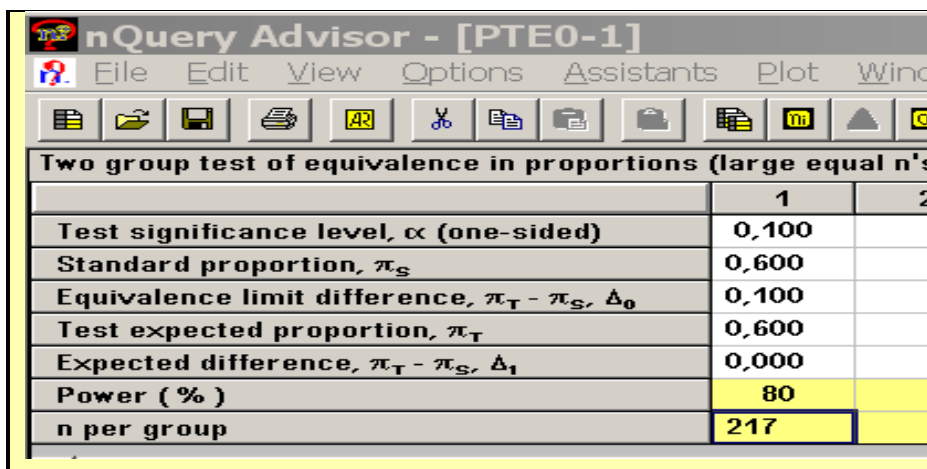
Huom.

Tavallisimmin kaavaa (6) käytettäessä asetetaan $\pi_E = \pi_C = \pi$ ja valitaan $\alpha=0,10$ (huom. yksipuolinen testi), $\beta = 0,20$ (eli voima=80 %) ja ϵ :lle annetaan arvoja väliltä (0,05, 0,20).

Esim.

Vaiheen I rintasyövän hoidot (Machin & Campbell, s.36): Mastectomia (C), Kyhmyyn poisto (E). Seuranta -aika 5 vuotta. Oletetaan, että $\pi_C = 0,60$ ja että hoidot katsotaan bioekvivalenteiksi, jos π_E on vähintään **0,50**, eli $\epsilon = 0,10$. Jos lisäksi vaaditaan, että $\alpha = 0,10$ (yksipuolinen testi) ja $1 - \beta = 0,80$, niin kuinka paljon potilaita kumpaankin ryhmään tarvitaan?

Ratkaisu: Asetetaan $\pi_E = \pi_C = 0,60$. Taulukon 1 perusteella $f(\alpha, \beta) = 4,5$. Sijoittamalla nämä arvot kaavaan (6) saadaan $n = 217$ (nQuery), eli näin monta potilasta tarvitaan kumpaankin hoitoryhmään.



Two group test of equivalence in proportions (large equal n's)		
	1	2
Test significance level, α (one-sided)	0,100	
Standard proportion, π_S	0,600	
Equivalence limit difference, $\pi_T - \pi_S, \Delta_0$	0,100	
Test expected proportion, π_T	0,600	
Expected difference, $\pi_T - \pi_S, \Delta_1$	0,000	
Power (%)	80	
n per group	217	

Huom.

Jos ei ole realistista tehdä oletusta $\pi_E = \pi_C$, vaan $\pi_E < \pi_C$, niin aineistokokoa joudutaan yleensä selvästi kasvattamaan, koska kaavassa (6) nimittäjä pienenee.

Kvantitatiivinen lopputulos, kahden keskiarvon vertailu

Usein kliinisten tutkimusten lopputulosmuuttuja on kvantitatiivinen, ja halutaan tutkia, onko ryhmien keskiarvojen välillä eroa. Mikäli kyseessä on yksi ryhmä ja kaksi ajankohtaa, käytetään joko **parittaista t-testiä** tai sen parametritonta vaihtoehtoa **Wilcoxonin testiä**. Jos taas kyseessä on toisistaan riippumattomat vertailtavat ryhmät, testinä käytetään **kahden otoksen t-testiä** tai **Mann–Whitney–Wilcoxon testiä**.

Yksi otos, yksi tai kaksi ajankohtaa tai periodia

Käyttötilanne:

- halutaan verrata aineistosta laskettua keskiarvoa hypoteettiseen arvoon μ_0 , tai
- halutaan tutkia jonkin kvantitatiivisen suureen (esimerkiksi verenpaine) muutosta kahden ajankohdan välillä

Tällaisessa tilanteessa voidaan tarvittava aineistokoko laskea kaavalla:

$$n = (z_{1-\alpha/2} + z_{1-\beta})^2 / \delta^2 + \frac{z_{1-\alpha/2}^2}{2} \quad (7)$$

Kaavassa δ on vaihtoehtoisen hypoteesin ja nollahypoteesin määrittelemien keskiarvojen μ_A ja μ_0 standardoitu erotus $\delta = |\mu_A - \mu_0| / \sigma$, missä σ on tarkasteltavan suureen perusjoukon standardipoikkeama. Osoitteessa:

www.mv.helsinki.fi/sarna/stats/YhdenKeskiarvonVertailu.xls olevalla laskimella voidaan laskea kaavaan 7 perustuva arvio. Laskimella voidaan laskea myös käänteisesti a) voima suureiden α , δ ja n funktiona tai b) efektin suuruus δ suureiden α , voima ja n funktiona.

Huom.

Mikäli standardipoikkeamaa ei kyetä arvioimaan, niin δ :lle voidaan käyttää arvoja väliltä 0,1 – 1,0 (Cohen 1988). Vaikutusta voidaan arvioida δ :n suhteen seuraavasti:

δ	Arvio
noin 0,2	Pieni kliininen vaikutus
noin 0,5	Kohtalainen kliininen vaikutus
noin 0,8	Suuri kliininen vaikutus

Jos sitä vastoin tunnetaan tarkasteltavan suureen vaihteluväli, niin standardipoikkeama voidaan karkeasti arvioida jakamalla vaihteluväli neljällä. Myös ristikkäistutkimuksissa voidaan soveltaa kaavaa (7).

Esim.

Oletetaan, että ”terve-lapsi” -tutkimuksessa on todettu lapsen keskimääräiseksi kävelemäänoppimisiäksi $\mu_0 = 12$ kk. Halutaan tutkia oppivatko synnyynnäistä sydänvikaa potevat lapset kävelemään myöhemmin kuin terveet lapset ja halutaan todeta 0,5 standardipoikkeaman ero kaksisuuntaisella testillä siten, että $\alpha = 0,05$ ja $1 - \beta = 0,95$. Oletetaan lisäksi, että tutkijoilla on niin vahva ennakkokäsitys eron

suunnaasta, että voidaan käyttää yksipuoleista testiä. Paljonko tutkittavia lapsia tarvitaan?

Yhden keskiarvon vertailu n:n arviointi $\alpha:n$, $\delta:n$ ja voiman perusteella		nQuery Advisor - [MOT0-1]	
Merkitsevyystaso α	0,050	One group t-test that a mean equals user	
1 tai 2 suuntainen testi	1		1
Nollahypoteesin mukainen keskiarvo μ_0	0,00	Test significance level, α	0,050
Vaihtoehtoinen keskiarvo μ_A		1 or 2 sided test?	1
Standardipoikkeama σ		Null hypothesis mean, μ_0	0,000
Efektin suuruus $\delta = \mu_A - \mu_0 / \sigma$	0,50	Alternative mean, μ_A	
Voima (%) $100(1 - \beta)$	95 %	Standard deviation, σ	
Aineistokoko (n)	45	Effect size, $\delta = \mu_A - \mu_0 / \sigma$	0,500
		Power (%)	95
		n	45

Tulkinta:

Kun aineistokoko on 45, niin yhden otoksen yksisuuntaisella t-testillä voidaan merkitsevyydellä 0,05 95 % voimalla todeta 0,5 standardipoikkeaman suuruinen ero.

Esim.

Jos aineistokoko kasvatetaan kaksinkertaiseksi, niin edellisessä esimerkissä voitaisiin todeta 0.35 standardipoikkeaman ero.

Yhden keskiarvon vertailu $\delta:n$ arviointi $n:n$, $\alpha:n$, voiman perusteella		nQuery Advisor - [MOT0-1]	
Merkitsevyystaso α	0,050	One group t-test that a mean equals user	
1 tai 2 suuntainen testi	1		1
Nollahypoteesin mukainen keskiarvo μ_0	0,00	Test significance level, α	0,050
Vaihtoehtoinen keskiarvo μ_A		1 or 2 sided test?	1
Yhteinen standardipoikkeama σ		Null hypothesis mean, μ_0	0,000
Efektin suuruus $\delta = \mu_A - \mu_0 / \sigma$	0,35	Alternative mean, μ_A	
Voima (%) $100(1 - \beta)$	95 %	Standard deviation, σ	
Aineistokoko (n)	90	Effect size, $\delta = \mu_A - \mu_0 / \sigma$	0,349
		Power (%)	95
		n	90

Esim.

Kahden periodin ristikkäistutkimusasetelma (nQueryn 6.0 manuaali s. 13.5). Suunnitellaan syöpätutkimusta, jonka tavoitteena on verrata uutta neutropeenian vähentämiseen tähtäävää hoitoa lumehoitoon. Lopputulosmuuttujana on niiden päivien lukumäärä, jolloin potilaalla kemoterapian seurauksena on neutropeniaa (neutrofiilien määrä alle tietyn rajan). Suunnitellaan, että käytetään **parittaista kaksisuuntaista t -testiä** merkitsevyydellä 5 %. Ristikkäistutkimusasetelman mukaisesti potilaat satunnaistetaan saamaan joko lume tai aktiivihoidoa

ensimmäisellä hoitoperiodilla ja puhdistusjakson ("wash-out period") jälkeen hoidot vaihtuvat. Oletetaan, että eri hoitosyklien (lume -hoito / hoito - lume) välillä ei ole siirtyvää vaikutusta ("carry-over effect"). Tutkija kysyy riittääkö **30 potilasta** tuottamaan riittävän voiman todeta neutropenian kestossa **reduktio 3 vrk**, kun aiemmassa pilottitutkimuksessa neutropenian on todettu kestävän keskimäärin **6 vrk** ja että erotusmuuttuja (hoito-lume) keskihajonta on **5,2**.

Yhden keskiarvon vertailu voiman arviointi n:n α:n, ja δ:n perusteella	
Merkitevyystaso α	0,050
1 tai 2 suuntainen testi	2
Nollahypoteesin mukainen keskiarvo μ ₀	6,00
Vaihtohtoinen keskiarvo μ _A	3,00
Yhteinen standardipoikkeama σ	5,20
Efektin suuruus δ = μ _A -μ ₀ / σ	0,58
Voima (%) 100(1 - β)	86 %
Aineistokoko (n)	30

nQuery Advisor - [MOT1-2]	
Paired t-test of mean difference equal to zero	
	1
Test significance level, α	0,050
1 or 2 sided test?	2
First condition mean, μ ₁	6,000
Second condition mean, μ ₂	3,000
Mean difference, μ _d = μ ₁ - μ ₂	3,000
Standard deviation of differences, σ _d	5,200
Effect size, δ = μ _d / σ _d	0,577
Power (%)	86
n	30

Tulkinta: 30 potilaan aineistolla saavutetaan 86 % voima todeta haluttu ero.

Kahden otoksen t-testi

Käyttötilanne:

Halutaan verrata kahden toisistaan riippumattoman ryhmän (esimerkiksi kahden hoitoryhmän) keskiarvoja toisiinsa t-testillä.

t-testi on moniin eri testausilanteisiin soveltuva parametrinen testi, joka perustuu t-jakaumaan. Testistä on olemassa monia eri versioita. Testin perusmuoto on:

$$t = \frac{(\text{otossuure}) - (\text{hypoteettinen arvo})}{(\text{otossuureen keskivirhe})}$$

Testistä on olemassa monia eri versioita riippuen nimittäjässä olevan keskivirheen arviointi- ja/tai vapausasteiden laskentatavasta.

t-jakauma "t-distribution" on yksi yleisimmin käytetyistä todennäköisyysjakaumista Normaalijakauman ohella. Jakauman muotoa säätelee parametri nimeltä vapausaste, jonka laskentakaava riippuu tutkimusaineiston koosta ja jakauman käyttötarkoituksesta. Jakaumasta käytetään myös nimitystä Studentin t-jakauma. Se on Normaalijakaumaa litteämpi ja pitkähäntäisempi.

Valitaan μ₁ :ksi esim. lumeryhmän keskiarvo ja μ₂:ksi hoitoryhmän keskiarvo. Kaava (7) modifioituu tällöin muotoon:

$$n = 2 \cdot (z_{1-\alpha/2} + z_{1-\beta})^2 / \delta^2 + \frac{z_{1-\alpha/2}^2}{4} \quad (8)$$

missä δ on vertailtavien ryhmien keskiarvojen μ₁ ja μ₂ standardoitu erotus δ = |μ₁ - μ₂| / σ, missä σ on ryhmien yhteinen perusjoukon standardipoikkeama.

Huom.

Jos $2\alpha = \beta = 0,05$, niin tällöin pätee likimain yhtälö $n = 26 / \delta^2$, eli jos standardoitu hoitoero on 1 (kliinisesti suuri ero) tarvitaan 26 potilasta kumpaankin ryhmään, ja jos $\delta = 0,1$ (kliinisesti hyvin pieni ero) tarvitaan potilaita 2600 kumpaankin ryhmään.

Osoitteessa www.mv.helsinki.fi/home/sarna/stats/KahdenKeskiarvonVertailu.xls on laskin, jolla kaavan 8 mukaiset laskut voidaan tehdä. Pienillä aineistoilla kaava (8) antaa aliestimaatin, koska tarkkaan ottaen laskentamenettelyssä pitäisi käyttää z:n asemesta t-jakauman prosenttipisteitä. Ne ovat kuitenkin aineistokoon funktioita, joten algebrallista ratkaisukaavaa ei silloin ole olemassa vaan pitäisi käyttää iteratiivista ratkaisumenettelyä. Kun arvioitu $n > 30$, niin virhe < 1 ja pienemmillä n:n arvoilla virhe on korkeintaan - 5 %.

Esim.

Suunnitellaan tutkimusta, jonka tavoitteena on vertailla lääkehoidon (ryhmä 2) ja lumehoidon (ryhmä 1) eroa systolisen verenpaineen keskimääräisen laskun suhteen. Oletetaan, että lumehoitoryhmässä ei tapahdu mitään muutosta, $\mu_1 = 0$. Sen sijaan lääkehoitoryhmässä oletetaan tapahtuvan laskua **10 mmHg ($=\mu_2$)**. Oletetaan, että aiempien vastaavanlaisen tutkimusten perusteella arvioidaan muutosta kuvaavan muuttujan hajonnaksi kummassakin ryhmässä $\sigma = 10$ mmHg, eli halutaan todeta standardoitu erotus $\delta = 1,0$, jos se on olemassa. Oletetaan, että käytetään **riippumattomien otosten t- testiä** ja että ero halutaan todeta **90 %:n** voimalla merkitsevyytason ollessa $\alpha = 0,05$ (kaksisuuntainen). Ohjelma nQuery käyttää iteratiivista ratkaisumenettelyä ja siksi pienillä n:n arvoilla tulee pieni ero verrattuna kaavalla 8 laskettuun tulokseen.

Kahden keskiarvon vertailu n:n arviointi α :n, δ :n ja voiman perusteella		nQuery Advisor - [MTT0-1]	
Merkitsevyytaso α	0,050	Two group t-test of equal means (equal n's)	
1 tai 2 suuntainen testi	2		1
Ryhmän 1 keskiarvo μ_1	0,00	Test significance level, α	0,050
Ryhmän 2 keskiarvo μ_2	10,00	1 or 2 sided test?	2
Keskiarvojen erotus $\mu_1 - \mu_2$	-10,00	Group 1 mean, μ_1	0,000
Yhteinen standardipoikkeama σ	10,00	Group 2 mean, μ_2	10,000
Efektin suuruus $\delta = \mu_1 - \mu_2 / \sigma$	1,00	Difference in means, $\mu_1 - \mu_2$	-10,000
Voima (%) $100(1 - \beta)$	90 %	Common standard deviation, σ	10,000
Aineistokoko (n)	22	Effect size, $\delta = \mu_1 - \mu_2 / \sigma$	1,000
		Power (%)	90
		n per group	23

Esim.

D -vitamiinihoito raskaana olevilla naisilla neonataalisen hypokalsemian ehkäisemiseksi (Cockburn et al 1980) Naiset satunnaistettiin kahteen ryhmään: D-vitamiinihoito ja lumehoito. Kalsiumin määrittäminen seerumista suoritettiin viikko lapsen syntymän jälkeen. Oletetaan, että $\mu_1 = 9,0$ mg/100 ml, $\sigma = 1,8$ mg/100 ml ja kliinisesti merkitsevä muutos olisi $|\mu_1 - \mu_2| = 0,5$ mg/100 ml. Päätetään, että käytetään arvoja: $\alpha = 0,05$, $1 - \beta = 0,95$. Tällöin standardoitu hoitovasteiden erotus δ on $0,5 / 1,8 = 0,28$ mg/100 ml.

Kahden keskiarvon vertailu n:n arviointi α :n, δ :n ja voiman perusteella	
Merkitsevyystaso α	0,050
1 tai 2 suuntainen testi	2
Ryhmän 1 keskiarvo μ_1	9,00
Ryhmän 2 keskiarvo μ_2	8,50
Keskiarvojen erotus $\mu_1 - \mu_2$	0,50
Yhteinen standardipoikkeama σ	1,80
Efektin suuruus $\delta = \mu_1 - \mu_2 / \sigma$	0,28
Voima (%) $100(1 - \beta)$	95 %
Aineistokoko (n)	338

nQuery Advisor - [MTT0-1]	
Two group t-test of equal means (equal n's)	
	1
Test significance level, α	0,050
1 or 2 sided test?	2
Group 1 mean, μ_1	9,000
Group 2 mean, μ_2	8,500
Difference in means, $\mu_1 - \mu_2$	0,500
Common standard deviation, σ	1,800
Effect size, $\delta = \mu_1 - \mu_2 / \sigma$	0,278
Power (%)	95
n per group	338

Huom.

μ_1 :n, σ :n ja δ :n valinta on käytännössä yleensä vaikeaa. Esimerkissä voitaisiin menetellä seuraavasti: Asetetaan lapsen hypokalsemian rajaksi se, että seerumin kalsium ylittää 7,4 mg/100 ml ja oletetaan, että $\pi_1 = 20\%$ (lumeryhmä) ja $\pi_2 = 10\%$ (D-vitamiinihoitoryhmä), $\Delta = 10\%$, $\alpha = 0,05$ ja $1 - \beta = 0,95$ Kaavaan (2) perustuva Excel-ohjelma (www.mv.helsinki.fi/home/sarna/stats/KahdenSuhtOsuudenVertailu.xls) antaa tuloksen $n = 325$. Tämä tulos perustuu olettamukseen, että testinä käytetään χ^2 -testiä ilman jatkuvuuskorjausta. Ohjelma nQuery käyttää jatkuvuuskorjattua testiä, jolloin tulos on: $n = 329$. Näistä edellinen on aliestimaatti ja jälkimmäinen yliestimaatti.

Parametrittomat testit

Mikäli t - testin asemesta käytetään joko **Wilcoxonin parittaista testiä** ("Wilcoxon matched pair test", "Wilcoxon pairwise test") tai **Wilcoxonin järjestyslukujen summatestiä** ("Wilcoxon Rank - Sum test") joka on sama kuin **Mann-Whitneyn U -testi**, voidaan voimalaskelmat tehdä edellä olevilla kaavoilla käyttäen yksinkertaista korjaustekijää.

Parametriton menetelmä "non-parametric method" on menetelmä, joka ei perustu mihinkään jakaumaoletuksiin, vaan siinä tarvittavat suureet lasketaan suoraan havaintoaineistosta esim. muodostamalla järjestyslukuja sekä niiden perusteella johdettuja suureita (esim. Wilcoxon-Mann-Whitney'n testisuure). Parametrittomia menetelmiä käytetään etenkin pienissä ja/tai hankalissa havaintoaineistoissa, esim. kun aineisto sisältää poikkeavia havaintoarvoja.

Korjaustekijä riippuu jakaumien todellisesta muodosta seuraavasti:

	Korjaustekijä
Tasainen jakauma	1
Kaksinkertainen eksponenttijakauma	$2/3 \approx 0,67$
Logistinen jakauma	$9/\pi^2 \approx 0,91$
Normaali jakauma	$\pi/3 \approx 1,05$

Todetaan, että normaalijakauman tapauksessa Mann–Whitneyn U-testiä käytettäessä tarvitaan noin 5 % suurempi aineisto kuin käytettäessä t - testiä. Yleensä Mann–Whitneyn testiä käytetään kun jakaumat ovat vinoja, jolloin Mann–Whitneyn testi vaatii pienemmän aineistokoon kuin t - testi; esimerkiksi logistisen jakauman tapauksessa vain 91 % t -testin vaatimasta aineistokoosta. Ohjelmassa nQuery on mahdollista arvioida aineistokokoa myös silloin kun käytetään parametrittomia testejä.

Esim.

Oletetaan, että edellä tarkastellussa verenpainetutkimusesimerkissä halutaan käyttää t -testin asemesta Mann–Whitneyn U-testiä. Aluksi arvioidaan vaikutuksen suuruus. Oletetaan että vertailtavien ryhmien havaintoarvot ovat x_{1i} , $i=1, \dots, n_1$ ja x_{2i} , $i=1, \dots, n_2$. Oletetaan, että $n_1 < n_2$. U -testisuurelle on seuraavanlainen käytännön tulkinta: Se ilmoittaa niiden havaintoparien lukumäärän (k), joille pätee $x_{1i} < x_{2i}$. Jos merkitään n :llä kaikkien mahdollisten havaintoparien määrää n :llä, niin vaikutuksen suuruus voidaan arvioida $p_1 = P(x_{1i} < x_{2i}) = k / n$. Standardoitua keskiarvojen erotusta 1,0 vastaavaksi arvioksi saadaan (nQuery) 0,760 ja tarvittavaksi potilasmääräksi tulee $n=26$ (nQuery). Tässä arviossa ei ole huomioitu jakauman muotoa, siksi n on suurempi kuin mitä t -testin yhteydessä saatiin.

MTT1-1					
Wilcoxon (Mann-Whitney) rank-sum test that $P(X<Y) = .5$ (continuous outcome)					
	1	2	3	4	5
Test significance level, α	0,050				
1 or 2 sided test?	2				
$p_1 = P(X<Y)$	0,760				
Power (%)	90				
n per group	26				

nQuery Advisor - [MTT1S-1]			
Estimate of $p_1 = P(X<Y)$ from means and SD assuming normality			
	1	2	3
Group 1 mean, μ_1	0,000		
Group 2 mean, μ_2	10,000		
Difference in means, $\mu_1 - \mu_2$	-10,000		
Common standard deviation, σ	10,000		
Effect size, $\delta = (\mu_1 - \mu_2) / \sigma$	-1,000		
$p_1 = P(X<Y)$	0,760		

Huom.

Mikäli ryhmien välillä ei olisi eroa, niin $p_1 = 0,500$. Standardoituja erotuksia 0,2 ("pieni vaikutus"), 0,5 ("kohtalainen vaikutus") ja 0,8 ("suuri vaikutus") vastaavat arvot ovat: 0,556, 0,638 ja 0,714.

Muita aineistokoon laskentatilanteita

Tutkimusasetelman ja -problematiikan johdosta joudutaan usein käyttämään myös muita kuin edellä esitettyjä kaavoja. Tällaisia tilanteita syntyy mm. silloin, kun halutaan

- vertailla samanaikaisesti **useampia ryhmäkeskiarvoja** (F-testi, Anova) keskenään (Esim. nQuery 6.0, Teoria: Day & Graham, 1989)

- **toistomittausten Anova** (Esim. nQuery 6.0)
- **satunnaistettujen lohkojen Anova** (Ohjelmat nQuery 6.0)
- seurata jonkin markkinoille tulleen uuden lääkkeen **sivuvaikutuksia** (Machin & Campbell, 1987)
- vertailla aikaa, joka kuluu jonkin tapahtuman ilmaantumiseen **logrank-testillä** (Esim. nQuery 6.0)
- käyttää **ositettua satunnaistamista**
- ottaa huomioon seurannasta kadotettujen vaikutus aineistokokoon
- arvioida aika, joka kuluu maksimaalisen hyödyn saavuttamiseen eri hoitomuodoilla
- verrata **korrelaatiokertoimia** keskenään. (Esim. nQuery 6.0)
- verrata ristikkäistutkimusasetelmassa preferenssejä (MedStat: SUPPSIZE).

Kommentteja

Kun tilastollinen ohjelmapaketti tai taulukoiden käyttö on tuottanut tutkijalle tarvittavaksi aineistokooksi esimerkiksi luvun 989, on seurauksena yleensä tutkijan turhautuminen; enhän minä noin suurta aineistoa voi saada kokoon. Tällöin on syytä muistaa, että huolimatta aineistokoon arvioinnin näennäisestä matemaattisesta tarkkuudesta, on taustalla kuitenkin monia ratkaisemattomia tieteellisiä ongelmia, kuten esimerkiksi seuraavat:

Yhden lopputulosmuuttujan käyttö laskentakaavoissa

- Käytännössä jokaisessa kliinisessä tutkimuksessa on useita sekä ”kovia” (esim. kuolema) että ”pehmeitä” (esim. arviointiskaalat) lopputulosmuuttujia. Aineistokoon arviointi pitäisi itse asiassa suorittaa käyttämällä monimuuttujamenetelmiä. Sellaisia menetelmiä, jotka toimisivat tyydyttävästi kliinisten tutkimusten yhteydessä, ei ole kuitenkaan ainakaan kirjoittajan tiedossa. Yleensä aineistokokolaskelmat suoritetaan tärkeimmän kovan lopputulosmuuttujan suhteen. Tällä menettelyllä on kuitenkin taipumus tuottaa tarpeettoman suuri aineistokoko pehmeiden muuttujien tarpeisiin, sillä saavutettavissa olevat kliinisesti merkitykselliset erot ovat yleensä kovissa muuttujissa suhteellisesti pienempiä kuin pehmeissä. Suositeltavaa onkin kokeilla aineistokokolaskelmissa tärkeimpiä molemmantyyppisiä muuttujia ennen lopullista päätöksentekoa.

Perusjoukon tuntemattomien suureiden arvojen valinta

- Tällaisia perusjoukon tuntemattomia suureita ovat esimerkiksi kliinisesti merkityksellinen ero (Δ), kontrolliryhmän suhteellinen osuus (P_c), riskisuhde (R), perusjoukon standardipoikkeama jne. Näiden suureiden valinta perustuu yleensä aikaisempiin vastaaviin tutkimuksiin tai on enemmän tai vähemmän hatusta otettua. Valinnalla saattaa olla kuitenkin huomattava merkitys aineistokoon kannalta. Vaikeimmin arvioitavissa ovat tavallisesti **standardipoikkeamat**, sillä tarkasteltavan suureen jakaumaa ei yleensä kovin hyvin tunneta, ja arvio joudutaan muodostamaan yhdistelemällä useiden erillistutkimusten tuloksia. Kannattaa kuitenkin mieluummin yrittää arviointia kuin tyytyä tekemään tarkasteltava suure kaksiarvoiseksi, jolloin menetetään paljon informaatiota.

Tilastollisen merkitsevyytason ja testin voimakkuuden valinta

- Nämä valinnat ovat täysin tutkijan määriteltävissä. Ei ole olemassa mitään matemaattista sääntöä, joka määräisi esimerkiksi, että merkitsevyytason pitää olla 5 %, jota tunnetulla biometrian oppi -isällä R. A. Fisherillä oli tapana käyttää ja josta myöhemmin muodostui koko tilastomaailman noudattama käytäntö. Lisäksi pitää ottaa huomioon tullaanko tutkimuksessa käyttämään **yksi - vai kaksisuuntaista testiä**. Yleisohje tähän tilanteeseen on seuraava:
 - käytetään kaksisuuntaista testiä, ellei aikaisempiin tutkimuksiin perustuva näyttö yksisuuntaisen testin puolesta ole erittäin vakuuttava
 - käytetään merkitsevyytasoja 5 % tai 1 %
 - käytetään testin voimakkuutta 80 % - 95 %.

Viitteet

Brown SE. Estrogen and endometrial cancer in a retirement community. NEJM 294: 1262-67, 1976.

Cappuccio FP et al. Lack of the effect of oral magnesium on high blood pressure. A double blind study. BMJ 291: 235-8, 1985

Cockburn F, **Belton** NR, **Purvis** RJ et al. Maternal vitamin D intake and mineral metabolism in mothers and their newborn infants. BMJ 281: 11-14, 1980.

Day SJ, **Graham** DF. Sample size and power for comparing two or more treatment groups in clinical trials. BMJ 1989; 299: 663-5.

Familiari L, **Postorino** S, **Turiano** S, **Luzza** G. Comparison of pirenzepine and trithiozine with placebo in treatment of peptic ulcer. Clinical Trial J. 18: 363-8, 1981.

Freiman KA, **Chalmers** TC, **Smith** H Jr, **Kuebler** RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. NEJM 299: 690-694, 1978.

Machin D, **Campbell** MJ. Statistical Tables for the Design of Clinical Trials. Blackwell Scientific Publications, Oxford 1987

Makuch R, **Simon** R. Sample size requirements for evaluating a conservative therapy. Cancer Treat, Rep. 621: 1037-40, 1980.

Pocock SJ. Clinical trials. A practical approach. ss.123-138. John Wiley and Sons, Chichester, 1983.

Smith CR, **Lewith** GT, **Machin** D. A preliminary study to establish a controlled method of assessing transcutaneous nerve stimulation (TNS) as a treatment for the pain caused by osteoarthritis (OA) of the knee. Physiotherapy 69: 266-8, 1983.

Taulukoita ja ohjelmia

Machin D, **Campbell** MJ, **Say-Beng** Tan, **Sze-Huey** Tan. Sample Size Tables for Clinical Studies. €67,50. <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1405146508.html>

Taulukkokokoelma, jossa on paljon esimerkkejä kliinisistä tutkimuksista.

nQuery Advisor 7.0: Statistical Solutions. <http://www.statsol.ie/index.php?pageID=2>
Monipuolinen yleisohjelma aineistokoon suunnitteluun, mutta kallis, hinta €1250.

SPSS Sample Power 2 for Windows, <http://www.academic-collegiate.com/SPSS-Sample-Power-2-p/112015.htm> Hinta \$525.

Egret[®]Siz. Sample Size and Power for Nonlinear Regression. Cytel Statistical Software & Services Dos-pohjainen ohjelma. <http://www.cytel.com/products/Egret/featurespec.asp>

Joissakin yleispaketeissakin, kuten esim. **NCSS:ssä** mukana "Power"-moduli

Netissä on myös monia web-sivuja, joilla voi tehdä aineistokoon arviointeja, esim. osoitteessa: <http://www.biostat.ucsf.edu/samplesize.html> tai <http://www.stat.uiowa.edu/~rlenth/Power/index.html>

Satunnaistaminen ja sokkouttaminen

Yleistä

Huomattava osa, etenkin faasi III:n, kliinisistä tutkimuksista on **vertailevia** ("**comparative trials**"), joissa yhden tai useamman hoidon tehoa verrataan johonkin yleisesti hyväksyttyyn/käytettyyn standardihoitoon, kontrolliin.

Uutta hoitoa saaneiden vertaaminen aikaisemmin hoidettuihin potilaisiin ("**historical controls**") voi antaa erittäin harhaisen tuloksen, koska olosuhteet ja hoitokäytännöt saattavat olla olennaisesti muuttuneet. Nykyisin historiallisten kontrollien käyttö tutkimuksissa on melko harvinaista. Joskus eettiset syyt saattavat pakottaa historiallisten kontrollien käyttöön.

Tieteellisesti ideaalista vertailevaa kliinistä koetta voidaan luonnehtia seuraavasti:

- Vertailtavat ryhmät tutkitaan samanaikaisesti.
- Vertailtavat ryhmät ovat tutkimuksen aloitushetkellä mahdollisimman samankaltaisia kaikkien muiden tekijöiden kuin tutkittavan uuden hoitomuodon tai lääkkeen suhteen.
- Vertailukelpoisuus tulisi säilyä tutkimuksen kuluessa.

satunnaistaminen "randomization" on kokeellisissa tutkimuksissa käytetty menetelmä, jonka tarkoituksena on:

1. tehdä vertailtavat ryhmät tausta- tai ennustetekijöiltään mahdollisimman samankaltaisiksi tutkimuksen alkaessa ja
2. estää subjektiivisten tekijöiden vaikutus tutkimuksen lopputulokseen.

Esim.

Satunnaistamisen tarpeellisuus. Monivitamiinihoidot ja hermostoputken sulkeutumishäiriöt. (Smithells et al. 1980, Smithells et al. 1981). Monivitamiinihoitoa raskauden aikana saaneita äitejä verrattiin ryhmään äitejä, jotka eivät olleet saaneet vitamiinisupplementaatiota. Kahdesta tutkimuksesta saatiin yhdistettynä seuraavat tulokset kun lopputulosmuuttujana ("outcome measure") oli syntyneen lapsen hermostoputken sulkeutumishäiriön suhteellinen määrä.

Yhdistetty lopputulos oli:

	<i>Monivitamiinihoito</i>	<i>Kontrolliryhmä</i>
<i>Synnytysten lukumäärä</i>	397	493
<i>Häiriöiden lukumäärä</i>	3 (0,8 %)	23 (4,7 %)

Tutkijat ilmoittivat johtopäätöksensä, että monivitamiinihoidosta on merkittävä hyöty ($P = 0.003$, yksisuuntainen testi) ja suosittelivat monivitamiinihoidon käyttöönottoa raskaana oleville naisille.

Ongelmana tässä tutkimuksessa on se, että kontrolliryhmä ei ollut satunnaistettu. Kontrolliryhmän lähempi tarkastelu osoitti mm. ,että siihen oli valikoitunut huomattavasti enemmän äitejä korkean riskin alueelta, Pohjois- Irlannista. Satunnaistamattomien tutkimusten perusteella ei yleensä voida saada riittävä

tieteellistä näyttöä todellisista hoitoeroista ja siten hoitokäytäntöjä ei niiden perusteella tulisi muuttaa.

Todellinen ongelma satunnaistamattomien tutkimusten suhteen on kuitenkin se, että ne saattavat estää tai olennaisesti haitata satunnaistettujen tieteelliset kriteerit täyttävien tutkimusten suorittamista, koska niitä saatetaan pitää epäeettisinä. Yllä mainitutkin tutkimukset herättivät huomattavaa keskustelua Brittein saarilla 1980-luvulla ja useita satunnaistettuja tutkimuksia aiheesta on jälkeinpäin suoritettu. Edelleen kuitenkin hermostoputken sukeutumishäiriön syy on tuntematon, mutta mm. foolihapolla on todettu olevan tärkeä merkitys sen ennaltaehkäisyssä.

Satunnaistamisen ohella toinen tärkeä asia kliinisissä kokeissa on sokkouttaminen.

Sokkouttaminen "blinding, masking" on menettely, jolla pyritään hoitokokeissa välttämään hoitohenkilökunnasta, tutkijoista ja potilaista johtuvaa informaatioharhaa. Sokkouttamismenettelyn suhteen kliinisiä kokeita on neljää eri tyyppiä:

- avoimet ("open"); ei käytetä sokkouttamista
- yksöissokko ("single-blind")
- kaksoissokko ("double blind")
- kolmoissokko ("triple blind")

Avoimissa tutkimuksissa satunnaistaminen suoritetaan esimerkiksi leikkausjärjestyksessä (ei yleensä sairaalaantulojärjestyksessä). Tällöin on erittäin tärkeää että, satunnaistamiskoodi ei ole kenelläkään hoitohenkilökunnasta tiedossa ennen kuin leikkaus on suoritettu. Sen jälkeen, yleensä jonkun tietyn protokollassa määritellyn ajan kuluttua päätetään, täyttääkö potilas jatkohoidon edellytykset ja pidetäänkö hänet tutkimuksessa, jonka jälkeen sekä potilas, että hoitohenkilökunta on tietoinen mitä jatkohoitoa kukin potilas saa. Tutkimusta valvovien henkilöiden vastuulla on seurata hoidon toteutumista ja asianmukaista kirjaamista. Potilaiden hoidossa tulee noudattaa tarkoin protokollaa, jotta subjektiiviset tekijät eivät pääsisi vaikuttamaan lopputulokseen. Kaikkia potilaita tulee tutkia yhtä usein, eikä esimerkiksi siten, että uutta hoitoa saavia tutkittaisiin useammin ja tarkemmin. Etenkin avoimissa tutkimuksissa randomisaatiokoodi kannattaa sulkea potilaskohtaisiin kirjekuoriin, jotka kulkevat potilaspapereiden mukana.

Yksöissokkotutkimuksissa ainoastaan potilas ei tiedä mitä hoitoa hän saa, mutta hoitohenkilökunta tietää. Valtaenemmistö vertailevista tutkimuksista toteutetaan kuitenkin **kaksoissokko**periaatteella, jolloin myöskään hoitohenkilökunta ei ole tietoinen potilaan saamasta hoidosta. Satunnaistamiskoodi avataan vasta, kun tutkimustiedosto on esitarkastettu, korjattu ja suljettu. Tutkimustuloksia käsittelevien henkilöidenkään ei tulisi tietää hoitokoodeista ennen kuin tärkeimmät tulokset on analysoitu. Tämä niin kutsuttu **kolmoissokkomenetelmä** ("triple blind") tarkoittaa, että tilastanalyysitkin suoritetaan käyttäen hoitoryhmistä peitekoodeja (esim. A, B, C, jne.) analysoinnin alkuvaiheessa.

Huomattava osa kliinisistä kokeista analysoidaan **hoitoaikeen mukaisesti**, eli nk. **"intention to treat (ITT)"** periaatteella. Se tarkoittaa sitä, että kukin potilas sisällytetään tilastanalyysissä siihen ryhmään, mihin hänet on satunnaistettu, riippumatta siitä toteutuuko hoito vai ei. Näin vältetään valikoitumisesta aiheutuvaa harhaa. Toinen analysoititapa on **protokollan mukainen analyysi**. Se tarkoittaa, että potilaat analysoidaan niissä ryhmissä, mitä hoitoa he todellisuudessa ovat saaneet ja analyysihin otetaan mukaan vain ne potilaat, joilla hoito on toteutunut. Tämä

analysointitapa arvioi monen tutkijan mielestä paremmin eri hoitojen biologisia vaikutuksia (**"efficacy"**).

Satunnaistetut ryhmät voivat olla kooltaan erisuuria. Tavallisinta on kuitenkin käyttää 1:1 satunnaistamissuhdetta, eli yhtä paljon potilaita uuteen ja kontrollihoitoon tai jos vertailtavia hoitoja on useita, niin käytetään samaa potilasmäärää kaikissa ryhmissä. Joskus eettisiin syihin vedoten vertailuryhmä saatetaan jättää pienemmäksi. Ainakin 1/3 potilaista tulisi kuitenkin olla satunnaistettuja kontrolleita.

Niitä menetelmiä, jotka turvaavat sen, että kaikissa ryhmissä on haluttu määrä tutkimuksen lopussa, kutsutaan rajoitetuiksi. Vastakohtan muodostavat rajoittamattomat menetelmät, joita käytettäessä saatetaan päätyä hyvinkin erisuuruisiin potilasmääriin.

Yksinkertainen satunnaistaminen

Yksinkertaisessa satunnaistamisessa ei aseteta ehtoja ryhmien koolle. Menetelmää harvemmin käytetään sellaisenaan hoitokokeessa.

<i>Esim.</i>	<i>Sijoitusryhmä</i>	<i>Luvut</i>
Kaksi ryhmää	A	0 – 4
	B	5 – 9
Kolme ryhmää	A	1 – 3
	B	4 – 6
	C	7 – 9
	–	0

	<i>1</i>	<i>2</i>	<i>3</i>	
Luku	0	5	2	7 ...
Järjestys	–	B	A	C ...

Tulkinta

Satunnaislukugeneraattori on tuottanut luvut: 0, 5, 2, 7, ... Lukua 0 ei käytetä. Ensimmäinen potilas satunnaistetaan hoitoryhmään B, toinen A:han ja kolmas C:hen jne.

Huom.

Mikäli käytetään satunnaislukujen taulukoita, niin ensimmäinen luku otetaan aina satunnaisesti valitusta kohdasta. Tilastopakettien satunnaislukugeneraattorit käyttävät satunnaisesti valittua siemenlukua tähän tarkoitukseen.

Ongelma:

Yleensä tuloksena yksinkertaisessa satunnaistamismenettelyssä on erisuuret ryhmäkoot.

Seuraavasta taulukosta ilmenee kuinka suuret ryhmäkokojen erot voivat olla todennäköisyydellä a) $P \geq 0.05$ tai b) $P \geq 0.01$ (Lähde: Pocock 1983).

Taulukko:

Mahdollinen epätasapaino ryhmäkokojen erossa käytettäessä yksinkertaista satunnaistamista kahteen hoitoryhmään.

Potilaiden määrä	$P \geq 0,05$	$P \geq 0,01$
	Ryhmäkoot	Ryhmäkoot
10	2 : 8	1 : 9
20	6 : 14	4 : 16
50	18 : 32	16 : 34
100	40 : 60	37 : 63
1000	469 : 531	459 : 541

Tulkinta: Esim. jos kokeeseen on päätetty ottaa yhteensä 50 potilasta, jotka satunnaistetaan kahteen hoitoryhmään käyttäen rajoittamatonta satunnaistamista, niin on yli 5 %:n mahdollisuus, että toiseen ryhmään tulee 18 ja toiseen 32 potilasta.

”Korttipakan sekoittamis” -menettely

Oletetaan, että kyseessä on kahta hoitoa A ja B vertaileva kliininen koe ja halutaan kumpaankin hoitoryhmään 20 potilasta. Tehdään pakka, jossa on 40 korttia siten, että 20 merkitty A:lla ja 20 B:llä. Pakka sekoitetaan ja kukin kortista suljetaan vuorotellen sinetöityihin kirjekuoriin, jotka numeroidaan 1 – 40 ja liitetään potilaspapereihin potilaiden sisäänottojärjestyksessä. Menetelmä takaa saman potilasmäärän vasta, kun kaikki potilaat on otettu sisään. Satunnaistaminen tätä menetelmää käyttäen voidaan tehdä esim. osoitteesta <http://graphpad.com/quickcalcs/RandMenu.cfm> löytyvällä ohjelmalla.

Esim. 20 henkilön satunnaistamislista kahden ryhmän tapauksessa.

Assign subjects to groups

Subject # Group Assigned

1	B
2	A
3	A
4	B
5	A
6	A
7	B
8	A
9	A
10	A
11	A
12	B
13	A
14	B
15	B
16	B
17	B
18	B
19	A
20	B

How it works: The random number generator is seeded with the time of day, so it works differently each time you use it. Each subject is first assigned to a group nonrandomly. Then the assignment of each subject is swapped with the group assignment of a randomly chosen subject. This should suffice, but the entire process is repeated twice to make sure it is really random. Note that you can copy and paste the values from the web page into Excel.

Satunnaiset permutoidut lohkot

Korttipakan sekoittamismenettelyn ongelmana oli, että potilasmäärät eivät välttämättä ole yhtä suuret ennen kuin kaikki potilaat ovat tulleet tutkimukseen. Tämän epäkohdan korjaamiseksi kliinisissä kokeissa satunnaistamismenetelmänä käytetäänkin tavallisimmin ns. satunnaisten lohkojen menetelmää. Se turvaa saman määrän potilaita hoitoryhmiin tietyin tasaisin lukumäärävälein. Esim. jos käytetään lohkokokoa 10 ja hoitoryhmiä on kaksi, niin aina 10:n potilaan välein ryhmäkoot ovat balanssissa, kun potilaita otetaan tutkimukseen sisään.

Esim. Kaksi ryhmää A ja B ja lohkon ("block") suuruus on 2.

Järjestys	Arvottu luku
AB	0 – 4
BA	5 – 9

Luku: 0, 5, 2, ...

Järjestys: AB, BA, AB, ...

Menettely:

Mikäli kyseessä on kahden samanaikaisen vertailevan hoidon tutkimus, niin satunnaistamislista muodostetaan siten, että ensimmäiselle potilaalle tulee hoidoksi A, toiselle B jne. Mikäli kyseessä on kahden periodin ristikkäistutkimus, niin satunnaistamislista voidaan muodostaa siten, että kirjainjärjestykset AB ja BA määräävät hoitojen antamisjärjestyksen. Ensimmäinen potilas aloittaa siten hoidolla A ja jatkaa toisella periodilla hoidolla B. Toinen potilas sitä vastoin aloittaa hoidolla B ja siirtyy hoidolle A toisella periodilla.

Kahden samanaikaisen hoidon vertailevissa tutkimuksissa lohkokokoa 2 pitäisi välttää (paitsi ositetussa satunnaistamisessa) ennustettavuuden vuoksi.

Mikäli kyseessä on kaksi vertailtavaa hoitoryhmää, A ja B ja halutaan ja samansuuruiset ryhmät aina neljän potilaan välein, lohkokooksi tulee siten neljä ja satunnaistamislista muodostetaan esim. seuraavasti. Kirjaimet A ja B voidaan järjestää kuudella eri tavalla neljän kirjaimen lohkoihin. Annetaan näille eri järjestyksille luvut 1 – 6 alla olevan taulukon mukaisesti.

Järjestys	Luku
AABB	1
ABBA	2
BBAA	3
BABA	4
ABAB	5
BAAB	6

Lukuja 7 - 9 ja 0 ei käytetä. Oletetaan, että on saatu satunnaisluvut 4, 2, 6, jne. ... Satunnaistamislistan alku muodostuu siten seuraavanlaiseksi: B, A, B, A, A, B, B, A, B, A, A, B, jne. ...

Mikäli käytetty lohkokokoko on pieni ja kyseessä on avoin tutkimus, niin hoitoryhmän ennustettavuus kunkin lohkon lopussa voi muodostua ongelmaksi. Esim. yllä olevassa listassa, jos tiedetään kolmen ensimmäisen potilaan samaa hoitoa, niin tiedetään myös

neljännen. Myöhemmin käsiteltävässä ositetussa satunnaistamisessa tätä ongelmaa voidaan pienentää siten, että peräkkäiset lohkot sijoitetaan eri ositteisiin.

Toteutus: Satunnaistamislista lohkosatunnaistamista käyttäen voidaan tehdä esim. osoitteesta <http://www.randomization.com> löytyvällä ohjelmalla. Tulosten voi helposti siirtää mm. Exceliin.

Harhaisen kolikon menetelmä

Satunnaistettujen lohkojen menettely takaa saman potilasmäärän aina kunkin lohkon jälkeen. Mikäli potilasmääriä ei välttämättä haluta aivan täsmälleen samoiksi, niin voidaan käyttää nk. harhaisen kolikon menetelmää.

Menettelytapa kahden ryhmän A ja B tapauksessa:

- Ensimmäiset 2 potilasta satunnaistetaan yksinkertaisella satunnaistamisella.
- Seuraava potilas sijoitetaan todennäköisyydellä $P > 0.5$ (esim. $P = 3/5$) pienempään ryhmään ja todennäköisyydellä $1-P$ ($2/5$) suurempaan ryhmään.
- Jos ryhmäkoot ovat yhtä suuret, käytetään yksinkertaista satunnaistamismenetelmää.

Esim.

Oletetaan, että käytetään arvoa $P = 3/5$. Jos ryhmät $k:n$ ($k \geq 2$) ensimmäisen potilaan jälkeen ovat erisuuret, niin potilas sijoitetaan pienempään ryhmään, mikäli saatu satunnaisluku on väliltä $0 - 5$ (eli $3/5$ luvuista $0 - 9$) ja suurempaan ryhmään, mikäli luku on väliltä $6 - 9$ ($2/5$). Jos ryhmät ovat yhtä suuret, niin käytetään vastaavasti lukuja: $0 - 4$ ryhmään A ja $5 - 9$ ryhmään B, eli samoja painoja (yksinkertainen satunnaistaminen).

Satunnaisluku	0	5	2	7	8	4	3	7	4	1	6
Ryhmä	A*	B	A*	A	A	B	B	A	B	B	B* ...

* Ryhmät ovat yhtä suuret, joten käytetään yksinkertaista satunnaistamista

Mikäli käytetään painokerrointa $3/5$, niin todennäköisyys sille, että jossain vaiheessa koetta ryhmäkoot eroaisivat enemmän kuin 10:llä on korkeintaan yksi kahdestakymmenestä, joten painokerroin $3/5$ on riittävä ainakin suurissa kokeissa ($n \geq 100$). Pienemmissä kokeissa kannattaa käyttää suurempaa painokerrointa, $3/4$ tai $2/3$.

Ositettu satunnaistaminen

Tavoite:

taata vertailtavien ryhmien samankaltaisuus (homogeenisuus) taustatekijöiden suhteen. Homogeenisuus on tärkeää mm. tilastollisen tehokkuuden kannalta.

Ositettu satunnaistaminen ("stratified randomization") ei ole suositeltavaa, mikäli

- tutkimus on suuri ja lähtöarvoilla ja aikaisilla tuloksilla ei ole relevanssia
- pienet organisatoriset resurssit
- on epävarmuutta relevanteista taustatekijöistä, jotka vaikuttavat lopputulokseen.

Menettelytapa:

- Muodostetaan ositteet tausta - tai ennustetekijöiden perusteella.

- Sovelletaan satunnaistettujen lohkojen menetelmää erikseen kuhunkin ositteeseen käyttäen melko pientä lohkokokoa (esim. 4).

Esim. Taustatekijät: x_1 ja x_2 saavat arvoja yhdestä kahteen.

Osite	1	2	3	4
x_1	1	1	2	2
x_2	1	2	1	2
	B	B	A	B
	A	B	A	A
	B	A	B	A
	A	A	B	B
	A	A	B	A
	B	A	A	B
	A	B	B	B
	B	B	A	A

Ongelma: Ositteiden määrän lisääntyessä tulee ositteita, joihin ei tule riittävästi tapauksia. **Ratkaisu: Minimointimenettely.**

Minimointimenettely

Minimointi "minimization" on menetelmä potilaiden allokoimiseksi hoitoryhmiin kliinisissä kokeissa siten, että vertailtavat hoitoryhmät olisivat mahdollisimman hyvin balanssissa ennusteeseen vaikuttavien tekijöiden suhteen. Vaihtoehto muille satunnaistamismenetelmille, esim. lohkosatunnaistamiselle.

Minimointimenettelyn (Taves 1974) tavoitteena on taata vertailtavien ryhmien samankaltaisuus, kun lopputulokseen vaikuttavia (korreloivia) taustatekijöitä on paljon. Sen toimintaperiaatetta voidaan kuvata seuraavalla yksinkertaistetulla esimerkillä.

Esim. Hoidot A ja B.

Tausta-/ prognostiset tekijät	Arvot
keskus	1 – 5
tauti	1 – 2
vaikeusaste	1 – 3

- Oletetaan, että on satunnaistettu 30 potilasta ja laskettu taustatekijöiden **keskus**, **tauti** ja **vaikeusaste** suhteen **frekvenssijakaumat ryhmissä A ja B**. (alla olevan taulukon rivit 2 ja 3).
- Oletetaan, että seuraavan potilaan (nro **31**:n) taustatekijät ovat: **keskus = 3**, **tauti = 1** ja **vaikeusaste = 3**.
- Riviltä 4 ilmenee silloin potilaan 31 aikaansaama lisäys frekvenssijakaumiin.
- Rivit 5 ja 8 ilmaisevat muuttuneet kohdat frekvenssijakaumissa, jos potilas 31 sijoitettaisiin ryhmään A tai ryhmään B.
- Rivit 6 ja 9 ilmaisevat vastaavat kohdat muuttumattomissa frekvenssijakaumissa.

- Rivi 7 on rivien 5 ja 6 erotuksen itseisarvo sarakkeittain ja rivi 10 on vastaava laskettuna riveistä 8 ja 9.
- Sarakkeeseen 5 on summattu nämä muutokset riveiltä 7 ja 10.

1: Ryhmät	Keskus					Tauti		Vaikeusaste			Summa
	1	2	3	4	5	1	2	1	2	3	
2: A	2	3	1	6	3	10	5	2	8	5	
3: B	3	1	3	4	4	9	6	3	6	6	
4: Potilas 31	0	0	1	0	0	1	0	0	0	1	
5: A + potilas 31	2					11		6			
6: B	3					9		6			
7: A + potilas 31 - B	1					2		0			3
8: B + potilas 31	4					10		7			
9: A	1					10		5			
10: B + potilas 31 - A	3					0		2			5

Tulkinta:

Summasarakkeesta voidaan todeta, että potilas 31 kannattaa sijoittaa hoitoryhmään A, koska silloin ryhmien A ja B taustamuuttujien frekvenssijakaumat tulevat lähemmäksi toisiaan.

Huom.

Kun potilaat sijoitetaan hoitoryhmiin käyttäen minimointimenettelyä, pitää lopputulos analysoida käyttäen kovarianssianalyysiä (tarkastellaan myöhemmin) siten, että kovariaateiksi otetaan minimoinnissa käytetyt taustatekijät. Jos käytettäisiin varianssianalyysiä, testauksen tuloksena saatava P -arvo olisi liian pieni; ylioptimistinen todelliseen α -tasoon verrattuna. Monen taustatekijän tapauksessa minimointimenettely yhdistettynä kovarianssianalyysiin kasvattaa testauksen voimaa verrattuna tavalliseen satunnaistamiseen. Edellytyksenä on luonnollisesti se, että kovarianssianalyysin käyttöedellytykset ovat voimassa.

Esim.

Minimointimenettelyn käyttö vertailevassa hoitotutkimuksessa (Meade ym., 1995.) Kyseessä on monikeskustutkimus, johon otettiin yhteensä 741 alaselkäoireyhtymää potevaa henkilöä, joiden ikä vaihteli välillä 18 -64 vuotta ja joilla manipulaatiohoito ei ollut kontraindisoitu. Tutkimuksen tavoitteena oli verrata sairaalan organisoimaa jatkohoitoa ja kiropraktikon suorittamaa hoitoa keskenään. Keskeisin lopputulosmuuttuja oli Oswestryn 10-osiainen selkäkipua mittaava kyselylomakeasteikko. Tulokset esitettiin skaalalla 0 -100. Tutkimuksen seuranta -aika oli kolme vuotta ja tutkimustulokset analysoitiin hoitoaikeen mukaisella periaatteella.

Hoitoaikeen mukainen analyysi, "intention to treat analysis", ITT -analyysi
 Kaikki satunnaistetut potilaat otetaan mukaan analyysiin, mikäli mahdollista. Menettelyllä pyritään välttämään kliinisestä kokeesta poispuodonneista potilaista ("dropouts") aiheutuvaa harhaa.

Satunnaistaminen suoritettiin soveltaen yllä kuvattua minimointiperiaatetta. Ennustavina tekijöinä käytettiin seuraavia luokkamuuttujia: lähettävä klinikka, episodin kesto (kaksiarvoisena; 1 kk tai alle tai yli 1 kk), aikaisemmat selkäkiput (kaksiarvoisena, kyllä - ei), Oswestry lähtötilanteessa (kaksiarvoisena; yli 40, 40

tai alle). Tutkimus antoi tulokseksi, että kiropraktikon hoidossa olleet saivat enemmän hyötyä hoidosta. Todennäköisin selitys tulokselle on se, että kiropraktikon hoidot jakaantuivat pidemmälle aikavälille. Kovin vakuuttavaa näyttöä hoitoerojen suuruudesta ei tutkimuksen perusteella voi kuitenkaan luotettavasti arvioida, mm. analyysimenetelmissä tulisi käyttää **toistomittausasetelmaa**. Tutkijat tyytyivät kuitenkin melko yksinkertaisiin perusmenetelmiin.

Toistomittausanalyysi “analysis of repeated measures” on toistomittausasetelmaa käyttävä analyysimenetelmä. **Toistomittausasetelma**, ”repeated measures design” on tutkimusasetelma, jossa pyritään vähentämään yksilöiden välisen vaihtelun vaikutusta tutkittavaan asiaan, esim. hoitojen vaikutukseen suorittamalla tutkittavista useita mittauksia esim. eri aikoina ja eri olosuhteissa.

Epäsymmetrisesti tasapainotetut menetelmät

Ylivoimaisesti tavallisinta kahden vertailtavan ryhmän kliinisissä kokeissa on, että satunnaistetaan sama määrä potilaita kumpaankin vertailtavaan ryhmään. Riippumatta lopputulos - tai vastemuuttujan luonteesta, näin aina onkin tehokkainta. Mikäli kuitenkin tavoitteena on saada esimerkiksi enemmän käytännön kokemusta uudesta hoidosta kuin standardihoidosta, saatetaan päätyä käyttämään epäsymmetrisesti tasapainotettuja satunnaistamismenetelmiä siten, että uutta hoitoa saavia potilaita otetaan tutkimukseen enemmän. Tällöin kuitenkin tutkimuksen voima laskee nopeasti, kun uutta hoitoa saavien osuus kasvaa yli 70 %:n (ks. Pocock, s. 88).

Puutteet satunnaistamisessa

Altman ja Dore (1990) tutkivat 80 artikkelia satunnaistetuista kliinisistä kokeista, jotka oli julkaistu jossakin seuraavista neljästä lehdestä: Annals of International Medicine, British Medical Journal, The Lancet tai New England Journal of Medicine. Arviointiraati kiinnitti erityistä huomioita virheisiin ja puutteisiin satunnaistamisen osalta.

Valikoituja tuloksia:

- 60 %:ssa artikkeleista metodiosasta puuttui maininta **mitä satunnaistamismenetelmää** oli käytetty.
- **Satunnaislukujen muodostamistapa** oli mainittu vain puolessa artikkeleista.
- **Lohkosatunnaistamista** käytettiin 30 %:ssa tutkimuksista.
- Vain 52 %:ssa tutkimuksista, joissa oli käytetty **osittamista**, käytettiin **lohkosatunnaistamista**. Osittaminen ei palvele tarkoitustaan, ellei satunnaistamisessa käytetä lohkomenetelyä.
- **Lohkokokoa** ei mainittu **35 %:ssa** tapauksista. Yhdessä **30** potilaan aineistossa käytettiin lohkokokoa **20!** Tämä lohkokoko on liian suuri kyseiseen tapaukseen. Lohkokoko tulee valita siten, että aineistokoko on jaollinen lohkokoolla. Tavallisimmin käytetyt lohkokoot ovat välillä 4 - 20. Liian pieniä lohkokokoja kannattaa välttää ennustettavuussyistä. 30 potilaan ja kahden hoitoryhmän tapauksessa sopiva lohkokoko olisi 6.
- **Hoidon allokointimenetelmä** oli puutteellisesti ilmoitettu **45 %:ssa** tapauksista,

ja loppuista 44 tutkimuksesta 16:ssa oli ilmoitettu, että on käytetty kirjekuoria, mutta ainoastaan kahdessa tutkimuksessa oli ilmoitettu, että ne olivat numeroituja, sinetöityjä ja läpinäkymättömiä, niin kuin pitäisi.

Ohjelmia

nQuery Advisor versio **6.0**: Statistical Solutions, LTD, 2005. Hinta: €750.

<http://www.statsol.ie/nquery/nquery.htm>

- Monipuolinen yleisohjelma aineistokoon arviointiin ja satunnaistamislistojen tekoon

Yksinkertaisia satunnaislistojen tekoon ja satunnaislukujen generointiin soveltuvia ohjelmia löytyy internetistä mm. sivuilta:

<http://www.randomization.com> ja

<http://graphpad.com/quickcalcs/RandMenu.cfm>

Viitteet

Altman DG. Comparability of randomized groups. *Statistician* 1985; 34: 125-36.

Altman DG, Dore CJ. Randomization and baseline comparisons in clinical trials. *Lancet* 1990; 335: 149-53.

Meade TW, Dyer S, Browne W, Frank AO. Randomized comparison of chiropractic and hospital outpatient management for low back pain: results from extended follow-up. *BMJ* 1995; 311: 349-351.

Smithells RW, Sheppard S, Schorah JC et al. Possible prevention of neural tube defects by periconceptional vitamin supplementation. *Lancet* 1980, i 339- 40.

Smithells RW, Sheppard S, Schorah JC et al. Vitamin supplementation and neural tube defects. *Lancet* 1981, ii 1425.

Pocock SJ. *Clinical trials; a practical approach*. John Wiley & Sons, Chichester 1983: 66-99.

Taves DR. Minimization: A new method of arranging patients to treatment and control groups. *Clin Pharmacol Ther* 1974; 15: 443- 53.

Aineiston kerääminen, tallentaminen ja käsittely

Otantamenetelmiä

Huomattava osa lääketieteellisistä tutkimuksista perustuu **otantaan ("sampling")**. Sen suorittamistapoja on useita. Seuraavassa esitellään lyhyesti kliinisissä tutkimuksissa tavallisimmin käytetyt menetelmät.

Otantatutkimus "sample study" on otantaan perustuva epäkokeellinen tutkimusasetelma, jonka tavoitteena on tehdä luotettavia perusjoukkoa koskevia johtopäätöksiä tutkimalla huolella valittu ja sopivan kokoinen osa perusjoukkoa. Otantatutkimuksen tulokset eivät koskaan ole tarkkoja vaan niihin sisältyy aina otannasta aiheutuvaa vaihtelua.

Yksinkertainen satunnaisotanta

Käyttö:

Kun perusjoukko äärellinen ja henkilöt ovat identifioituja, kuuluvat esim. johonkin rekisteriin.

Oletetaan, että perusjoukkoon kuuluu N henkilöä (tai muuta havaintoyksikköä), joille on annettu tunnukset $1, 2, \dots, N$ ja että tästä perusjoukosta halutaan kerätä n henkilöä käsittävä tutkimusaineisto (otos).

Menettelytapa:

Muodostetaan tasaisesti jakautuneita satunnaislukuja väliltä $(1, N)$ käyttäen satunnaislukujen taulukoita tai tietokoneiden satunnaislukujen muodostamisohjelmia.

Yksinkertaisessa satunnaisotannassa ("simple random sampling") jokaisella mahdollisella $n:n$ henkilön suuruisella otoksella, joka on valittavissa tästä N henkilöä käsittävästä perusjoukosta, on siten sama todennäköisyys tulla tutkijan aineistoksi. Tässä suhteessa on selvä ero kliinisissä kokeissa käytettyyn menettelyyn, missä tutkimukseen kelpuutetut ("eligible") henkilöt satunnaistetaan vertailtaviin ryhmiin, jolloin kullakin potilaalla on yhtä suuri todennäköisyys joutua mihinkä tahansa vertailtavista ryhmistä. Tällöin vertailtavat ryhmät eivät ole satunnaisotoksia mistään perusjoukosta. Usein myöskään tutkimukseen kelpuutettujen $N:n$ henkilön ryhmä, ei ole satunnaisotos mistään perusjoukosta, vaan se on valittu muilla kriteereillä.

Systemaattinen otanta

Käyttö:

Kun perusjoukko on ääretön tai siihen kuuluvia henkilöitä ei etukäteen tiedetä. Esim. poliklinikkatutkimuksissa on usein tämä tilanne.

Menettelytapa:

Otetaan mukaan joka k:s, henkilö satunnaisesti valitusta ajanhetkestä alkaen.

Systemaattisessa ("systematic sampling") otantamenettelyssä perusjoukko voi olla myös ääretön. Otanta aloitetaan satunnaisesti valitusta kohdasta (esim. ajanhetkestä) ja sen jälkeen otokseen valitaan joka k:s (esim. joka kolmas tai joka neljäs) henkilö.

Verrattuna yksinkertaiseen satunnaisotantaan systemaattinen otanta on helpompi suorittaa; ei edellytä olemassa olevaa rekisteriä perusjoukosta. Tätä menetelmää voidaan käyttää esim., kun halutaan tutkia tapaturma-asetalle tuotavia potilaita ja erityisesti silloin, kun tutkimus edellyttää mm. tavanomaisesta rutiinista poikkeavia näytteenottoja ja lisämäärytyksiä, joita ei kaikille potilaille tehdä. Tällöin systemaattinen otanta on ainoa keino, koska potilaita ei ole ennakkoon rekisteröity ja koska tutkimukseen halutut tiedot eivät ole sellaisia, joita olisi systemaattisesti tallennettu potilaskertomuksiin. Toinen selvä etu on, että systemaattinen otos jakaantuu yleensä tasaisemmin yli perusjoukon, mistä johtuen systemaattinen otanta on usein huomattavastikin tarkempi kuin yksinkertainen satunnaisotanta.

Systemaattisuus voi kuitenkin joissakin tapauksissa aiheuttaa virheellisyyttä. Esim. tapaturma -asematutkimuksessa on aiheellista valita k siten, että otos ei valikoidu esim. kellonajan, viikonpäivän, vuodenajan tai muiden sellaisten tekijöiden suhteen, jotka ovat haitallisia tutkimustulosten yleistettävyyden kannalta. Tämän asian varmistamiseksi onkin aiheellista menetellä siten, että kerätään useita systemaattisia osaotoksia valitsemalla useita satunnaisaloituskohtia, joista alkaen otantaa suoritetaan, esim. keräämällä eri vuodenaikoina omat systemaattiset otokset.

Ositettu otanta

Käyttö:

Perusjoukko on äärellinen ja halutaan turvata edustavuus jonkin tekijän suhteen.

Menettelytapa:

Perusjoukko jaetaan äärelliseen määrään ositteita ("strata", perusjoukon osajoukkoja), joista kustakin erikseen suoritetaan satunnaisotanta. Ositteisiin jaon perusteina voi olla esim. ikäryhmitys tai keskussairaalapirijako.

Ositettua otantaa ("stratified sampling") käytetään mm. seuraavista syistä johtuen:

- halutaan parantaa koko otoksesta laskettavien suureiden tarkkuutta verrattuna yksinkertaisella satunnaisotannalla saataviin arvioihin
- halutaan turvata edustavuus niiden tekijöiden suhteen, jotka ovat ositteisiin jaon perustana
- halutaan laskea arvioita tutkittaville suureille kussakin ositteessa.

Ensimmäisessä tapauksessa on optimaalista valita ositteiden aineistokoot siten, että ne ovat niin hyvin kuin mahdollista verrannollisia tutkittavan muuttujan hajontaan perusjoukon vastaavassa ositteessa. Mikäli ei ole mitään tietoa perusjoukon hajonnoista eri ositteissa tai ollaan ennen kaikkea kiinnostuneita jälkimmäisestä vaihtoehdosta, niin otantasuhde on yleensä tällöin syytä valita vakioksi kaikissa ositteissa, mikä merkitsee, että kustakin ositteesta valittava otoksen koko on suhteessa ositteen kokoon perusjoukossa.

Esim.

Halutaan arvioida jonkin/joidenkin sairauksien yleisyyttä Suomen väestössä otantatutkimuksen avulla ja oletetaan, että on näyttöä siitä, että kyseiset sairaudet ovat epätasaisesti jakautuneita maan eri osien suhteen. Tällöin voidaan käyttää ositettua otantaa. Ositteisiin jaon perusteeksi valitaan esim. lääni tai keskussairaalaapiiri. Kustakin näin saatavasta ositteesta suoritetaan satunnaisotanta siten, että aineistokoko on suhteessa vastaavaan väestömäärään. Tämä menettely parantaa ennen kaikkea sairauksien esiintyvyyttä mittaavien suureiden tarkkuutta (hajontaa) verrattuna yksinkertaisella satunnaisotannalla saataviin arvoihin.

Kolmannessa tapauksessa voi käydä niin, että joudutaan poikkeamaan optimaalisesta otantasuunnitelmasta, koska se tuottaisi joihinkin ositteisiin liian vähän havaintoja. Tällöin käytetään nk. rikastusmenettelyä, mikä tarkoittaa, että aineistokoko kasvatetaan pienissä ositteissa. Esim., jos ositteisiin jaon perusteena on sosiaaliasemaryhmä, niin ylimmän sosiaaliasemaryhmän osuutta voidaan joutua kasvattamaan otosta valittaessa, koska sen osuus perusjoukosta on yleensä muita sosiaaliasemaryhmiä pienempi.

Monivaiheinen otanta**Käyttö:**

Menetelmää käytetään pääasiassa kustannusten säästämiseksi joissakin melko harvinaisissa tutkimustilanteissa.

Menettelytapa:

Perusjoukko jaetaan ensiksi ensimmäisen vaiheen otosyksiköiksi. Tämän jälkeen käytetään jotain edellä kuvatuista otantamenettelyistä ja valitaan otos näistä ensimmäisen vaiheen otosyksiköistä. Kukin otokseen valituista yksiköistä jaetaan edelleen toisen vaiheen otosyksiköiksi, joista seuraavaksi valitaan otos. Näin voidaan jatkaa niin kauan kuin on tarpeellista.

Monivaiheisen otantamenettelyn ("multiphase sampling") **etuna** kustannusten säästön ohella olla joissakin tilanteissa otantaprosessin helpottuminen. Se ei myöskään edellytä yhteistä rekisteriä koko perusjoukosta. **Haittana** voidaan pitää sitä, että yleensä arvioitavien suureiden tarkkuudessa menetetään.

Esim.

Jos halutaan suorittaa jonkin tietyn alueen (esim. pääkaupunkiseudun) koululaisiin kohdistuvaa tutkimusta, niin ensimmäisessä vaiheessa otantayksiköiksi valitaan koulut ja toisen vaiheen otantayksiköiksi luokat ja sitten kolmannen vaiheen otantayksiköiksi oppilaat. Tämyntyyppiseen järjestelyyn otannan osalta joudutaan turvautumaan ennen kaikkea käytännön järjestelyiden helpottamiseksi.

Aineiston kuvaaminen, tallentaminen ja käsittely

Tavallisesti tutkimuksissa joudutaan keräämään suuri määrä tietoa, useista tutkittavista ja monista tutkimuksen kannalta relevanteista suureista (**muuttujista**, "variables"). Jotta tämä tutkimustieto olisi mielekkäästi tulkittavissa, se joudutaan tiivistämään **yhdistelmä tiedoksi** ("aggregate data"), **tunnusluvuiksi** ("characteristic"), **taulukoiiksi** ("cross-tables") tai **graafisiksi kuvaajiksi** ("graphical displays").

Yhdistelmä tieto "aggregate data" on tieto, joka on muodostettu yhdistämällä tietoa useista eri tietueista, esim. muodostamalla jonkin henkilön kaikista sairaalakäynneistä yhteenlasketut hoitopäivät.

Ennen analyysivaihetta kannattaa suorittaa tutkimusaineiston perusteellinen tarkastelu ("data screening"), joka tulisi suunnitella yhtä huolellisesti kuin varsinainen tilastokäsittelykin. Aineiston esikäsittelyssä pyritään paikallistamaan aineistossa mahdollisesti esiintyvät virheellisyydet ja **poikkeavat havaintoarvot** ("outliers"), ja puuttuvien tietojen määrä ja laatu. Lisäksi jakaumien muotoon ja perusräppövuussuhteisiin kannattaa perehtyä huolella.

Hypoteettinen tietokanta TRIAL, Vertaileva tutkimus kolmen hoidon P, Q ja R välillä. (Lähde: Wulff HR, Sclichting P, 1991)

Nimi	Kuvaus	Muuttujan tyyppi
'Patno'	potilaan tunnistekoodi	Numeerinen
'Gender'	Sukupuoli: 1=Mies, 2=Nainen	Dikotominen
'Age'	Ikä	Jatkuva
'Therapy'	Hoito: P(=1), Q (=2) ja R (=3)	Laadullinen
'Smoker'	Tupakointi 0=Ei (N), 1=Kyllä (Y)	Dikotominen
'Alcohol'	Alkoholin kulutus 0=Ei, 1=Vähän, 2=Kohtalaisesti, 3="Runsaasti"	Järjestysasteikollinen
'Obs_1'	Maksan suurentuma (hepatomegalia); onko maksa palpoitavissa (= 1) vai ei (= 0) havainnoitsijan1 mukaan	Dikotominen
'Obs_2'	Hepatomegalia havainnoitsijan 2 mukaan	Dikotominen
'Imag_S'	Kuviteltu laboratoriomittaus hoidon alussa	Jatkuva
'Imag_1'	Kuviteltu laboratoriomittaus viikon kuluttua	Jatkuva
'Imag_2'	Kuviteltu laboratoriomittaus kahden viikon kuluttua	Jatkuva
'Status'	Potilaan tila seurannan päätyttyä: 0=ei tapahtumaa (R), 1=tapahtuma (E), 2=kadotettu (L)	Laadullinen
'Date_1'	Seurannan aloituspäivämäärä	Päivämäärä
'Date_2'	Seurannan lopetuspäivämäärä	Päivämäärä
'Time'	Seuranta-aika	Numeerinen

Pat no	G e n d e r	A g e	T h e r a p y	S m o k e r	A l c o h o l	O b s _ 1	O b s _ 2	Imag _S	Imag _1	Imag _2	Date_1	Date_2	S t a t u s	Time
501	1	29	2	1	1	1	0	15,1	12,1	18,1	11.12.1978	31.08.1979	1	263
502	2	56	2	1	1	1	1	24,4	23,6	29,4	03.01.1979	23.04.1979	1	110
503	2	64	3	1	1	0	0	54,7	59,2	65,5	11.01.1979	07.09.1979	1	239
505	2	37	3	1	2	0	1	21,4	35,2	34,2	11.01.1979	01.05.1979	1	110
506	1	70	3	0	2	1	1	39,5	51,9	70,5	14.03.1979	15.10.1979	1	215
507	2	50	1	1	2	0	0	33,2	28,5	30,3	20.03.1979	20.06.1979	1	92
601	2	52	2	1	2	1	1	39,5	50,6	53,3	01.04.1979	03.07.1979	1	93
602	2	18	2	0	0	1	1	21,4	22,1	22,1	20.04.1979	15.11.1979	2	209
603	1	59	3	1	3	1	0	46,6	67,1	105	25.04.1979	23.10.1979	1	181
604	2	50	3	0	2	0	.	22,8	26,8	26,8	27.04.1979	12.03.1980	1	320
605	1	65	1	0	3	0	0	22,8	28,5	22,1	09.05.1979	15.11.1979	1	190
606	1	42	2	1	2	1	1	24,4	23,6	27,6	10.05.1979	04.09.1979	1	117
607	1	61	2	1	2	0	1	29,4	31,2	38,4	17.05.1979	22.01.1980	1	250

608	1	28	1	1	3	0	0	50,6	63,9	62,3	12.06.1979	13.09.1979	1	93
609	1	24	3	1	1	1	1	14,6	14,6	14,1	30.06.1979	25.10.1979	2	117
611	2	30	1	1	2	0	0	15,1	12,1	11,2	11.07.1979	26.01.1980	1	199
612	2	36	2	1	3	0	0	.	20	24,4	31.07.1979	15.11.1979	1	107
613	2	52	3	1	2	1	0	41,8	40,6	57,7	02.08.1979	08.03.1980	1	219
701	1	38	2	0	3	0	0	23,6	24,4	29,4	17.08.1979	24.11.1979	1	99
702	2	46	3	1	3	0	0	28,5	37,3	53,3	23.08.1979	24.08.1980	2	367
703	1	60	2	1	2	1	0	40,6	37,3	51,9	05.09.1979	10.01.1980	1	127
704	1	47	2	1	2	1	1	23,6	26	28,5	10.09.1979	16.06.1980	1	280
705	1	41	1	1	1	1	1	35,2	34,2	37,3	30.09.1979	24.12.1979	2	85
706	2	56	1	1	1	0	0	32,2	30,3	24,4	22.10.1979	31.01.1980	1	101
707	1	73	3	0	0	1	1	56,2	59,2	70,5	24.10.1979	01.12.1980	0	404
708	1	37	3	1	2	0	0	18,1	26	26	27.10.1979	12.08.1980	1	290
709	1	45	2	1	3	0	0	22,1	28,5	30,3	12.11.1979	07.04.1980	1	147
711	1	39	1	1	3	.	1	27,6	20,7	31,2	12.11.1979	12.03.1980	1	121
712	2	51	2	0	2	0	0	35,2	41,8	57,7	25.11.1979	05.06.1980	2	193
713	2	67	1	1	0	0	1	72,3	85,5	81	13.12.1979	27.06.1980	1	197
714	1	43	1	1	3	0	0	24,4	23,6	25,2	10.01.1980	15.07.1980	1	187
715	2	37	1	0	0	1	1	15,7	.	.	.	31.12.1980	.	.
716	1	62	1	1	3	1	1	37,3	38,4	35,2	23.01.1980	03.05.1980	1	101
717	2	53	2	1	1	0	0	25,2	24,4	41,8	25.01.1980	05.11.1980	1	285
718	2	72	3	0	0	1	1	42,9	49,2	46,6	03.02.1980	01.12.1980	0	302
719	2	75	3	0	0	.	.	57,7	72,3	105	07.03.1980	01.02.1981	0	331
720	1	47	3	1	2	1	0	22,1	28,5	28,5	08.03.1980	01.12.1980	0	268
721	2	58	3	0	1	1	1	47,9	59,2	89,5	24.03.1980	15.07.1980	1	113
801	2	48	3	1	2	0	0	23,6	20,7	32,2	11.04.1980	14.11.1980	1	217
802	1	51	3	1	3	0	1	32,2	39,5	51,9	12.04.1980	22.10.1980	1	193
803	1	81	2	1	3	0	0	74,1	68,8	83,5	22.04.1980	27.11.1980	1	219
804	1	35	1	0	1	1	1	19,3	20	18,7	07.05.1980	18.08.1980	2	103
805	2	64	3	0	1	1	1	46,6	62,3	70,5	14.05.1980	01.12.1980	0	201
806	1	55	3	0	1	1	0	22,8	22,8	22,8	15.05.1980	01.12.1980	0	200
807	2	44	1	0	3	1	1	26	24,4	16,3	03.06.1980	20.09.1980	1	109
808	2	59	1	1	1	0	0	36,2	31,2	38,4	18.06.1980	19.10.1980	1	123
901	1	41	1	1	3	0	0	22,1	20,7	23,6	19.06.1980	03.11.1980	1	137
902	2	61	2	0	0	0	1	54,7	60,7	70,5	22.07.1980	01.12.1980	0	132
903	2	55	2	0	2	1	1	24,4	23,6	38,4	23.07.1980	01.12.1980	0	131
904	2	49	1	1	3	1	1	20,7	16,3	17,4	01.08.1980	01.12.1980	0	122

Muuttujaluettelo

Tilastokäsittelyä varten kannattaa tutkimustiedostosta/tietokannasta laatia riittävän selkeä kuvaus (suoraan tiedostoon ja/tai erilliseksi dokumentiksi) muuttujien sisällöstä, mitta-asteikosta ja tiedon esittämistavasta. Tämä myös estää tiedon muuttumisen käyttökelvottomaksi pitkään kestävä tutkimuksen aikana unohduksen seurauksena. Kuvaukseen kannattaa sisällyttää muuttujista ainakin seuraavat asiat:

- **Muuttujan nimilyhenne;** Nimen tulee olla yksikäsitteinen; ts. kahdella muuttujalla ei voi olla samaa nimeä. Perinteisesti monet tilasto-ohjelmat ovat sallineet korkeintaan 8-merkkisiä nimiä. Uusimmissa ohjelmaversioissa tämä rajoitus on poistettu, SPSS sallii 64-merkkisiä nimiä. Kovin pitkien nimien käyttö ei kuitenkaan ole suositeltavaa; se saattaa helposti aiheuttaa sekaannuksia ja hankaloittaa mm. muunnosten tekoa. Muuttujanimen pitää alkaa kirjaimella. Muuttujanimessä voi esiintyä sekä isoja että pieniä kirjaimia. Erikoismerkkejä @ # _ ja \$ voi käyttää. Sen sijaan tyhjä, väliviiva, kauttaviiva ja kertomerkki ovat kiellettyjä. Piste on sallittu, paitsi muuttujanimen lopussa. Ohjelmointikielissä on

tiettyjä varattuja nimiä, joita ei muuttujaniminä voi käyttää; esim. ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO WITH.

- **Muuttujan tyyppi;** Ennen tietojen tallentamista muuttujan tyyppi ("Number", "Date" jne.) pitää määrittellä.
- **Muuttujan varaama tila** (=kentän pituus); Numeerisen muuttujan kohdalla kentän pituudessa pitää ottaa huomioon mahdollisen desimaalipisteen ja etumerkin vaatima tila. Myös desimaalien määrä pitää määrittellä.
- **Muuttujan otsikko;** Useimmissa ohjelmistoissa on kullekin muuttujilla varattu nimen ohella myös otsikkokenttä ("Label"), johon voi tallentaa merkkimuotoista tekstiä tulostusasun luettavuuden parantamiseksi. (SPSS: "Variable View" ► "Label")
- **Muuttujan koodiarvot;** Luokitelluille ja usein myös järjestysasteikollisille muuttujille kannattaa tiedostoon määrittellä koodiarvot ja niiden merkitys tulosteiden luettavuuden parantamiseksi. Koodiarvot ovat tutkijan vapaasti valittavissa. Usein kuitenkin jonkin oireen tai ominaisuuden puuttuminen on tilastokäsittelyn kannalta hyvä koodata 0:ksi ja sen olemassaolo 1:ksi. (SPSS: "Variable View" ► "Values")
- **Puuttuvien tietojen merkintä;** Esim. tietokantaohjelmissa dBase merkki * (tai ** jne.) tulkitaan puuttuvaksi tiedoksi. Tyhjä tulkitaan kiinteässä tallennusmuodossa ("fixed format") puuttuvaksi tiedoksi, mutta ei vapaassa tallennusmuodossa ("free format"). Ohjelmassa dBase tyhjä onkin ongelmallinen; se muuntuu helposti eri käsittelyvaiheissa nolaksi aiheuttaen sekaannuksia. Yleensä ohjelmissa tyhjäksi jätetty kenttä tulkitaan puuttuvaksi tiedoksi (esim. Excel, SPSS) ellei toisin määrittellä. Puuttuvien tietojen koodi voidaan kuitenkin määrittellä tiedostoon kunkin muuttujan osalta erikseen (SPSS: "Variable View" ► "Missing"), esim. valitsemalla puuttuvan tiedon koodiksi jokin muuttujan vaihtelualan ulkopuolelle jäävä arvo.
- **Muuttujan asteikko;** Tilastokäsittelyssä käytettävä asteikkotyyppi (SPSS: "Variable View" ► "Measure")
- **Erytishuomautukset;** Mittaustarkkuus, pyöristys, typistys, jne.

Lisätietoja käsittelyä varten:

- **Muuttujien luokitukset;** Jatkuvat muuttujat joudutaan usein aineistoa kuvattaessa ja joissakin malleissa estimointisyistä luokittelemaan. Tästä tosin saattaa aiheutua jonkin verran informaation menetystä. Jatkuvien muuttujien tiedot kannattaa kuitenkin tallentaa luokittelemattomina, sillä luokitukset voidaan suorittaa tilasto -ohjelmalla. Yleensä joudutaan kokeilemaan monia eri luokituksia.
- **Muuttujien muunnokset;** Erotus - ja summamuuttujat, indeksit ym. johdetut muuttujat
- **Ehdot loogisuustarkistuksia varten;** Jo tietojen tallennusvaiheessa kannattaa tietokantaohjelmaan rakentaa joukko ohjelmointikielen käskyjä, jonka avulla

voidaan tarkistaa loogisten virheiden olemassaoloa. Ennen tilastokäsittelyä kannattaa tarkistaa minimi- ja maksimit ja loogiset mahdottomuudet.

Asteikot

Muuttujan mitta -asteikko ratkaisee, minkälaisia deskriptiivisiä keinoja (tilastollisia tunnuslukuja, graafisia kuvaajia ja taulukoita), tilastollisia analyysimenetelmiä ja testejä on tarkoituksenmukaista käyttää.

- **Välimatka -asteikko** ("Interval scale") Asteikko, missä minkä tahansa kahden numeerisen arvon erotuksella on kvantitatiivisesti sama merkitys missä tahansa kohdassa skaalaa, esim. lämpötila.
- **Suhdeasteikko** ("Ratio scale") Välimatka -asteikko, missä 0-arvo merkitsee ominaisuuden puuttumista, esim. alkoholin kulutus, riskisuhde.
- **Järjestysasteikko** ("Ordinal scale") Kolmi- tai useampiluokkainen asteikko, missä luokkien suhteellinen sijainti toisiinsa nähden muodostaa luonnollisen järjestyksen, esim. oireen voimakkuus: ei oiretta, lievä, kohtalainen, vaikea.
- **Laatueroasteikko** ("Nominal scale", "Multichotomy") Kolmi- tai useampiluokkainen asteikko, missä luokkien välillä ei oleteta olevan mitään luonnollista järjestystä tutkittavassa asiayhteydessä, esim. siviilisääty, tautiluokitus.
- **Kaksiluokkainen asteikko** ("Dichotomy") On erikoistapaus järjestysasteikosta. Siinä on kaksi luokkaa, joiden järjestys keskenään on sopimuskysymys, esim. sukupuoli, taudin olemassaolo.

Monissa tilastollisissa malleissa välimatka - tai suhdeasteikollinen muuttuja joudutaan luokittelemaan järjestysasteikolliseksi, jotta mallin parametrien arviointi olisi mahdollista ja tulokset tulkittavissa.

Laatueroasteikollista muuttujaa voidaan testausmielessä tarkastella tilastollisissa malleissa järjestysasteikollisena, kun halutaan tutkia, muodostavatko sen luokat luonnollisen järjestyksen (trenditestit).

Laatueroasteikollisesta muuttujasta voidaan ja usein joudutaan muodostamaan tilastollisiin malleihin useita kaksiarvoisia **ilmais-** tai **mallintamismuuttujia**, ("indicator"-, "dummy"- tai "design" - muuttujat), joilla yhdessä on sama tietosisältö kuin alkuperäisellä muuttujalla.

Välimatka - tai suhdeasteikollista muuttujaa (esim. ikä) voidaan tilastollisissa malleissa käsitellä kuten laatueroasteikollista muuttujaa, kun halutaan esimerkiksi testata muuttujan mahdollisia ei-monotonisia yhteyksiä mallin muihin muuttujiin. Tällöin on kuitenkin syytä huomata, että **katkaisukohdilla** ("cut-off points") on suuri merkitys tulosten kannalta. Huono katkaisukohtien valinta voi peittää alleen tilastollisen yhteyden.

Tiedot kannattaa alkuperäiseen tiedostoon kerätä mahdollisimman tarkasti, jotta tilastokäsittelyssä niitä voidaan sopivasti yhdistellä tutkimushypoteesin edellyttämällä tavalla, esimerkiksi laboratoriomittausarvot kannattaa ottaa tarkkoina eikä pelkästään kaksiarvoisina (normaali, epänormaali), samoin ikä kannattaa ottaa ainakin vuoden tarkkuudella.

Puuttuvat tiedot

Puuttuvat tiedot muodostavat monissa kliinisissä tutkimuksissa ison ongelman. Niitä syntyy olosuhteiden pakosta ja yleensä niille pitäisi jotain tehdä ennen tilastokäsittelyä. Puuttuvien tietojen seurauksena tutkimuksen voima saattaa heiketä ja estimaattien tarkkuus voi huonontua. Näin käy erityisesti silloin, jos tilastokäsittelyyn otetaan vain ns. ”täydelliset tapaukset”, eli potilaat, joilta on saatu kaikki tiedot (”complete-case analysis”, ”available-case analysis”). Mikäli puuttuvat tiedot eivät tällaisessa tilanteessa jakaannu satunnaisesti aineiston sisällä, saattaa tuloksiin lisäksi syntyä harhaa.

Monet tutkijat pitävät karkeana nyrkkisääntönä, että enempää kuin 5 % muuttujan arvoista ei saisi puuttua. Monissa kliinisissä tutkimuksissa suoritetaan potilaasta toistomittauksia ja Tällöin puuttuvat tiedot ovat erityisen ongelmallisia.

Useimpiin tilastollisiin ohjelmistoihin sisältyy puuttuvien tietojen korvaamismahdollisuus. Siihen tarkoitukseen on myös erikoisohjelmia, kuten esim. **SOLAS** (Solas For Missing Data Analysis 3.0. Statistical Solutions. <http://www.statsol.ie/index.php?pageID=5>)

SPSS:ssä puuttuvien tietojen korvaaminen voidaan tehdä valikoista **”Transform”** ► **”Replace missing values”** valitsemalla kohdasta **”Method”** tarkoituksenmukaisin menetelmä.

Tavallisimpia puuttuvien tietojen korvaamismenetelyjä (”imputation”) ovat:

- **Koko aineiston keskiarvolla** tai **-mediaanilla** korvaaminen. SPSS ”Series mean”
- **Ryhmäkeskiarvolla tai -mediaanilla korvaaminen.** SPSS ”Mean of nearby points” tai ”Median of nearby points”.
- **Viivallinen interpolaatio.** Esim. Potilaan joltain tutkimuskäynniltä puuttuvaksi jäänyt arvo korvataan edeltävän ja sitä seuraavan käynnin arvojen perusteella viivallisesti interpoloidulla arvolla. SPSS ”linear interpolation”.
- **Rergressiotekniikkaan** perustuva menetelmä. Käytettävissä olevien havaintopisteiden avulla ennustetaan puuttuva arvo. Eräs tämän tekniikan modifikaatio sisältyy SPSS:ään
- **Viimeisen havaintoarvon siirtäminen eteenpäin (LVCF- menetelmä** ”last value carried forward”). Käytetään etenkin toistomittausasetelmissä. Tämän menetelmän kanssa tulisi kuitenkin olla erityisen varovainen, se voi johtaa pahasti harhaiseen lopputulokseen.
- **Lähimmällä kaltaistetulla havaintoarvolla korvaaminen** (”hot-deck imputation”). Käytetään esim. kyselytutkimuksissa siten, että jos jokin kysymyssarja käsittää 10 osiota ja henkilöltä A puuttuu vastaus osioon 4, niin etsitään sellaisen henkilön B vastaus, joka on vastannut mahdollisimman samantyyppisesti kuin henkilö A muihin 10 osiosta ja siirretään B:n vastaus A:n tietueeseen puuttuvaan osioon 4..

Mikäli puuttuvia tietoja korvataan, niin tutkimustuloksia raportoitaessa pitää aina ilmoittaa mitä menetelmää on käytetty ja kuinka paljon puuttuvia tietoja on korvattu.

Aineiston tallentaminen ja muokkaaminen

Suuret aineistot kannattaa tallentaa erillisillä tietokantaohjelmilla (Access, dBase, jne.), joista kulloinkin tilastokäsittelyyn tarvittavat tiedostot voidaan muodostaa. Pienet tiedostot voidaan tallentaa suoraan tilasto-ohjelmistojen Data Editoria käyttäen. Myös taulukkolaskimiin tallennetut tiedot siirtyvät helposti tilasto-ohjelmiin, jos ne on tallennettu yllä olevan suorakaidetiedoston muotoon, joka on tavallisin tiedostoformaatti. Excelissä muuttujanimet kannattaa tallentaa ensimmäiselle riville, niin silloin ne siirtyvät tilasto-ohjelmiin. Mikäli kyseessä on numeerinen muuttuja, niin muuttujan arvot pitää muistaa määrittellä Excelissä kohdassa "Format Cells" numeeriseen muotoon ja määrittellä myös desimaalien määrä ennen siirtoa tilasto-ohjelmalle.

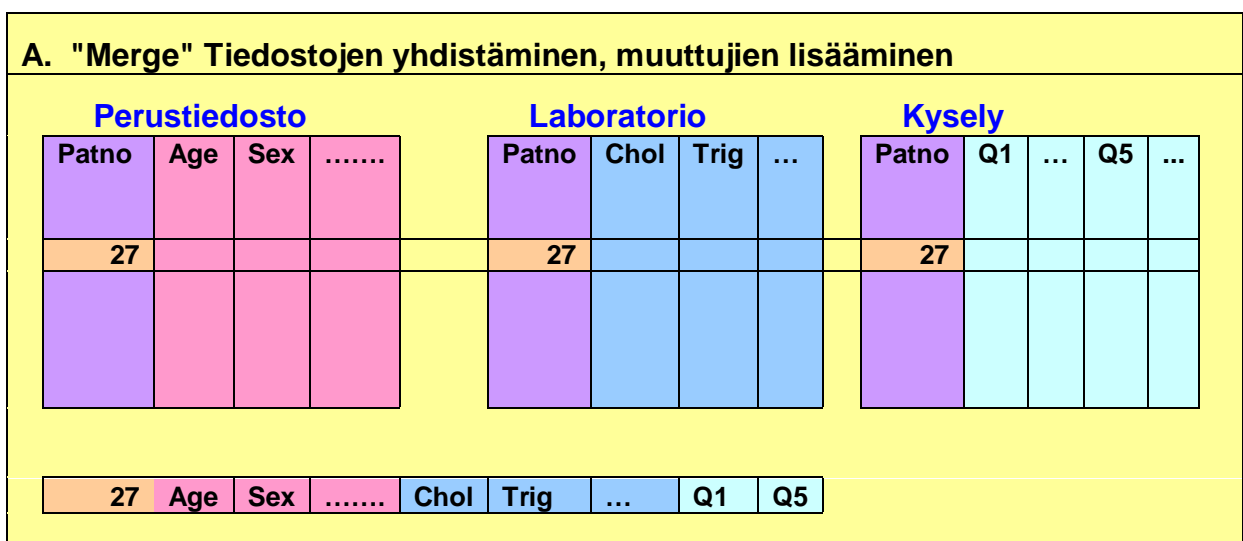
Aikaviiveiden laskemista varten tietokantaan kannattaa sisällyttää tarkat päivämäärät "date"-muodossa. Suositeltavia tallennusmuotoja ovat: **dd.mm.yyyy** tai **yyyy.mm.dd**. Näistä jälkimmäinen on lajitteluissa helpompi. Sellainen muoto, missä kuukausi on esitetty kolmella kirjaimella yleensä tuottaa ongelmia tiedostokonversioissa mm. Excelin ja SPSS:n välillä.

Huom.

Kun SPSS:llä lasketaan "date"-muodossa olevien päivämäärien erotuksia, lopputulos on ilmoitettu sekunteina. Ne saadaan esim. vuosiksi jakamalla luvulla 31536000 (=60*60*24*365). SPSS:ssä on lisäksi monipuolinen apuohjelmisto päivämäärätietojen aikatietojen käsittelyyn kohdassa "**Transform**" ► "**Date and Time Wisard**"

Osa-aineistojen yhdistäminen

Tietokantaohjelmissa on erittäin monipuoliset mahdollisuudet tiedostojen käsittelyyn, erilaisiin loogisuustarkistuksiin ja tietojen yhdistelyyn. Nykyisten tilasto-ohjelmistojenkin Data Editorit tarjoavat kuitenkin usein riittävät mahdollisuudet tiedostojen esikäsitteilyyn, ja tiedostojen yhdistelyyn. Tällöin on tärkeää, että kuhunkin tiedostoon on tallennettu samassa muodossa oleva havaintotunniste, esim. potilasnumero, jota voidaan linkkausavaimena. SPSS:ssä aineistojen yhdistämisiä voi tehdä kohdassa "**Data**" ► "**Merge Files**" ► "**Add Variables**" tai ► "**Add Cases**"



B. "Merge" Tiedostojen yhdistäminen, havaintojen lisääminen

Osite 1

Patno	Strata	Var2
	1		
	1		
	1		
	1		

Osite 2

Patno	Strata	Var2
	2		
	2		
	2		
	2		

Osite 3

Patno	Strata	Var2
	3		
	3		
	3		

Patno	Strata	Var2
	1		
	1		
	1		
	1		
	2		
	2		
	2		
	2		
	3		
	3		
	3		

Poikkeavat havaintoarvot

Biologisessa materiaalissa esiintyy aina poikkeuksellisen pieniä tai suuria muuttujan arvoja. Tärkeä vaihe aineiston esitarkastelussa on pyrkiä paikallistamaan tällaiset arvot, kenellä henkilöllä niitä esiintyy ja missä muuttujissa. Poikkeavia havaintoarvoja esiintyy seuraavista syistä johtuen:

- Virheet näytteidenotossa, mittaamisessa tai tietojen rekisteröinnissä. Nämä poikkeavat havaintoarvot pyritään yleensä korvaamaan oikealla arvolla, mikäli se suinkin on mahdollista. Muussa tapauksessa havaintoarvo tai koko havaintoyksikkö saatetaan joutua sulkemaan pois analyysistä.
- Äärimmäiset biologisen, psykologisen tai ympäristövariaation aiheuttamat arvot. Yleensä analyysit ja arvioinnit joudutaan tällaisissa tapauksissa suorittamaan sekä näiden poikkeavien arvojen kanssa että ilman niitä. Lisäksi on olemassa menetelmiä arvioida poikkeavien arvojen vaikutusta tuloksiin.
- Otantaongelmat tai virheet tutkimusten sisäänotto - tai poissuljentakriteereissä saattavat tuoda mukaan aineistoon henkilöitä, jotka poikkeavat muista tutkittavista selvästi. Koska nämä henkilöt eivät kuulu tutkimuksen perusjoukkoon, lienee tavallisinta, että tapaukset vain raportoidaan ja suljetaan tilastokäsittelyn ulkopuolelle.

Esim.

Aineiston **Trial** muuttujalle **"Imag-S"** mahdolliset poikkeavat arvot saadaan **SPSS:llä** valikoista: **"Analyze" ► "Descriptive statistics" ► "Explore"** asettamalla lokeroon **"Dependent List"** **"Imag-S"** ja lokeroon **"Label cases by"** potilaan tunnuskoodeksi **"Patno"** ja klikkaamalla **"Outliers"** kohdasta **"Statistics"**.

The screenshot shows the SPSS Explore dialog box. On the left, a list of variables includes Gender, Age, Therapy, Smoking status [S...], Alcohol, Hepatomegaly 1 [...], Hepatomegaly 2 [...], Imag_1, and Imagn_2. The 'Dependent List' contains 'Imag_s'. The 'Factor List' is empty. The 'Label Cases by' field contains 'Patno'. At the bottom, the 'Statistics' radio button is selected. On the right, the 'Explore: Statistics' sub-dialog shows 'Descriptives' checked with a 'Confidence Interval for Mean' of 95%. 'M-estimators' and 'Percentiles' are unchecked, while 'Outliers' is checked.

Tulos:**Extreme Values**

			Case Number	Patno	Value
Imag_s	Highest	1	41	803	74,1
		2	30	713	72,3
		3	36	719	57,7
		4	25	707	56,2
		5	3	503	54,7 ^a
	Lowest	1	15	609	14,6
		2	16	611	15,1
		3	1	501	15,1
		4	32	715	15,7
		5	26	708	18,1

a. Only a partial list of cases with the value 54,7 are shown in the table of upper extremes.

Keskiluvut

Jakauman sijainnin kuvaamisessa käytetään tavallisimmin seuraavia keskilukuja:

Aritmeettinen keskiarvo ”arithmetic mean” on tavallisin välimatka- ja suhteasteikollisen muuttujan jakauman keskeisyyden mitoista. Se on herkkä poikkeaville havaintoarvoille.

Mediaani ”median” on jakauman sijaintia, keskeisyyttä kuvaava suure, jonka alapuolella on puolet ja yläpuolella on puolet havaintoarvoista, **50 % -piste**. Mediaani ei ole herkkä poikkeaville havaintoarvoille kuten aritmeettinen keskiarvo.

Geometrinen keskiarvo ”geometric mean” on keskiluku, jota käytetään vinojen

jakaumien yhteydessä. Se on logaritimuunnettujen havaintoarvojen aritmeettisen keskiarvon antilogaritmi.

Moodi ”mode” on jakauman sijaintia kuvastava suure. Se on havaintoarvojoukossa useimmin esiintyvä arvo. Mitta on harvoin käyttökelpoinen lääketieteellisissä tutkimuksissa. **Harmoninen keskiarvo** on käänteislukujen aritmeettinen keskiarvo ja sitä käytetään mm. fysiikan sovelluksissa. **Viritettyjä keskiarvoja** (”trimmed means”) käytetään, kun halutaan eliminoida poikkeavien havaintoarvojen vaikutus keskiarvoon. Niillä on merkitystä lähinnä deskriptiivisessä mielessä. SPSS:ssä ne ovat nimellä ”M-estimators”

SPSS -ohjelmasta nämä keskiluvut löytyvät valikoista **”Analyze”** ▶ **”Descriptive statistics”** ▶ **”Descriptives”** tai **”Explore”** kuten myös **”Analyze”** ▶ **”Compare means”** ▶ **”Means”**.

Esim.

Tiedosto **Trial**. Muuttujan **”Age”** keskiluvut hoitoryhmittäin. Kohdassa **”Means”** lokeroon **”Dependent List”** asetetaan **”Age”** ja lokeroon **”Independent List”** luokitteleva muuttuja **”Therapy”**.

Tulos:

Age				
Therapy	Mean	Median	Geometric Mean	N
Group P	46,63	43,50	45,12	16
Group Q	49,06	51,50	46,61	16
Group R	54,56	53,50	52,64	18
Total	50,26	50,50	48,19	50

Keskilukujen käyttökelpoisuutta eri asteikkotyyppien yhteydessä kuvaa seuraava taulukko:

	Asteikot			
Tunnusluku	Välimatka/ suhde	Järjestys	Laatuero	Dikotomia
Keskiarvo	+	(+)	-	-
Mediaani	+	+	-	-
Moodi	+	+	+	+

Selitykset: “+” = “käyttökelpoinen”, “(+)” = “varauksellisesti käyttökelpoinen”, “-” = “ei sopiva”

Varauksellisesti käyttökelpoinen tarkoittaa, että tunnusluku on mielekäs deskriptiivisesti vain, mikäli muuttujien arvot ovat yhtä etäällä toisistaan **biologisessa mielessä**.

Keskilukujen valinta käyttö riippuu **käyttötarkoituksesta** eikä jakauman muodosta. Esimerkiksi alkoholin tuottajaa saattaa kiinnostaa alkoholin vuosittainen keskikulutus tietyssä väestössä, jolloin luonnollinen valinta on keskiarvo. Alkoholitutkija saattaa sen sijaan olla kiinnostuneempi **mediaanista**, eli kuinka paljon alkoholia puolet kyseisestä väestöstä käyttää. Jos jakauma on symmetrinen, niin keskiarvo ja mediaani ovat samat.

Kun julkaistaan keskiarvoja, kannattaa pitää mielessä mikä on alkuperäisen datan numeerinen tarkkuus. Keskiarvoihin kannattaa ottaa vain yksi ylimääräinen desimaali.

Useimmissa analyyseissa dikotominen ja järjestysasteikollinen muuttuja toimivat laskennallisesti kuten välimatka -/suhdeasteikollinen muuttuja, mutta viimeksi mainituille kehitetyt testit toimivat vain likimääräisesti. Dikotomisten muuttujien keskiarvot eivät kuitenkaan ole **deskriptiivisesti** mielekkäitä, koska keskiarvo on riippuvainen valituista koodiarvoista. Prosenttiosuuskien ilmaiseminen on mielekkäämpää. Esim. jos muuttuja on dikotominen ja koodattu (0,1), niin keskiarvo tarkoittaa samaa kuin arvojen 1 prosenttiosuus.

Mikäli jakauma on **positiivisesti vino** (eli pidempi häntä oikealle), niin keskiluvuksi kannattaa usein valita logaritmuunnokseen perustuva **geometrinen keskiarvo**, eli **antilogaritmi logaritmien arvojen aritmeettisesta keskiarvosta**. Tällöin nolla-arvot pitää korvata mittaustarkkuuden alarajalla. Geometrinen keskiarvoa käytetään yleisesti mm. sero-bakteriologisissa tutkimuksissa. Sen käyttö on varsin luonnollista esim. tiitterien yhteydessä.

Mikäli jakauma on **negatiivisesti vino** (eli pidempi häntä vasemmalle), voidaan käyttää muunnosta eksponenttimuunnosta **exp(x)** ennen keskiarvon laskemista. Jos jakaumassa on **hännät molempiin suuntiin**, voidaan käyttää muunnosta **1/x**. Tällöin keskeisyyttä mittaa **harmoninen keskiarvo**, eli käänteisarvojen aritmeettinen keskiarvo.

Mikäli aineistossa muuttujan arvoina esiintyy poikkeuksellisen pieniä tai suuria arvoja (“outliers”), voidaan käyttää **viritettyjä** (“trimmed”) keskiarvon estimaatteja. Ne jättävät jakauman ääripäistä tietyn määrän havaintoarvoja pois.

Toinen tapa hallita poikkeavia havaintoarvojen ylikorostuvaa vaikutusta on käyttää **havaintoarvojen painottamista** siten, että estimaatin lähellä olevat arvot saavat suuremman painon kuin ääripäissä olevat arvot. Esim. SPSS -paketissa tällaisia keskiarvon estimaatteja on neljää eri tyyppiä, jotka eroavat toisistaan havaintojen painotusmekanismin suhteen. Tilastollisten testien kannalta nämä viritetyt tai painotetut keskiarvot eivät ole kovin käyttökelpoisia.

Hajonnan mitat

Numeerinen tapa kvantifioida tutkittavissa suureissa esiintyvää vaihtelua ("variability") eli hajontaa ("dispersion") on käyttää seuraavia mittalukuja:

Keskihajonta, standardipoikkeama "standard deviation", SD, on mitta, joka kuvaa tutkittavan muuttujan havaintoarvojen jakautumista tietyssä tutkimusaineistossa. Se soveltuu parhaiten hajonnan mitaksi symmetrisille jakaumille. Se on herkkä poikkeaville havaintoarvoille.

Vaihteluväli "range" on muuttujan suurimman ja pienimmän arvon erotus. Poikkeavien havaintojen tapauksessa tämän 100 % -vaihteluvälin sijasta käytetään usein esimerkiksi 95 % -vaihteluväliä.

Prosenttipiste (QP %) "percentile" on jakaumien kuvaamisessa käytetty suure, joka jakaa aineiston tietyn suuruisiin prosenttiosuuksiin, esim. 95 % -piste merkitsee että 5 % jakauman arvoista on yhtä suuria tai sitä suurempia. Tavallisimmin käytettyjä prosenttipisteitä ovat: tertiilit, kvartiilit ja desiilit. 50 %:n piste on sama kuin mediaani.

Kvartiiliväli "interquartile range" (IQR) on karkea vaihtelun mitta, joka ei ole herkkä poikkeaville havaintoarvoille. Kvartiiliväli on ylemmän ja alemman kvartiilin välinen alue. Laatikko-janakuviossa se on sama kuin laatikko, joten se pitää sisällään 50 % havaintoarvoista. Kvartiiliväliä käytetään kuvaamaan suureiden vaihtelua silloin, kun standardipoikkeama ei siihen sovellu; esim. kun jakaumat ovat kovin vinoja.

Variaatiokerroin CV "coefficient of variation" on skaalasta riippumaton vaihtelun mitta. Se lasketaan jakamalla mittausvirheiden standardipoikkeama keskiarvolla.

Viitearvoväli "reference interval" on väli, joka ilmaisee ns. normaalivaihtelua esim. laboratoriotesteissä tai lasten kasvukäyrissä.

SPSS -ohjelmassa erilaiset hajontamitat saadaan valikoista "**Analyze**" ► "**Descriptive statistics**" ► "**Explore**"

Esim.

Tiedosto **Trial**. Muuttujan "**Age**" hajonnan mitat hoitoryhmittäin.
Kohdassa "**Means**" lokeroon "**Dependent List**" asetetaan "**Age**" ja lokeroon "**Factor List**" luokitteleva muuttuja "**Therapy**". Tulos:

Explore

Dependent List:
Age

Factor List:
Therapy

Label Cases by:

Display:
 Both Statistics Plots

Explore: Statistics

Descriptives
Confidence Interval for Mean: 95 %

M-estimators

Outliers

Percentiles

		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
Age	Group P	16	100,0%	0	,0%	16	100,0%
	Group Q	16	100,0%	0	,0%	16	100,0%
	Group R	18	100,0%	0	,0%	18	100,0%

Descriptives					
Therapy			Statistic	Std. Error	
Age	Group P	Mean	46,63	3,055	
		95% Confidence Interval for Mean	Lower Bound	40,11	
			Upper Bound	53,14	
		5% Trimmed Mean	46,53		
		Median	43,50		
		Variance	149,317		
		Std. Deviation	12,220		
		Minimum	28		
		Maximum	67		
		Range	39		
		Interquartile Range	21		
		Skewness	,279	,564	
		Kurtosis	-1,015	1,091	
	Group Q	Mean	49,06	3,689	
95% Confidence Interval Lower Bound		41,20			

		for Mean	Upper Bound	56,93					
		5% Trimmed Mean		49,01					
		Median		51,50					
		Variance		217,796					
		Std. Deviation		14,758					
		Minimum		18					
		Maximum		81					
		Range		63					
		Interquartile Range		20					
		Skewness		-,098	,564				
		Kurtosis		,994	1,091				
		Group R	Mean		54,56	3,271			
			95% Confidence Interval for Mean	Lower Bound	47,65				
	Upper Bound			61,46					
	5% Trimmed Mean		55,12						
	Median		53,50						
	Variance		192,614						
	Std. Deviation		13,879						
	Minimum		24						
	Maximum		75						
	Range		51						
	Interquartile Range		19						
Skewness			-,377	,536					
Kurtosis			-,158	1,038					
			Percentiles						
Therapy			5	10	25	50	75	90	95
Weighted Average(Defini tion 1)	Age	Group P	28,00	29,40	37,50	43,50	58,25	65,60	.
		Group Q	18,00	25,70	39,00	51,50	59,00	67,00	.
		Group R	24,00	35,70	46,75	53,50	65,50	73,20	.
Tukey's Hinges	Age	Group P			38,00	43,50	57,50		
		Group Q			40,00	51,50	58,00		
		Group R			47,00	53,50	64,00		

Laskennallisesti standardipoikkeama on hyvä ja tavallisimmin käytetty hajonnan mittaluku. Merkintätapaa keskiarvo \pm SD tulisi välttää. Mikäli jakauma on normaali, tämä merkintä tarkoittaa noin 68 % vaihteluväliä. Väliä keskiarvo \pm 2·SD tai 95 % vaihteluväliä käytetään usein kuvaamaan niin sanottua normaalia vaihtelua, viitearvoväliä ("reference interval").

Riittävä numeerinen tarkkuus hajonnan ilmoittamisessa on korkeintaan kaksi ylimääräistä desimaalia verrattuna alkuperäisen datan tarkkuuteen.

Variaatiokerrointa (CV) käytetään mittausmenetelmän toistettavuuden mittana. Se lasketaan tekemällä kyseessä olevalla menetelmällä toistettuja mittauksia esim. eri näytteistä tai koehenkilöistä ja jakamalla mittausvirheiden SD kaikkien mittausten keskiarvolla. Mittausvirheiden SD voidaan laskea yksisuuntaisella varianssianalyysillä (tarkastellaan myöhemmin) siten, että ryhmittelevänä tekijänä on näyte tai koehenkilö ja kustakin/kultakin otetut toistomittaukset (voi olla vaihteleva määrä per näyte tai koehenkilö) edustavat ryhmien sisäistä vaihtelua, jonka määrää varianssianalyysissä mittaa keskineliösumma MS_w ("within group mean square"). Tämän suureen positiivinen neliöjuuri on mittausvirheiden SD. Variaatiokerroin on yksiköistä riippumaton ja sitä käytetään kun halutaan vertailla vaihtelua eri aineistoissa. Se ilmoitetaan tavallisesti prosenttilukuna.

Keskihajonta vs. keskivirhe

Käsitteet **keskihajonta** ("standard deviation", SD) ja **keskivirhe** ("standard error", SE) menevät tutkijoilta helposti sekaisin ja siten myös tieteellisissä artikkeleissa näitä käsitteitä käytetään usein väärin ja väärässä yhteydessä.

Keskihajonta standardipoikkeama, SD(x), on mitta, joka kuvaa tutkittavan muuttujan x havaintoarvojen jakautumista tietyssä tutkimusaineistossa. Soveltuu parhaiten hajonnan mitaksi symmetrisille jakaumille. Mitta on herkkä poikkeaville havaintoarvoille.

Keskihajonta mittaa suureen x hajontaa yksittäisessä tutkimusaineistossa, ts. kuinka paljon esim. potilaskohtaisissa havaintoarvoissa esiintyy vaihtelua. Yleensä julkaisujen ensimmäisessä taulukossa tutkijan käyttämä aineisto kuvataan tunnuslukujen, prosenttiosuuksien, keskilukujen ja hajontalukujen muodossa ja siinä yhteydessä pitäisi käyttää SD:tä, mikäli suureen x jakauma on edes suunnilleen symmetrinen. Muissa tapauksissa kannattaa turvautua edellä kuvattuihin parametrittomiin hajontamittoihin.

Keskivirhe "standard error", SE(·), mittaa tutkimusaineiston perusteella **lasketun**, minkä tahansa, **otossuureen** (esim. keskiarvon, prosenttiosuuden, matemaattisen mallin parametrin jne.) **luotettavuutta**, eli hajontaa kyseisen **suureen otosjakaumassa**; ts. kuinka paljon suure voisi vaihdella, jos tutkimus toteutettaisiin toistuvasti samalla aineistokoolla ja tutkimusasetelmalla useita kertoja.

Keskivirhe liittyy siten estimointiin, tuntemattomien suureiden arviointiin ja luottamusvälien laskentaan. Toisin kuin SD keskivirhe SE jää tutkijalle pakostakin jonkin verran abstraktiksi teoreettiseksi käsitteeksi, koska havaittua otosjakaumaa ei yleensä voida muodostaa muuta kuin tietokonesimulaatiota käyttäen. Tätä menettelyä tosin käytetäänkin nk. Monte Carlo- menetelmiin perustuvissa luottamusväleissä.

Parametriset keskivirheiden laskentakaavat perustuvat tunnettuihin teoreettisiin todennäköisyysjakaumiin. Esim. keskiarvon keskivirhe lasketaan kaavalla SD/\sqrt{n} , missä

n on havaintoarvojen lukumäärä (aineistokoko). Tämä laskentakaava perustuu siihen, että keskiarvon otosjakauma on likimain normaalin jo melko pienellä aineistokoollakin riippumatta suureen x jakaumasta. Tämä perustuu nk. todennäköisyyslaskennan keskeiseen raja-arvoväittämään.

Keskivirheitä tai etenkin niihin perustuvia luottamusvälejä tulisi käyttää erityisesti julkaisujen tulososassa, kun kuvataan tutkimuksen kannalta keskeisimpien suureiden luotettavuutta, ts. sitä kuinka paljon suureeseen sisältyy otannasta ja otoksen koosta johtuvaa vaihtelua, virheellisyyttä.

Jakauman vinous ja huipukkuus

Jakauman **vinous** ("skewness") paljastuu parhaiten graafisessa tarkastelussa. Tietokoneohjelmissa on olemassa useita jakauman vinouden eri piirteitä korostavia mittalukuja, joista mikään ei ole ihanteellinen. Tavallisimmin käytetty lienee kolmanteen momenttiin ja sen keskivirheeseen perustuva parametrinen vinouskerroin ja sen testi, joka sisältyy useimpiin tilastopaketteihin. Myös parametrin vinouskerroin on laskettavissa prosenttipisteiden perusteella, mutta sitä näkee harvemmin käytettävän. Parametrinen huipukkuuskerroin ("kurtosis") ja sen testi perustuu neljänteen momenttiin ja sen keskivirheeseen.

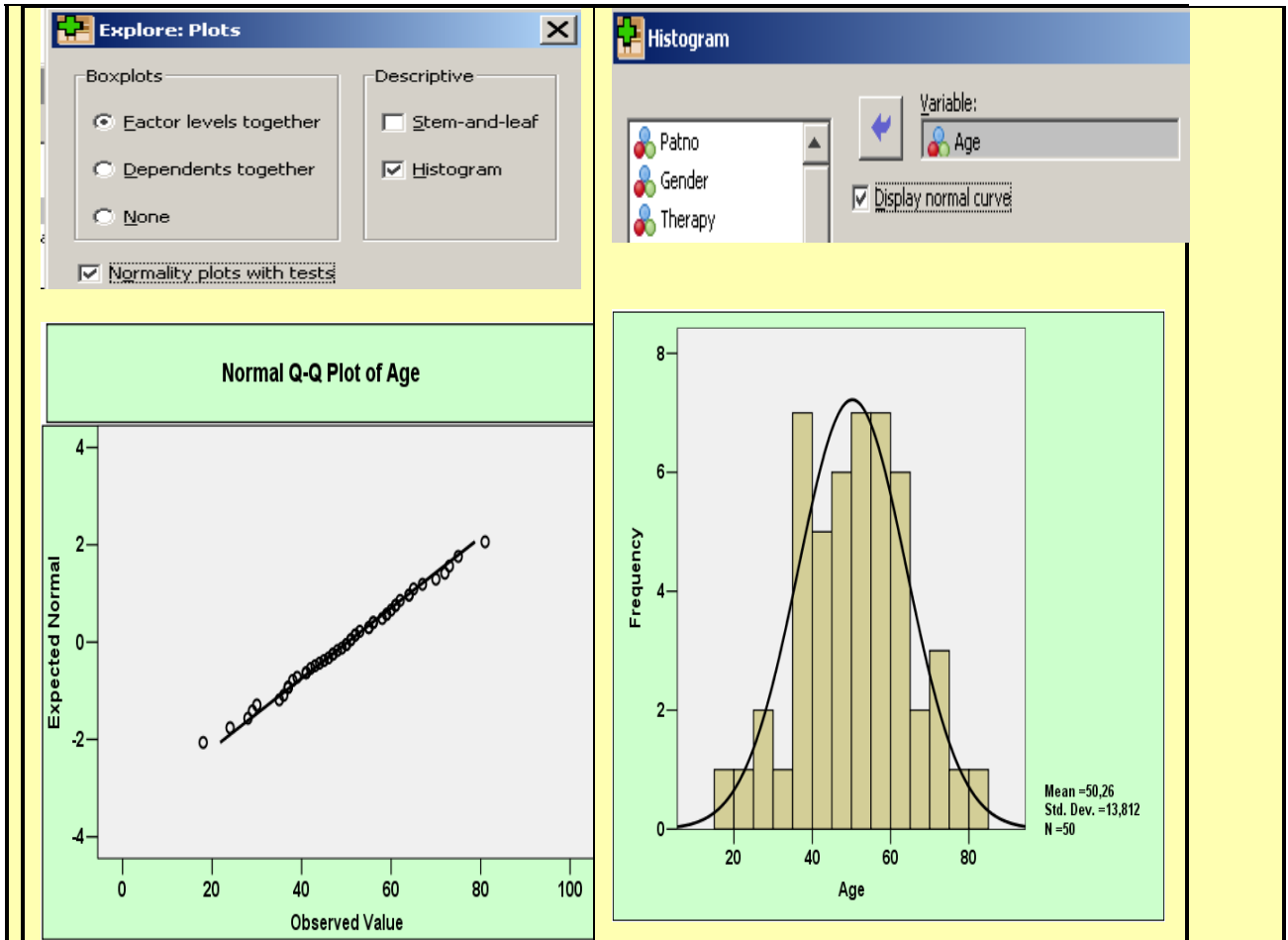
Jakauman normalisuus

Normaalijakauma, "**Normal distribution**", "**Gaussian distribution**", "**bell curve**", on yksi yleisimmistä tilastollisissa testeissä ja menetelmissä käytetyistä jatkuvista jakaumista. Normaalijakauma on keskiarvon suhteen symmetrinen jakauma siten, että noin 95 % havaintoarvoista on välillä (keskiarvo \pm 1.96·SD)

Monet nk.parametrisista testeistä ja menetelmistä perustuvat normaalisuusoletukseen. Jakauman normalisuutta voidaan testata monin eri tavoin. Graafisesti sitä voidaan tutkia esim. "**probit-plot**"- menetelmällä. Y -akselilla on tällöin kumulatiivisen normaalijakauman mukaiset odotusarvot ja x -akselilla havaitun kumulatiivisen jakauman arvot.

Esim.

Aineiston Trial muuttujalle "**Age**" saadaan **probit-plot**-kuvio valikoista: "**Analyze**" ► "**Descriptive statistics**" ► "**Explore**" laittamalla lokeroon "**Dependent List**" "**Age**" ja klikkaamalla kohdasta "**Plots**" "**Normality plots with tests**". Normaalijakaumakäyrällä varustettu histogrammi saadaan valikoista: "**Graphs**" ► "**Legacy Dialogs**" ► "**Histogram**" asettamalla lokeroon "**Variable**" muuttuja "**Age**" ja klikkaamalla kohta "**Display normal curve**". Kumpiakin kuvia on jonkin verran muotoiltu SPSS:n editorilla.



Tulkinta:

Normaalisuus on sitä parempi mitä paremmin pisteet sijoittuvat 45° kulmassa (x, y) - koordinaatistoon piirretylle suoralle viivalle, joten tässä tapauksessa silmämääräisesti arvioiden jakauma on melko hyvin normaalin.

Graafisten tarkasteluiden ohella on myös kvantitatiivisia keinoja arvioida normalisuutta. Voidaan käyttää esim. **Shapiro-Wilkin W tai Kolmogorov–Smirnovin** testejä, joita käsitellään kohdassa tilastolliset testit.

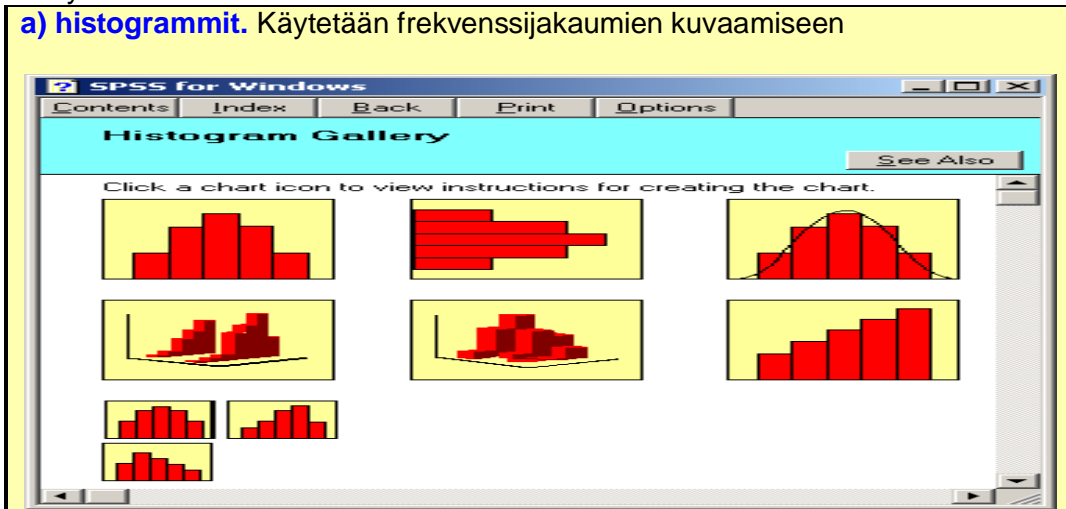
Graafiset kuvaajat

Yleisohjeita kuvien rakentamiseen:

- Kuvien ja kuvatekstien tulee olla itsensä selittäviä; kuvat pitää olla ymmärrettävissä artikkelin sisällöstä riippumatta
- Vaaka - ja pystyakselit tulee nimetä selkeästi ja myös yksiköt pitää olla näkyvissä
- Ei liian paljon tietoa yhteen kuvaan
- Kuvien tarkoitus on antaa nopeasti selkeä käsitys esitettävästä asiasta eikä esitellä detaljitietoutta
- Ei liian paljon numerotietoutta kuvaan
- Luottamusvälit suositeltavampia kuin keskivirheistä muodostuvat virhejanat ("error bars")

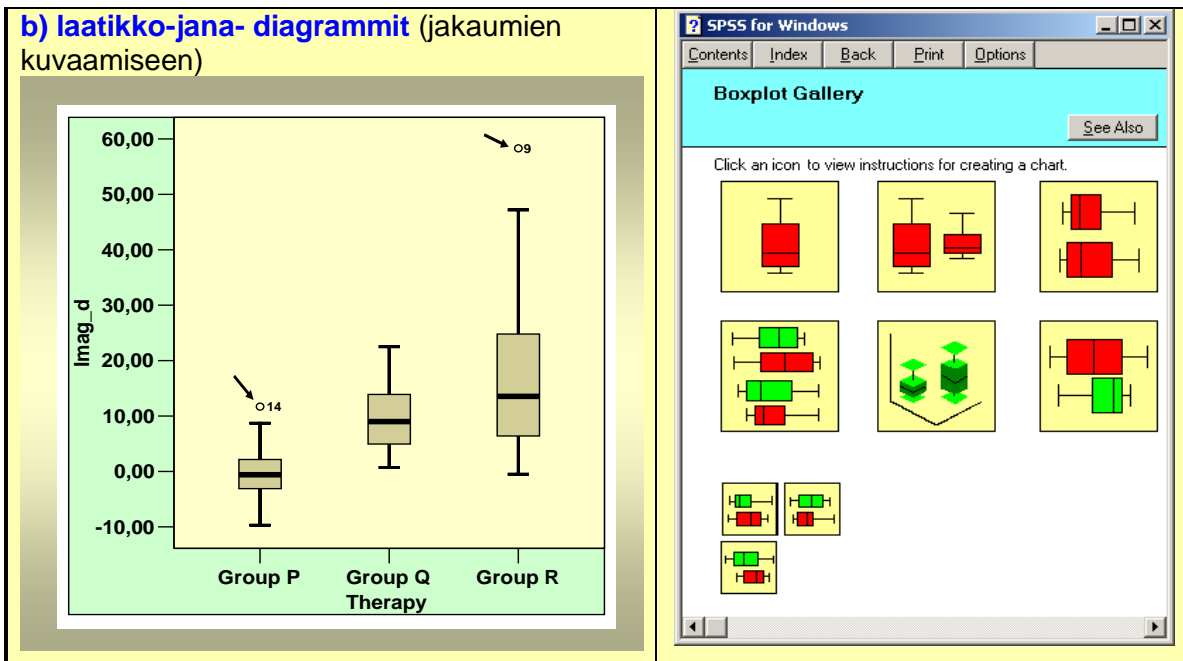
Korkealaatuisen julkaisugrafiikan teko edellyttää erityisohjelmistojen käyttöä. Kohtalaista grafiikkaa voi kuitenkin tehdä ihan tavallisilla taulukkolaskimilla, kuten Excel tai yleistilastopaketeilla, kuten SPSS, jonka graafisiin esitystapoihin kuuluvat mm. seuraavat esitysmuodot:

a) histogrammit. Käytetään frekvenssijakaumien kuvaamiseen



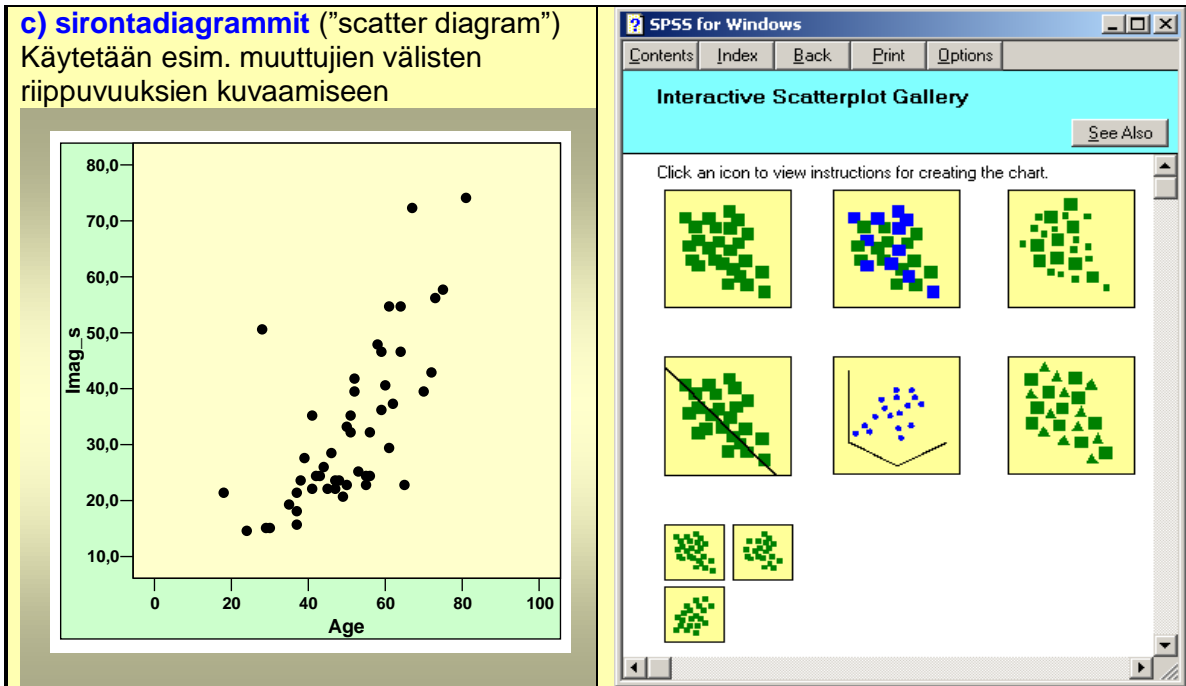
Kuvattaessa frekvenssijakaumaa histogrammille on vaihtoehtona **frekvenssimurtoviiva** ("frequency polygon"), jonka etuna histogrammiin verrattuna on se, että samaan kuvaan voidaan sijoittaa esimerkiksi useampien ryhmien frekvenssijakaumat vertailujen helpottamiseksi, ja myös jakauman muoto hahmottuu paremmin. Frekvenssijakaumissa kannattaa aina käyttää tasavälisiä luokituksia.

b) laatikko-jana- diagrammit (jakaumien kuvaamiseen)

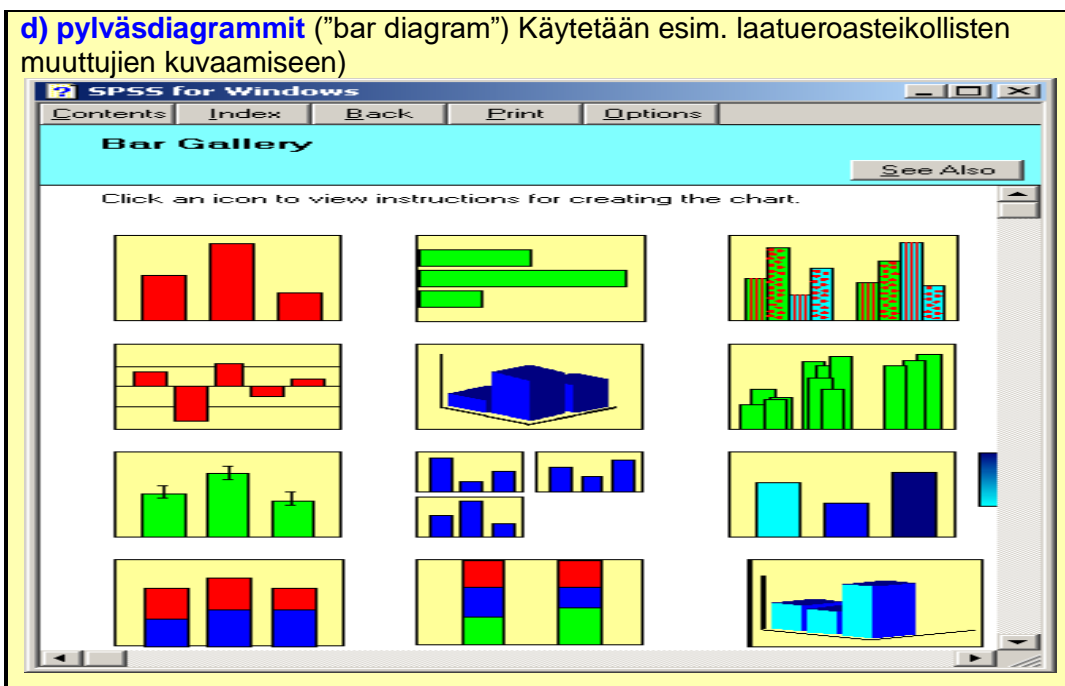


Laatikko-jana-diagrammilla voidaan tutkia esim. luokittelevan muuttujan ja jatkuvan lopputulosmuuttujan välistä yhteyttä. Laatikon alaraja on 25 %:n piste, keskellä oleva viiva on mediaani ja laatikon yläraja on 75 %:n piste. Viiksien tulkinta on joissakin tilanteissa hankalaa. Viiksissä alimmainen viiva kuvaa pienintä arvoa ja ylimmäinen suurinta arvoa, jota ohjelma ei ole tulkinut poikkeavan pieneksi tai suureksi arvoksi ("outlier"). Tulosteessa on kahden tyyppisiä poikkeavien havaintojen määrittelyitä:

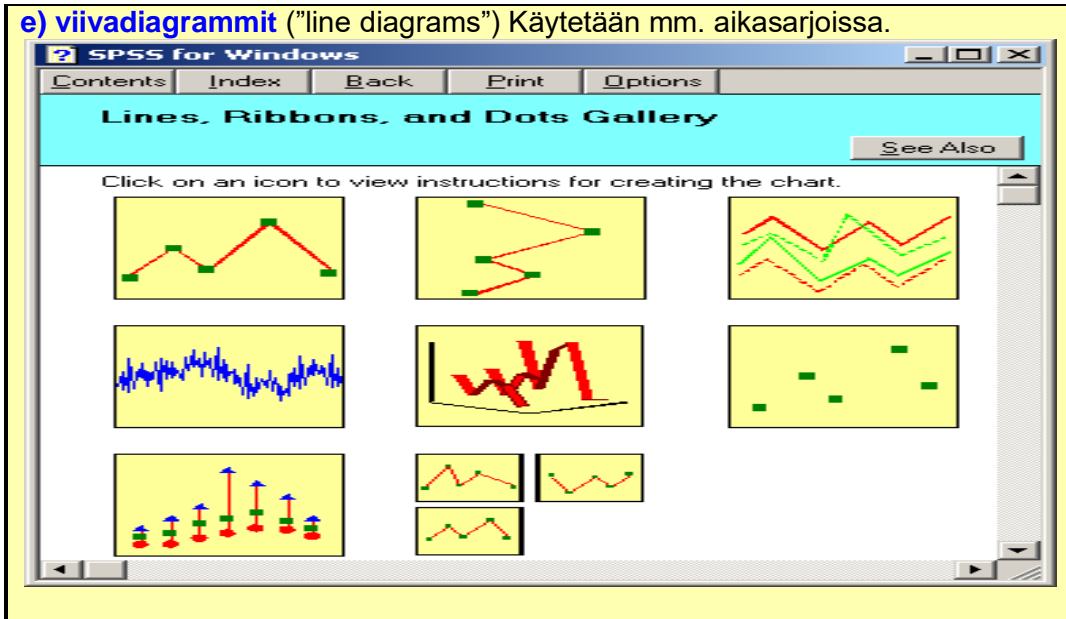
Symboli "o" tarkoittaa, että arvo poikkeaa laatikosta enemmän kuin 1,5 laatikon piteuden verran ja symboli "*" tarkoittaa poikkeamaa, joka on enemmän kuin 3 0 laatikon piteutta. Mikäli tällaisia poikkeavia havaintoja ei ole, niin viikset määrittelevät vaihteluvälin, eli pienimmän ja suurimman arvon.



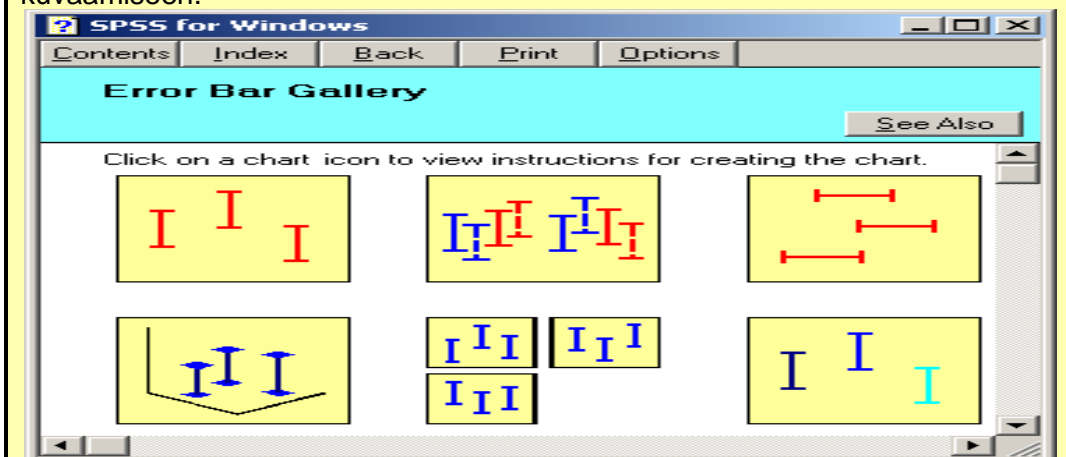
Kun kuvataan kahden muuttujan välisiä yhteyksiä sirontadiagrammeilla, niin kaikki havaintoaineiston pisteet tulisi esittää. Joissakin tilanteissa kuvantamissymboli voidaan korvata lukuarvolla, joka ilmaisee yhtenevien pisteiden määrän. Etenkin pienissä aineistoissa yksittäisten havaintopisteiden esittäminen hajontadiagrammeissa on suositeltavaa.



e) viivadiagrammit ("line diagrams") Käytetään mm. aikasarjoissa.



f) virhejanat "error bars" Käytetään esim. keskilukujen otantavirheen kuvaamiseen.



Viitteet

Stenman S. Tieteellisen artikkelin kuvitus I. Piirrostyypit ja niiden käyttö. Suomen lääkärilehti 1995; 50: 4: 379 -385.

Stenman S. Tieteellisen artikkelin kuvitus II. Mitä kuuluu kuvioon ja mitä tekstiin. Suomen lääkärilehti 1995; 50: 5: 499 -501.

Stenman S. Tieteellisen artikkelin kuvitus III. Tietokonegrafiikan sudenkuoppia. Suomen lääkärilehti 1995; 50: 6: 629 -634. Stenmanin artikkelisarja on suositeltavaa luettavaa.

Wulff HR, Sclichting P. MedStat 2.12 manual. Department of Medical Gastroenterology C Herlev University Hospital, Denmark, 1991.

Luottamusväliestimointi

Yleistä

Kaikille tärkeimmille tutkimustuloksille tulisi aina laskea luottamusvälit. Ne antavat käsityksen tulosten varmuudesta, ts. kuinka paljon ilmoitetut tulokset voisivat vaihdella otantavaihtelusta johtuen. Useimmat arvostetuimmista yleislehdistä, BMJ, JAMA, Lancet, New England, jne. korostavat luottamusvälien käyttöä, mm. kirjoitusohjeissaan. Erityisalojen lehdissä käytäntö vaihtelee. Esimerkiksi American Journal of Epidemiologyssä suorastaan edellytetään, että kaikille keskeisimmille tuloksille on laskettu luottamusvälit ja American Journal of Physiologyssä tehdyn selvityksen (Curran-Everett D ym.1995) mukaan vain **yhdessä** 370:ssä artikkelista oli annettu luottamusvälit!

Valitettavasti monissa tavallisimmissa tilastopaketeissakin on puutteita luottamusvälien laskennan suhteen. Tällöin joudutaan turvautumaan erityisohjelmistoihin. Altmanin, Machinin, Bryantin ja Gardnerin vuonna 2000 toimittaman luottamusvälien erityiskirjan oheistuotteena olevan CIA-ohjelman avulla voidaan kätevästi laskea luottamusvälit useimpiin käytännön tilanteisiin. Tiedot tähän Windows-ohjelmaan voidaan viedä joko näppäilemällä tai siirtää esimerkiksi Excelistä.

Luottamusvälien laskenta on osa estimointia.

<p>Estimoinnilla "estimation" tarkoitetaan tutkittavan ilmiön kuvaamiseen käytetyn mallin tai jakauman sisältämien tuntemattomien suureiden, parametrien, arviointia tutkimusaineiston (otoksen/otosten) perusteella. Estimointitapoja on kahta päätyyppiä:</p> <ul style="list-style-type: none"> • piste-estimointia ("point estimation") ja • luottamusväliestimointia ("confidence interval estimation")
<p>Piste-estimointi on tuntemattoman suureen, eli parametrin, (esim. jonkin taudin yleisyys) arviointi yhdellä lukuarvolla (estimaatilla), joka on laskettu havaintoaineistosta.</p>
<p>Otossuure "sample statistic" on otoksen havaintoarvojen perusteella laskettu suure, havaintoarvojen funktio.</p>
<p>Estimaattori "sample statistic" on otossuure, jolla tuntematonta parametria arvioidaan.</p>
<p>Parametri "parameter" on tuntematon suure, joka säätelee tutkittavan ilmiön kuvaamisessa käytettyä mallia, esim. jotain teoreettista jakaumaa. Arvioidaan tutkimusaineiston perusteella.</p>

Samaa parametria voidaan estimoida monella eri estimaattorilla tilanteesta riippuen. Esimerkiksi jakauman keskeisyyttä voidaan arvioida erilaisilla keskiluvuilla (tavallinen aritmeettinen keskiarvo, viritetyt keskiarvot jne.). **Estimaatti** on otoksen perusteella laskettu estimaattorin arvo. Se on aina aineistokohtainen. Estimaatin luotettavuutta, ts. siihen sisältyvää virhettä, mitataan keskivirheellä.

Keskivirhe "standard error" mittaa tutkimusaineiston perusteella lasketun minkä tahansa otossuureen hajontaa (luotettavuutta), eli kuinka paljon suure voisi vaihdella, jos tutkimus toteutettaisiin toistuvasti samalla aineistokoolla ja tutkimusasetelmalla.

Luottamusväli "confidence interval" (CI) on kahden otossuureen määrittelemä väli, johon arvioitavana oleva tuntematon perusjoukon suure (esim. todellinen hoitoero)

sijoittuu **luottamustason $100(1-\alpha)$ %** ilmoittamalla varmuudella. Varmuusväli olisi kuvaavampi suomenkielinen termi.

Tavallisimmin luottamustasoksi ("nominal confidence level") valitaan jokin arvoista: 0,90 tai **0,95** tai 0,99. Valinta ei perustu kuitenkaan mihinkään matemaattiseen tosiseikkaan. Kysymys on siitä, kuinka suuren riskin tutkija johtopäätöksilleen katsoo voivansa ottaa. Käytännössä eri menetelmillä laskettujen luottamusvälien todellinen luottamustaso, eli menetelmän **peittävyys**, poikkeaa jonkin verran suuntaan tai toiseen tästä nimellisestä tasosta.

Luottamusvälin nk. **frekventistinen** tulkinta tarkoittaa, että jos sama tutkimus toistettaisiin useita kertoja ja jokaisesta tutkimusaineistosta (otoksesta) laskettaisiin samalla laskentamenetelmällä kyseessä olevan suureen θ , esim. jonkin asian tai ominaisuuden suhteellisen osuuden, $100(1-\alpha)$ %:n luottamusväli. Tällöin $100(1-\alpha)$ %:ssa tapauksia tuntematon perusjoukon suhteellinen osuus sisältyy näihin väleihin.

Otantavaihtelusta johtuen eri toisto-otoksista laskettujen välien $[\hat{\theta}_1, \hat{\theta}_2]_n$ pituus vaihtelee otoksesta toiseen. Se pienenee, kun aineistokokoa n kasvaa, ts. varmuus lisääntyy. Vastaavasti, kun n pidetään vakiona ja luottamustasoa $1 - \alpha$ kasvatetaan, niin väli pitenee. Esim. 95 %:n väli on pidempi kuin 90 % väli, kun n pidetään kiinteänä.

Huom.

Luottamusvälin kaikki arvot eivät ole yhtä todennäköisiä! Luottamusvälin keskellä olevat arvot ovat todennäköisempiä kuin sen laidalla olevat. Todennäköisyydet määräytyvät tarkasteltavan suureen otosjakauman perusteella.

Frekventisten välien ohella on myös Bayesiläisiä välejä, jotka perustuvat **a priori** ja **a posteriori** todennäköisyystiheyksien yhdistelmään. **A priori** todennäköisyystiheys kuvastaa tutkijan subjektiivista ennakkokäsitystä arvioitavana olevan parametrin θ arvojen todennäköisyyksistä ennen tutkimusaineiston analysointia, ja **a posteriori** todennäköisyydet lasketaan tutkimusaineiston perusteella.

Luottamusvälit voidaan lisäksi jakaa

- **tarkkoihin** ("exact") ja
- **likimääräisiin** ("approximative")

väleihin riippuen siitä käytetäänkö välien laskennassa tarkasteltavan suureen tarkkaa vai likimääräistä todennäköisyysjakaumaa. Likimääräiset välit perustuvat tavallisesti kyseessä olevan suureen keskivirheeseen ja ovat yleensä optimistia, eli antavat todellista väliä kapeamman välin. Niiden antama peittävyys on alempi kuin nominaalinen taso, esim. 95 %. Vastaavasti eksaktit menetelmät antavat enemmän tai vähemmän konservatiivisen, eli liian leveän välin. Eri laskentamenetelmät eroavat toisistaan melkoisestikin keskimääräisen peittävyyden suhteen.

Esim.

Oraalinen magnesiumhoito hypertensioon. (Cappuccio et al. 1985).

Tutkimusaineisto: 17 potilasta. Tutkimusasetelma: Ristikkäistutkimus, Hoitojaksot olivat yhden kuukauden mittaisia. Annos: 15 mmol magnesiumia päivässä.

Tutkimuksella pyrittiin näyttämään, ettei hoidolla ole vaikutusta.

Tulos:

Ei eroa lumehoitoon verrattuna. Luottamusväli olisi paljastanut tutkimuksen alhaisen voiman, mutta sitä ei artikkelissa ilmoitettu!

Toisin kuin P-arvo luottamusväli antaa kliinisissä tutkimuksissa tietoa myös hoitoeron suuruudesta.

Esim.**Luottamusvälit vs. merkitsevyytestit. Sepelvaltimotaudin**

seurantatutkimus, Lipid Research Program, 1984. Aineistona 3806 miestä joilla oli korkeat lipidiarvot. Seuranta-aika oli keskimäärin 7,4 vuotta. Kyseessä oli lumehoitokontrolloitu kaksoissokkotutkimus. Ryhmät: A = Lumehoito, B = Kolestyramiinihoito.

Tulos:	Ryhmä A	Ryhmä B
Miesten määrä	1900	1906
CHD-tapaukset	187 (9,8 %)	155 (8,1 %)

Ilmoitettu tulos:

1. $P < 0,05$ (yksiipuoleinen testi)
2. CHD - riskin reduktio: $100 \cdot (9,8 - 8,1) / 9,8 = 17 \%$
3. 90 % luottamusväli: [3 %, 32 %]

Huom. riskisuhde $RR=0,83$

Suhteellinen riskin reduktio "relative risk reduction" (RRR), on mm. hoitotutkimuksien tehokkuuden arvioinnissa käytetty suure, joka määritellään: $RRR = (riski_2 - riski_1) / riski_2 = 1 - (riski_1 / riski_2)$, missä $risk_i$ on tapahtuman riski vertailuryhmässä ja $risk_i$ vastaavasti tapahtuman riski hoitoryhmässä. RRR ilmoitetaan yleensä prosenttilukuna. Riskiä arvioidaan tavallisesti kumulatiivisella ilmaantuvuudella.

Vaihtoehtoinen tapa raportoida tulos on ilmoittaa absoluuttinen CHD-riskin reduktio, eli **riskiero** ("absolute risk difference" ARD), joka määritellään: $ARD = riski_2 - riski_1$, joten

2. CHD-riskin reduktio: $9,8 \% - 8,1 \% = 1,7 \%$
 3. 95 % luottamusväli: [-0,1 %, 3,5 %]
- (1. $P > 0,05$, N.S.)

Hoitokokeissa vaikutuksen arvioimiseksi käytetään usein myös mittaa, joka kuvaa, kuinka monta potilasta täytyy aktiivihoidolla hoitaa, jotta vältetään tapahtuma, joka muuten tapahtuisi, jos potilasta esimerkiksi hoidetaan lumelääkkeellä. Tämä **NNT-luku** ("number needed to treat") lasketaan absoluuttisen riskieron perusteella.

Absoluuttinen riskiero "absolute risk difference" (ARD) ARD tai absoluuttinen riskin reduktio (ARR) mittaa vertailevissa tutkimuksissa tietyn tapahtuman riskin eroa toisiinsa vertailtavissa ryhmissä. Mitta kuvaa suhteellista riskiä paremmin hoidon kansanterveydellistä tai terveystaloustieteellistä merkitystä, koska se riippuu perusriskistä eli taudin yleisyydestä. Määritellään: $ARD = riski_2 - riski_1$, missä $risk_i$ on tapahtuman riski vertailuryhmässä ja $risk_i$ vastaavasti tapahtuman riski hoitoryhmässä. NNT-luku lasketaan ARD:n perusteella.

NNT-luku "number needed to treat" "hoitohyötyluku" on vertailevissa hoitokokeissa vaikutuksen arvioimiseksi käytetty mitta, joka kuvastaa sitä kuinka monta potilasta tarvitsee hoitaa, jotta vältettäisiin tapahtuma (esim. kuolema), joka ilman hoitoa tapahtuisi. Lasketaan kaavalla: $NNT = 1 / (riski_2 - riski_1)$. NNT-luku voidaan laskea sellaisista hoitokokeista, joissa lopputulosmuuttuja on kaksiarvoinen. Suuri hoitoeffekti

absoluuttisella skaalalla johtaa pieneen NNT arvoon. Negatiivinen NNT-luku, "hoitohaittaluku" (NNTH) tarkoittaa, että hoito on haitallinen.

Yllä olevassa esimerkissä $NNT = 1 / 0,017 = 58,8$, eli jos hoidetaan 59 potilasta aktiivihoidolla, säästytään lumehoitoon verrattuna yhdeltä CHD-kuolemalta tai jos hoidetaan 1000 potilasta, niin säästytään 17 kuolemantapaukselta.

Kun NNT-lukuja vertaillaan eri sairauksien hoidon välillä, kannattaa muistaa, että ne ovat vertailukelpoisia vain, mikäli perussairastumisriskit ovat vertailun kohteena olevissa hoidoissa samat. Täten NNT-luku pätee vain tutkitun joukon eri riskiryhmien ns. keskimääräisen potilaan kohdalla. Hoitavan lääkärin on hyvä huomata, että tätä lukua ei voi soveltaa yksittäiseen potilaaseen, koska mikäli yksittäisen potilaan sairastumisriski on suurempi tai pienempi kuin tämä ns. keskimääräisen potilaan kohdalla, ei NNT-luku anna oikeaa tietoa tämän yksilön kohdalla.

Esim.

Oletetaan, että sairauden A hoidossa saadaan jollain uudella hoidolla sairastumisriskin reduktio $100\% \rightarrow 90\%$ perinteiseen hoitoon verrattuna ja oletetaan, että vastaavasti sairauden B hoidossa saadaan reduktio $50\% \rightarrow 40\%$. Kummankin sairauden hoidossa saadaan uuden hoidon NNT-luvuksi $1 / 0,10$, eli 10, vaikka uuden hoidon merkitys voi näissä tapauksissa olla huomattavastikin erilainen.

NNT- luvulle voidaan helposti laskea myös luottamusväli ottamalla käänteisluku $ARD:n$ luottamusvälin ylä- ja alarajoista. Mikäli hoitoeffekti ei ole tilastollisesti merkitsevä, eli nolla sisältyy hoitoeron 95% :n luottamusväliin, joudutaan NNT- luvun luottamusvälin kanssa tulkinnallisiin vaikeuksiin; luottamusvälin alaraja olisi negatiivinen (hoidolla haitallinen vaikutus) eikä saatu NNT-arvio (piste-estimaatti) sisälly väliin. Internetissä osoitteessa www.healthcare.ubc.ca/calc/clinsig.html on kätevä laskin NNT-lukujen ja muiden kliinisen merkittävyyden arviointiin käytettyjen suureiden luottamusvälien laskemiseen.

Esim.

Lievän hypertension hoitokoe (MRC-tutkimus, 1985). Tutkimusaineistoon kuului 17 354 potilasta, jotka olivat iältään 35 -64 vuotiaita ja, joilla diastolinen paine oli välillä 90 -109 mmHg. Aktiivihoidoryhmiä oli kaksi (bendrofluazide ja propranolol) ja lisäksi oli lumehoidoryhmä. Seuranta-aika oli 5,5 vuotta ja Päälopputulosmuuttujana oli aivoinfarktien määrä. **Tulos:**

	Aktiivihoidot yhteensä	Lumehoito	Reduktio	CL 95 %
Tapahtumat	60	109	45 %	(25 %, 61 %)
Seurantavuodet	42911	42661		
Ilmaantuvuus per 1000 henkilövuotta	1,4	2,6	1,2	(0,56, 1,76)
NNT-luku			833	(556, 1662)

Reduktion luottamusväli lasketaan Poisson-jakauman perusteella. Laskettu väli on melko laeva mikä merkitsee sitä, että päättelyyn sisältyy epävarmuutta. Eron tilastolliseksi merkitsevyydeksi tutkijat ilmoittivat $P < 0,01$.

NNT voidaan arvioida myös elossaoloanalyysien perusteella. Silloin se ei kuitenkaan ole yksittäinen luku, vaan ajasta riippuva suure. Esim. Kaplan-Meier-analyyseista tai Coxin

mallista saatavien elossaolotodennäköisyyksien tai riskisuhteiden (HR) perusteella NNT voidaan laskea mihin tahansa hoidon aloittamisen jälkeiseen ajankohtaan liittyvänä. Silloinkin NNT-luvulle kannattaa laskea luottamusväli.

Esimerkkejä luottamusvälien laskentatilanteista

Keskiarvot ja niiden erotukset <ul style="list-style-type: none"> • yksi otos • kaksi otosta (riippumattomat, parittaiset)
Suhteelliset osuudet ja niiden erotukset <ul style="list-style-type: none"> • yksi otos (eri jakaumat: binomi, Poisson, jne.) • kaksi otosta (riippumattomat, parittaiset)
Regressio ja korrelaatio <ul style="list-style-type: none"> • yksi otos • kaksi otosta
Eloojäämisanalyysit <ul style="list-style-type: none"> • yksi suhteellinen osuus • suhteellisten osuuksien erotukset
Parametrittomat analyysit <ul style="list-style-type: none"> • yksi otos (mediaani, muut kvartiilit) • kaksi otosta mediaaninen ero (riippumattomat, pareittaiset)

Normaalijakaumaan perustuvat luottamusvälit

Monet likimääräisistä luottamusväleistä perustuvat **normaalijakaumaan**. Luottamusvälien laskentakaava on muotoa:

$$\text{Alaraja} : \hat{\theta} - z_{1-\alpha/2} \cdot \text{SE}(\hat{\theta}), \quad \text{Yläraja} : \hat{\theta} + z_{1-\alpha/2} \cdot \text{SE}(\hat{\theta}),$$

Kaavassa $\hat{\theta}$ on tarkasteltavan suureen θ piste-estimaatti, $\text{SE}(\hat{\theta})$ sen keskivirhe.

Esimerkiksi aritmeettiselle keskiarvolle $\text{SE}(\bar{x}) = \sigma / \sqrt{n}$, jossa σ hajonta perusjoukossa ja n on aineistokoko. Mediaanin keskivirhe saadaan kaavasta: $\text{SE}(\text{Med}) = (x_{(i)} - x_{(j)}) / (2\sqrt{3})$. (Tukey, 1977). Kaavassa $x_{(i)}$ ja $x_{(j)}$ ovat i . ja j . arvo järjestetyssä otoksessa $(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ siten, että $i = \text{INT}((n + \sqrt{3n})/2 + 1)$, $j = \text{INT}((n - \sqrt{3n})/2 + 1)$. Funktio $\text{INT}(\cdot)$ ottaa sulussa olevasta lausekkeesta kokonaisosan; esim. $\text{INT}(6,8) = 6$.

Luku $z_{1-\alpha/2}$ on standardin normaalijakauman (keskiarvo 0, hajonta 1) prosenttipiste. Esim. 95 %:n luottamusvälissä $z_{0,975} = 1,96$. Nämä prosenttipisteet saadaan esim. Excelillä valikosta "Insert Function" valitsemalla funktio **NORMINV** sijoittamalla lokeroon "Probability" arvo 0,975, "Mean" arvo 0 ja "Standard_dev" arvo 1

Function Arguments ? X

NORMINV

Probability 0,975 = 0,975

Mean 0 = 0

Standard_dev 1 = 1

= 1,959962787

Returns the inverse of the normal cumulative distribution for the specified mean and standard deviation.

Standard_dev is the standard deviation of the distribution, a positive number.

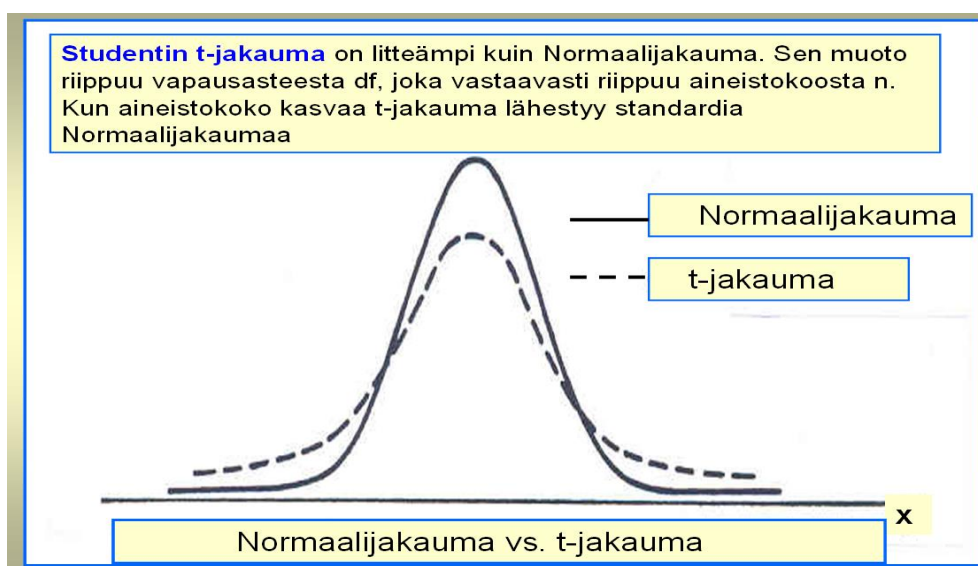
Formula result = 1,959962787

Keskiarvon luottamusväli

Tärkeimmille tutkimustuloksille, lopputulosmuuttujille, on tapana ilmoittaa piste-estimaatin ohella myös luottamusväli. Mikäli kyseessä on jatkuva muuttuja ja sen aritmeettinen keskiarvo, niin $100(1-\alpha)$ %:n luottamusväli lasketaan kaavalla:

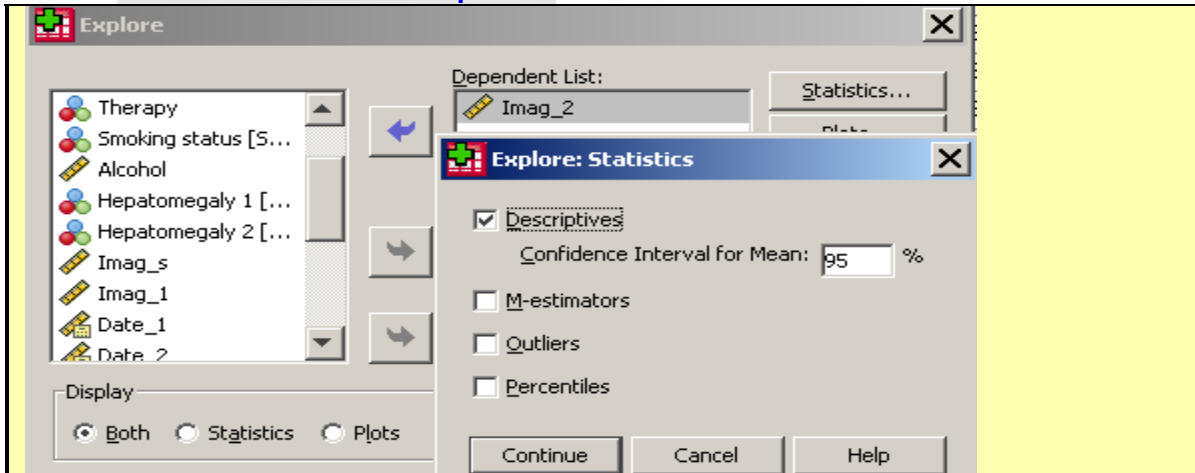
$$\text{Alaraja} : \bar{x} - t_{1-\alpha/2(n-1)} \cdot SD/\sqrt{n}, \quad \text{Yläraja} : \bar{x} + t_{1-\alpha/2(n-1)} \cdot SD/\sqrt{n}$$

Kaavassa $t_{1-\alpha/2(n-1)}$ on **Studentin t-jakauman** prosenttipiste vapausastein $n-1$, \bar{x} on aineistosta (x_1, x_2, \dots, x_n) laskettu aritmeettinen keskiarvo ja SD/\sqrt{n} on keskiarvon keskivirheen arvio.



Esim.

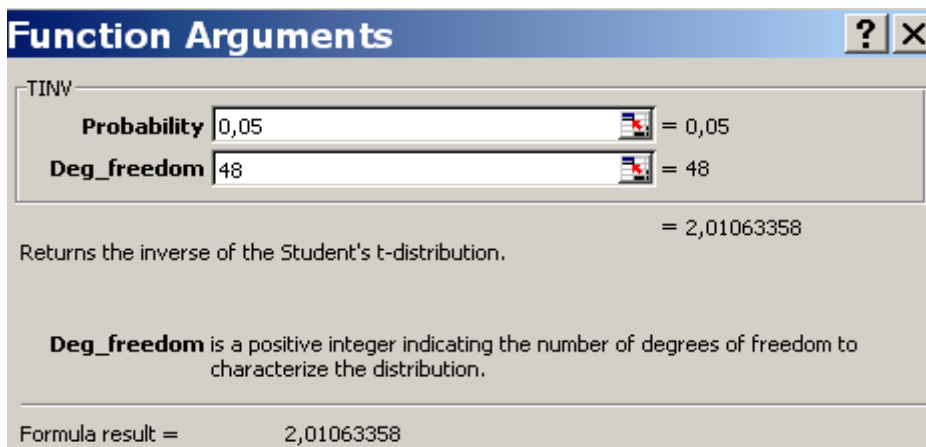
Tietokanta **TRIAL**. Muuttuja: **"Imag_2"**, joka on hypoteettinen laboratoriomittaus hoidon lopussa. Halutaan laskea muuttujan keskiarvo ja 95 %:n luottamusväli.
SPSS: Valikot: **"Analyze"** ▶ **"Descriptive Statistics"** ▶ **"Explore"**. Kohdasta Statistics valitaan **"Descriptives"**



Tulos:

Cases	Valid		Missing		Total	
	n	Percent	n	Percent	n	Percent
Imag_2	49	98,0%	1	2,0%	50	100,0%
Descriptives					Statistic	Std. Error
Imag_2	Mean				42,659	3,3950
	95% Confidence Interval for Mean				Lower Bound	35,833
					Upper Bound	49,485

Tämän välin laskennassa on käytetty t-jakauman arvoa $t_{0,975}(48) = 2,011$, joka Excelillä saadaan valikosta **"Insert Function"** valitsemalla funktio **TINV**. Lokeroon **"Probability"** sijoitetaan arvo $\alpha=1-0,95$ ja lokeroon **"Deg_freedom"** arvo $n-1 = 49-1 = 48$.



Mikäli alkuperäistä havaintoaineistoa ei olisi käytettävissä vaan ainoastaan tunnusluvut: keskiarvo ja keskihajonta, niin luottamusväli voitaisiin laskea esim. ohjelmalla CIA.

Keskiarvojen erotuksen luottamusväli

Vertailevissa kliinisissä tutkimuksissa hoidon lopputulosta mittaava suure on usein jatkuva ja hoitoero ilmoitetaan keskiarvojen erotuksena. Oletetaan, että ryhmät ovat toisistaan riippumattomia, eli muodostuvat eri potilaista. Merkitään ryhmäkokoja n_1 ja n_2 , keskiarvoja \bar{x}_1 ja \bar{x}_2 , sekä keskihajontoja SD_1 ja SD_2 . Muodostetaan **yhdistetty** ("pooled") **keskihajonta** kaavalla:

$$SD_p = \sqrt{\frac{(n_1 - 1) \cdot SD_1^2 + (n_2 - 1) \cdot SD_2^2}{n_1 + n_2 - 2}}$$

Keskiarvojen erotuksen keskivirhe:

$$SE(\bar{x}_1 - \bar{x}_2) = SD_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Keskiarvojen erotuksen luottamusväli:

$$\begin{aligned} \text{Alaraja} &: (\bar{x}_1 - \bar{x}_2) - t_{1-\alpha/2}(n_1 + n_2 - 2) \cdot SE(\bar{x}_1 - \bar{x}_2), \\ \text{Yläraja} &: (\bar{x}_1 - \bar{x}_2) + t_{1-\alpha/2}(n_1 + n_2 - 2) \cdot SE(\bar{x}_1 - \bar{x}_2) \end{aligned}$$

Kaavassa $t_{1-\alpha/2}(n_1+n_2-2)$ on t-jakauman prosenttipiste ja α on luottamustaso.

Esim.

Tietokanta **Trial**. Tarkastellaan muuttujaa "**Imag_2**" sukupuolen "**Gender**" mukaan ryhmitettynä. **Halutaan** laskea ryhmien välinen ero ja erotuksen 95 % luottamusväli. **SPSS**: Valikot: "**Analyze**" ► "**Compare Means**" ► "**Independent-Samples T Tests**". Kohdassa "Grouping variable" annetaan ryhmittelevän muuttujan koodiarvot ja kohdassa "Options" annetaan luottamustasoksi 95 % (oletusarvo)

The screenshot shows the SPSS 'Independent-Samples T Test' dialog box. In the 'Test Variable(s):' field, 'Imag_2' is selected. In the 'Grouping Variable:' field, 'Gender(1 2)' is selected. The 'Options...' sub-dialog box is open, showing 'Confidence Interval:' set to 95%. Under 'Missing Values', the radio button for 'Exclude cases analysis by analysis' is selected. The 'Independent Samples Test' table at the bottom shows 'Imag_2' as the variable being tested.

Independent Samples Test	
	Imag_2

		Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F	,223		
	Sig.	,639		
t-test for Equality of Means	t	-,855	-,856	
	df	47	46,996	
	Sig. (2-tailed)	,397	,396	
	Mean Difference	-5,8245	-5,8245	
	Std. Error Difference	6,8104	6,8032	
	95% Confidence Interval of the Difference	Lower	-19,5253	-19,5108
		Upper	7,8763	7,8618

Huom.

Mikäli ryhmien 1 ja 2 variansseja ei oleteta yhtä suuriksi, niin yhdistettyä varianssia SD_p ei muodosteta ja keskiarvojen erotuksen keskivirhe lasketaan kaavalla $(SE_1^2 + SE_2^2)^{1/2}$. Varianssien yhtä suuruus voidaan testata **Levenen mediaanitestillä**. Testi antaisi tässä tapauksessa P-arvoksi **0,639**, joka on paljon suurempi kuin yleisesti käytetty merkitsevyytaso 0,05, joten varianssit voidaan olettaa yhtä suuriksi ja käyttää keskivirheenä arvoa 6,810. Molempien ryhmien hajontojen tulisivatkin olla yhtäsuuria, jotta luottamusvälin laskenta oli luotettavaa.

Tulkinta

Keskiarvojen erotuksen arvio yhteen desimaaliin pyöristettynä on **-5,8** ja sen 95 % luottamusväli (**-19,5, 7,9**). Väli on melko laava, mikä tarkoittaa sitä, että arvioon sisältyy paljon tilastollista vaihtelua, epävarmuutta. Arvo nolla sisältyy kuitenkin selvästi väliin, mikä merkitsee sitä, että miesten ja naisten välinen ero ei ole tilastollisesti merkitsevä. Tilastollisen merkitsevyyden arviointi ei kuitenkaan ole luottamusvälien pääasiallinen käyttötarkoitus, vaan tarkasteltavan suureen otantavaihtelun arviointi. Luottamusvälin kaavassa esiintyvä t-arvo on tässä tapauksessa $t_{0,975}(47) = 2,0118$ (StaTable tai Excel).

Mikäli kyseessä on toisistaan riippuvat ryhmät 1 ja 2, esimerkiksi samat potilaat ja kaksi ajankohtaa, tai kaltaistetut parit joista on mitattu havaintoarvot (x_{1i}, x_{2i}) , $i = 1, \dots, n$, niin ajankohtien välisen muutoksen tai parien välisen erotuksen $d_i = x_{2i} - x_{1i}$ keskiarvon luottamusväli lasketaan kuten edellä yhden tutkimusryhmän tapauksessa.

Esim.

Aineisto: Tietokanta **Trial**.. Tarkastellaan erotusmuuttujaa **Imag_D = Imag_1 - Imag_S**, missä **Imag_S** on mittaus hoidon alussa ja **Imag_1** on mittaus viikon kuluttua. SPSS- ohjelmalla saadaan **Imag_D**:n keskiarvoksi 3,2646 ja keskivirheeksi 0,9396. Aineistokoko $n = 48$, koska yhdeltä henkilöltä puuttui alkuarvo ja yhdeltä puuttui loppuarvo. 95 %:n luottamusväliksi keskiarvojen erotukselle tulee: (**1.3744, 5.1548**).

Mikäli käytettävissä ei olisi alkuperäistä havaintoaineistoa, niin luottamusväli voitaisiin laskea esim. ohjelmalla **CIA**.

Mediaanien ja muiden prosenttipisteiden luottamusvälit

Usein kliinisissä tutkimuksissa jakaumat ovat vinoja ja keskilukuna käytetään silloin tavallisesti mediaania. Mediaanin luottamusväli voidaan laskea seuraavasti:

Lajitellaan aineisto (x_1, \dots, x_n) nousevaan järjestykseen $(x_{(1)}, \dots, x_{(n)})$. Lasketaan luvut r ja s siten, että:

$$r = n/2 - z_{1-\alpha/2} \sqrt{n}/2 \text{ ja } s = 1 + n/2 + z_{1-\alpha/2} \sqrt{n}/2$$

Luottamusväli: $CL_{95\%}: [x_{(r)}, x_{(s)}]$

Esim.

Tietokanta **Trial**. Muuttuja: **Imag_2**. $n = 49$. Ohjelmalla SPSS saadaan mediaaniksi 34,2. Ohjelma ei kuitenkaan laske mediaanin luottamusväliä. Se voidaan laskea osoitteessa: http://www.mv.helsinki.fi/home/sarna/stats/Prosenttipisteen_luottamusväli.xls olevalla laskimella. Tulokseksi saadaan $r = 17,6$ ja $s = 32,4$, eli pyöristettynä $r = 18$ ja $s = 32$. Tämä merkitsee sitä, että mediaanin 95 %:n väli on lajitellussa aineistossa **18.** ja **32.** arvo. Tämä luottamusväli ei ole tarkalleen 95 %:n väli vaan tässä tapauksessa **95,6 %**:n väli.

Ohjelma CIA antaa kaksi vaihtoehtoista laskentatapaa mediaanin luottamusvälin laskemiseksi: 1) tietojen suora syöttö (esim. työpöydän kautta Excelistä tai muista Windows-ohjelmista) tai 2) laskenta pelkän aineistokoon (n) perusteella.

Ohjelmassa on kaksi vaihtoehtoista menetelmää luottamusvälin laskemiseksi: **binomijakaumaan perustuvan** välin ja **Wilcoxonin testisuureeseen** perustuvan välin.

Imag_2:n mediaanille saadaan binomialisella menetelmällä 95,6 %:n luottamusväliksi (28,5, 46,6) ja Wilcoxonin menetelmällä 95,1 %:n väliksi: (32,3, 48,05).

Medians and their differences				
Single Sample Unpaired Samples Paired Samples				
Single Sample - Median and other Quantiles				
Sample size				49
Calculation Method			Quantiles	
<input checked="" type="radio"/> Binomial <input type="radio"/> Wilcoxon			Median 50	
Binomial method		Median	95.6% Confidence Interval for the Median Observation	
		34,2	28.5	to 46.6
Wilcoxon method		Median	95.1% Confidence Interval for the Median Average	
		40,5	32.3	to 48.0
				K* 416

Muille prosenttipisteille (q) r ja s korvataan arvoilla:

$$r' = n \cdot q - z_{1-\alpha/2} \sqrt{n \cdot q \cdot (1-q)} \quad \text{ja} \quad s' = 1 + n \cdot q + z_{1-\alpha/2} \sqrt{n \cdot q \cdot (1-q)}$$

Esim.

Tietokanta Trial. Muuttuja Imag_2. Ylimmän kvartiiliin, 75 % - pisteen ($q = 0,75$) 95 %:n luottamusväliksi saadaan ohjelmalla CIA:

Binomial method	Median	95% Confidence Interval for the 75th Quantile Observation	to	95th Quantile Observation
	34.2	41.8		70.5

Mediaanieron luottamusväli

Oletetaan, että kyseessä on kahden toisistaan riippumattoman ryhmän vertaileva kliininen tutkimus, jossa hoitoeroa halutaan mitata mediaanieron, esim. jakauman vinoudesta johtuen. Mediaanieroa ja sen luottamusväliä useimmat tilastopakettit (esim. SPSS) eivät laske. Sen sijaan erityisohjelma CIA laskee.

Oletetaan, että tarkasteltava muuttuja on x ja ryhmäkoot ovat n_1 ja n_2 . Mediaaniero ja sen luottamusväli lasketaan seuraavasti:

1. Muodostetaan yhdistetystä aineistosta kaikki mahdolliset muuttujan x havaintoarvojen erotukset $d_i = x_{i1} - x_{j2}$, $i = 1, \dots, n_1$, $j = 1, \dots, n_2$. Näitä erotuksia tulee yhteensä $N = n_1 \cdot n_2$, joten l saa arvot $1, \dots, N$.
2. Lajitellaan erotukset nousevaan järjestykseen $d_{(1)}, \dots, d_{(N)}$
3. Kaikkien erotusten d_l mediaani on ryhmien 1 ja 2 välinen mediaaniero.

Mediaanieron likimain $100 \cdot (1-\alpha) \%$:n luottamusväli saadaan seuraavasti:

$[d_{(k)}, d_{(N-k)}]$, missä $d_{(k)}$ on k . pienin ja $d_{(N-k)}$ on k . suurin erotus ja luku k saadaan kaavasta: $k = W_{\alpha/2} - n_1 \cdot (n_1+1)/2$, missä $W_{\alpha/2}$ on Mann–Whitney:n testisuureen jakauman $(100 \cdot \alpha/2)$ -prosenttipiste.

Huom.

Samoin kuin mediaaninkin tapauksessa tämä menettely ei välttämättä tuota täsmälleen 95 %:n luottamusväliä. Parittaisessa tapauksessa voidaan käyttää mediaanin luottamusvälejä erotuksille

Esim.

a) Riippumattomat otokset, $n_1 = n_2 = 30$ ja $\alpha = 0,05$. Erotuksia on siten $30 \cdot 30 = 900$. Ohjelmalla CIA saadaan: $k = 318$ joten likimääräiseksi 95 %:n luottamusväliksi tulee: $CL_{95\%} = [d_{(318)}, d_{(583)}]$

b) Parittaiset otokset, $n = 30$, $\alpha = 0,05$. Erotuksia on $n \cdot (n+1)/2 = 465$. Ohjelmalla CIA saadaan $k = 138$ joten likimääräiseksi 95 %:n luottamusväliksi tulee: $CL_{95\%} = [d_{(138)}, d_{(328)}]$ (Wilcoxonin testiin perustuva menetelmä)

Esim.

Tietokanta **Trial**. Muuttuja: **Imag_2**. Ohjelmalla **CIA** saadaan seuraavat tulokset ryhmien "Therapy"=1 ja "Therapy"=2 välillä. Ryhmien koot: $n_1 = 15$, $n_2 = 16$, mediaani₁ = **25,20**, mediaani₂ = **34,35**. Mediaaniero = **- 7,30** (Huom. Tämä ei ole aivan sama kuin mediaanien ero, mutta käytännössä numeerisesti likimain sama). Mediaanieron **95,1 %**:n luottamusväli on **[-19,70, 1,80]**

Suhteellisen osuuden luottamusväli

Oletetaan, että tutkittavan ryhmän koko on n ja k on niiden henkilöiden lukumäärä, joilla on jokin tietty ominaisuus. Arvioitu suhteellinen osuus kyseisen ominaisuuden esiintyvyydelle on siten $p = k/n$. Tämä on nk. piste-estimaatti vastaavalle perusjoukon suureelle. Tarkempi kuva ominaisuuden yleisyydestä perusjoukossa saadaan laskemalla lisäksi myös suhteellisen osuuden luottamusväli. Sen laskemiseksi on käytettävissä useita eri menetelmiä. Sopivimman menetelmän valinta eri tilanteissa ei ole triviaali asia. Siitä huolimatta monissa tilastopaketeissa, kuten SPSS:ssä, menetelmätarjonta kyseiseen ongelmaan on olematon. Sen sijaan erityisohjelmilla kuten CIA ja StatXact 8 laskenta kuhunkin tilanteeseen parhaiten soveltuvalla menetelmällä onnistuu.

Niin kutsuttu **perinteinen $100 \cdot (1 - \alpha) \%:n$ luottamusväli** perustuu Normaalijakauma-approksimaatioon ja lasketaan kaavoilla:

$$\text{Alaraja : } p - z_{1-\alpha/2} \sqrt{p(1-p)/n} \qquad \text{Yläraja : } p + z_{1-\alpha/2} \sqrt{p(1-p)/n}$$

Tätä väliä ei pidä käyttää, jos suhteellinen osuus on kovin pieni tai suuri, esim. kun k tai $n-k$ on alle 5.

Esimerkiksi CIA-ohjelmistoon sisältyvä **Wilsonin menetelmä** (Wilson 1927) on perinteistä menetelmää suositeltavampi useimmissa käytännön tilanteissa edullisempien tilastollisten ominaisuuksiensa vuoksi. Wilsonin menetelmässä lasketaan ensin suureet:

$$A = 2 \cdot k + z_{1-\alpha/2}^2, \quad B = z_{1-\alpha/2} \sqrt{z_{1-\alpha/2}^2 + 4 \cdot k \cdot \left(1 - \frac{k}{n}\right)} \quad \text{ja} \quad C = 2 \cdot (n + z_{1-\alpha/2}^2)$$

missä $z_{1-\alpha/2}$ on kuten aiemminkin Normaalijakauman prosenttipiste, jonka arvo 95 %:n välissä on 1,96. Perusjoukon suhteellisen osuuden luottamusväli (**L, U**) saadaan kaavoista:

$$\text{Alaraja: } L = (A-B)/C \quad \text{ja} \quad \text{Yläraja: } U = (A+B)/C$$

Menetelmällä on todettu olevan monia hyviä ominaisuuksia ja sen käyttömahdollisuudet ovat laajat. Myös silloin kun $k=0$ sitä voidaan käyttää. Tämä väli voidaan helposti laskea myös Excelillä (<http://www.mv.helsinki.fi/home/sarna/stats/WilsonCL.xls>).

Esim.

Bakteeriurian esiintyvyys synnyttäjillä. Tutkimusaineistoon kuului 196 synnyttäjää ja todettiin $k = 11$ bakteeriuriatapausta, joten esiintyvyyden piste-estimaatti on: $p = 11/196 = 0,056$.

Wilsonin menetelmä			
Osoittaja (k)	11		
Nimittäjä (n)	196		
Suhdeluku (p)	0,0561		
$1-\alpha$	0,9500	$z_{1-\alpha/2} =$	1,96
$(1-\alpha)\%$ CL	0,0316	0,0977	

Tuloksia eri ohjelmilla perinteisellä ja nk. eksakteilla menetelmillä

- **CIA: (0,023, 0,088)** Perinteinen ja **(0,032, 0,098)** Wilson
- **StatXact 8: (0,028, 0,098)** Clopper-Pearson, **(0,029, 0,097)** Blyth-Still-Casella

Todetaan, että eksakteilla menetelmillä lasketut välit ovat leveämpiä kuin perinteisellä menetelmällä laskettu likimääräinen väli.

Esim.

Kahden periodin ristikkäistutkimusasetelma, vertaillaan hoitoja A ja B ja potilaiden määrä on n. Lopputulos kolmiarvoinen: A parempi, B parempi, yhtä hyviä.

Lopputulos	Lukumäärä
A parempi	a
B parempi	b
Yhtä hyviä	$n - (a+b)$
Yhteensä	n

Voidaan näyttää, että suhteellisen erotuksen (preferenssin) $|a - b| / n$ likimääräinen $100 \cdot (1 - \alpha) \%:n$ luottamusväli voidaan laskea kaavalla:

$$\frac{|a - b|}{n} \pm z_{1-\alpha/2} \cdot \sqrt{\frac{a + b}{n}}$$

Normaalijakaumaan perustuvana likimääräinen testisuure on: $z = \frac{|a - b|}{\sqrt{a + b}}$

Num. esim.

40 potilaan aineistossa saatiin tulos: kolme potilaista piti lääkettä A parempana kuin B:tä, 13 oli päinvastaista mieltä ja loppujen mielestä lääkkeet olivat yhtä hyviä. Tällöin $a = 3$ ja $b = 13$ ja B:n suhteellinen preferenssi A:han verrattuna on $|3 - 13| / 40$, eli **25 %** ja luottamusväli **CL95 %** on: **(5,4 %, 44,6 %)**. ($z = 2,50$, **P = 0,0248**).

Esim.

Diagnostisten testisuureiden luottamusvälien laskeminen Wilsonin menetelmällä. Eräaseen tutkimusaineistoon kuului 245 ylävatsaoireista potilasta. Heiltä oli mitattu seerumin pepsinogeeni I:n taso ja lisäksi heille oli suoritettu endoskopiatus. Yhteensä 40 henkilöllä todettiin korpusatrofia (**"Disease +"**) ja lopuilla endoskopiatus oli joku muu kuin korpusatrofia (**"Disease -"**). PGI luokiteltiin **"PGI < 30 µg/l" = "Test +"** ja **"PGI ≥ 30 µg/l = "Test -"**.

Halutaan laskea diagnostiset suureet ja niiden 95 % luottamusvälit.

Sensitiivisyys, "Sensitivity" Seulonta- tai diagnostisen testin herkkyys taudin toteamiseksi. Se on ehdollinen todennäköisyys sille, että testi antaa oikean positiivisen tuloksen silloin, kun testattavalla henkilöllä todella on testauksen kohteena oleva tauti; ilmaisee testin kyvyn löytää sairait.

Spesifisyys, "Specificity" Seulonta- tai diagnostisen testin tarkkuus taudin toteamiseksi. Se on ehdollinen todennäköisyys sille, että testi antaa oikean negatiivisen tuloksen silloin, kun testattavalla henkilöllä todellisuudessa ei ole testauksen kohteena olevaa tautia; ilmaisee testin kyvyn tunnistaa terveet.

Ennustearvot mittaavat diagnostisten testien yhteydessä kuinka hyödyllinen testi on käytännössä. **Positiivinen ennustearvo (PPV)** on todennäköisyys, että henkilöllä on diagnosoimien kohteena oleva tauti (tai tila) ehdolla, että testitulokset on ollut positiivinen. Vastaavasti **negatiivinen ennustearvo (NPV)** on todennäköisyys, että henkilöllä ei ole diagnosoimien kohteena olevaa tautia tai tilaa ehdolla, että testitulokset on ollut negatiivinen.

PPV = (Sensitiivisyys) · (Vallitsevuus) / ((Sensitiivisyys · Vallitsevuus) + (1 - Spesifisyys) · (1 - Vallitsevuus)) ja

NPV = (Spesifisyys) · (1 - Vallitsevuus) / ((1 - Sensitiivisyys) · (Vallitsevuus) + (Spesifisyys) · (1 - Vallitsevuus))

Tulos:

Diagnostic Studies

Sensitivity & Specificity | Likelihood Ratio | ROC curve | Kappa

Sensitivity, specificity and prognostic values

		Disease		Total
		Positive	Negative	
Test	Positive	33	5	38
	Negative	7	200	207
Total		40	205	245

Which calculation method?

Recommended (Wilson) Traditional

	95% Confidence Interval		
Sensitivity	0.825	0.680	to 0.913
Specificity	0.976	0.944	to 0.990
Positive Predictive Value	0.868	0.727	to 0.942
Negative Predictive Value	0.966	0.932	to 0.984
Recommended (Wilson) method		Normal Value	1.960

Taudin tai epänormaalin tilan vallitsevuus (prevalenssi) vaikuttaa oleellisesti ennustearvoihin toisin kuin sensitiivisyyteen ja spesifisyyteen. Eri kliinisissä asetelmissa epänormaalin tilan yleisyys voi vaihdella huomattavastikin ja siten lasketut ennustearvot eivät ole yleispäteviä. Mitä harvinaisempi epänormaalin tilan vallitsevuus on, sitä varmempia yleensä voidaan olla, että negatiivinen testitulos kuvastaa hyvin potilaan normaalia tilaa ja sitä epävarmempia voidaan olla, että positiivinen testitulos kuvastaisi potilaan epänormaalia tilaa.

Vallitsevuus voidaan tulkita myös kyseisen tilan ennako(a priori) todennäköisyydeksi ennen testin suorittamista. PPV ja NPV ovat testin suorittamisen jälkeisiä, a posteriori, todennäköisyyksiä. Vertaamalla a priori ja a posteriori todennäköisyyksiä voidaan päätellä testin hyödyllisyydestä.

Todennäköisyyksien asemesta voidaan myös käyttää "mahdollisuuksien" ("odds") suhteita. Ennakkokäsitys epänormaalista tilasta ennen testiä saadaan suhteesta: Vallitsevuus/(1-Vallitsevuus). Diagnostinen testi on sitä parempi mitä enemmän testi korjaa tätä a priori "oddsia" ja siitä kertoo uskottavuussuhde.

Uskottavuussuhde (LR) "likelihood ratio" määrittelee kuinka todennäköinen jokin tapahtuma A on tietyn ehdon B ollessa voimassa verrattuna siihen, että ehto B ei ole voimassa. Kaavana: $LR = P(A | B) / P(A | \text{ei } B)$. Esim. Millä todennäköisyydellä vatsahaavapotilaalla on helicobakteeri+ verrattuna muihin ylävatsaoireita poteviin potilaisiin, eli niihin joilla ei ole todettu vatsahaavaa. Positiivinen ennustearvo lasketaan: **LR+ = Sensitiivisyys/(1-Spesifisyys)** ja negatiivinen ennustearvo vastaavasti **LR- = (1-Sensitiivisyys)/Spesifisyys**

Esim. (a posteriori odds "+") = (LR+)·(a priori odds) ja vastaavasti negatiiviselle tulokselle (a posteriori odds "-") = (LR-)·(a priori odds). Hyvissä testeissä LR+ on korkea ja vertaamalla keskenään suureita LR+ ja LR- voidaan päätellä diagnostisen testin hyödyllisyydestä. Ohjelmalla CIA saadaan seuraavat tulokset:

Diagnostic Studies

Sensitivity & Specificity Likelihood Ratio ROC curve Kappa

Likelihood Ratio

		Disease		Total
		Positive	Negative	
Test	Positive	33	5	38
	Negative	7	200	207
Total		40	205	245

Which calculation method?

Log Method Score Method

Likelihood Ratio for	Score Method	95% Confidence Interval	
Positive test	33.825	14.581	to 79.642
Negative test	0.179	0.090	to 0.328

Huom.

"Score"-menetelmän on todettu toimivan käytännössä jonkin verran paremmin kuin "Log"-menetelmän (Nam 1995).

Suhteellisten osuuksien erotuksen luottamusväli

Riippumattomat otokset

Oletetaan, että vertailtavien ryhmien lopputulosta, esim. hoitovastetta kuvaavat suhteelliset osuudet ovat: $p_1 = k_1 / n_1$, $p_2 = k_2 / n_2$

	Hoitovaste		Σ
	+	-	
Ryhmä 1	k_1	$n_1 - k_1$	n_1
Ryhmä 2	k_2	$n_2 - k_2$	n_2

Niin kutsuttu **perinteinen** $100 \cdot (1 - \alpha) \%$:n luottamusväli perustuu Normaalijakauma-approksimaatioon ja lasketaan kaavoilla:

$$\text{Alaraja} : (p_1 - p_2) - z_{1-\alpha/2} \cdot \sqrt{\frac{p_1 \cdot (1-p_1)}{n_1} + \frac{p_2 \cdot (1-p_2)}{n_2}}$$

$$\text{Yläraja} : (p_1 - p_2) + z_{1-\alpha/2} \cdot \sqrt{\frac{p_1 \cdot (1-p_1)}{n_1} + \frac{p_2 \cdot (1-p_2)}{n_2}}$$

Tätä likimääräistä väliä ei pitäisi kuitenkaan käyttää ainakaan a) pienillä aineistoilla (< 30 henkilöä kummassakin ryhmässä) tai b) mikäli suhteelliset osuudet ovat välin (0.1 – 0.9) ulkopuolella. Väli voidaan laskea esim. CIA:lla tai Excelillä osoitteessa:

<http://www.mv.helsinki.fi/home/sarna/stats/PerinteinenSuhtOsErotusCL.xls>

Kaikille suhteellisille osuuksille käyttökelpoinen ja luotettavampi suhteellisten osuuksien erotuksen luottamusvälin laskentamenetelmän on esittänyt **Newcombe** (1998, menetelmä 10)

Menetelmä on seuraavanlainen: Lasketaan ensin **Wilsonin menetelmällä** kummallekin ryhmälle erikseen suhteellisen osuuden luottamusvälit ja merkitään niitä (L_1, U_1) ja (L_2, U_2). Suhteellisten osuuksien erotuksen $100 \cdot (1 - \alpha) \%$:n luottamusväli saadaan kaavoista:

$$\text{Alaraja: } (p_1 - p_2) - \sqrt{(p_1 - L_1)^2 + (U_2 - p_2)^2}$$

$$\text{Yläaraja: } (p_1 - p_2) + \sqrt{(p_2 - L_2)^2 + (U_1 - p_1)^2}$$

Väli voidaan laskea esim. CIA:lla tai Excelillä osoitteessa:

<http://www.mv.helsinki.fi/home/sarna/stats/NewcombeSuhtOsErotusCL.xls>

Esim. Hoitoeron luottamusväli kahden hoidon välillä

Hoito	Parantui	Ei parantunut	Σ
1	21	8	29
2	12	19	31

Suhteelliset osuudet: $p_1 = 21 / 29 = 0,724$, $p_2 = 12 / 31 = 0,387$, $p_1 - p_2 = 0,337$ ja $SE(p_1 - p_2) = 0,121$, joten likimääräiseksi luottamusväliksi saadaan: $0,337 - 1,96 \cdot 0,121 = 0,101$, $0,337 + 1,96 \cdot 0,121 = 0,573$, eli (10,1 %, 57,3 %).

Wilsonin menetelmällä ryhmän 1 luottamusväli (L_1, U_1) = (**0,543, 0,853**) ja vastaavasti ryhmän 2 luottamusväli (L_2, U_2) = (**0,237, 0,562**). Sijoittamalla nämä arvot yllä oleviin kaavoihin saadaan Newcomben menetelmällä lasketuksi hoitoeron 95 %:n luottamusväliksi (**0,085, 0,535**).

Newcomben menetelmä			
	Ryhmä 1	Ryhmä 2	Erotus
Osoittaja (k)	21	12	
Nimittäjä (n)	29	31	
Suhteellinen osuus (p)	0,724	0,387	0,337
l	0,543	0,237	
p-l	0,181	0,150	
(p-l) ²	0,033	0,022	
u	0,853	0,562	
u-p	0,129	0,175	
(u-p) ²	0,017	0,031	
Erotus	0,337	0,337	
Sqrt	0,252	0,198	
1- α	0,950	$Z_{1-\alpha/2} =$	1,960
Erotuksen(1- α)% CL	0,085	0,535	

Parittaiset otokset

Käyttö:

- kaksi ajankohtaa, samat henkilöt
- kaltaistetut parit

Tarkastellaan ominaisuutta, joka on tyyppiä: kyllä, ei

1	2	Henkilöiden lukumäärä	Suhteelliset osuudet
Kyllä	Kyllä	r	$p_1 = \frac{r + s}{n}$
Kyllä	Ei	s	
Ei	Kyllä	t	$p_2 = \frac{r + t}{n}$
Ei	Ei	u	

Suhteellisten osuuksien erotus: $p_1 - p_2 = \frac{s - t}{n}$ ja sen keskivirhe:

$$SE(p_1 - p_2) = \frac{1}{n} \cdot \sqrt{s + t - \frac{(s - t)^2}{n}}$$

Suhteellisten osuuksien erotuksen luottamusvälin laskentamenetelmiä on useita. Useimmissa käytännön tilanteissa **Newcomben** esittämä menetelmä kuitenkin on todettu muita tarkemmaksi ja sopivammaksi. Sen laskentamenettely on seuraavanlainen:

- 1) Lasketaan aluksi apusuureet A, B ja C:

$$\begin{aligned} A &= (r+s) \cdot (t+u) \cdot (r+t) \cdot (s+u), \\ B &= r \cdot u - s \cdot t, \\ C &= B - n/2, \text{ jos } B > n/2, C = 0, \text{ jos } 0 \leq B \leq n/2, C = B, \text{ jos } B < 0 \\ \Phi &= C / \text{sqrt}(A) \end{aligned}$$

Suure Φ kuvastaa suhteellisten osuuksien p_1 ja p_2 välistä riippuvuutta. Se on läheisesti yhteydessä χ^2 -suureeseen. Tavallisesti se on positiivinen ja korjaa siten luottamusväliä kapeammaksi verrattuna nk. perinteisellä menetelmällä laskettuun väliin.

- 2) Lasketaan suhteellisten osuuksien p_1 ja p_2 luottamusvälit (l_1, u_1) ja (l_2, u_2) Newcomben menetelmällä kuten toisistaan riippumattomille suhteellisille osuuksille edellä.
- 3) Suhteellisten osuuksien erotuksen $100(1-\alpha) \%$:n luottamusväli saadaan kaavoilla:

$$\begin{aligned} \text{Alaraja} &: D - \sqrt{(p_1 - l_1)^2 - 2\phi(p_1 - l_1)(u_2 - p_2) + (u_2 - p_2)^2} \\ \text{Yläraja} &: D + \sqrt{(p_2 - l_2)^2 - 2\phi(p_2 - l_2)(u_1 - p_1) + (u_1 - p_1)^2} \end{aligned}$$

Nämä välit voidaan laskea CIA:lla tai Excelillä osoitteessa:

<http://www.mv.helsinki.fi/home/sarna/stats/NewcombeParittainenCL.xls>

Esim.

Tehohoidon monitoroinnin arviointitutkimus. (Eason J, Markowe HLJ. Controlled investigation of deaths from asthma in hospitals in the North East Thames region. BMJ 1987; 294: 1255-8).

Kyseessä oli parikaltaistettu tapaus-verrokkitutkimus. Aineistona oli 35 sairaalassa kuollutta astmapotilasta ja heille valitut ikä-sukupuoli-kaltaistetut parit, jotka olivat olleet potilaina samassa sairaalassa edellisenä vuonna ja poistettu sieltä elävinä. Kaikki olivat olleet tehohoidossa. Tutkimuksen kohteena oli monitoroinnin riittämättömyys.

Arvioinnin suoritti asiantuntijaraati käyttäen skaalaa (kyllä, ei). Koska kyseessä on pareittainen aineisto, niin kunkin parin kohdalla **lopputulosvaihtoehtoja on neljä:**

- Sekä astmapotilaan että verrokin monitorointi arvioitiin riittämättömäksi (**kyllä, kyllä**).
- Astmapotilaan monitorointi on ollut riittämätöntä, muttei verrokillä (**kyllä, ei**).
- Astmapotilaan monitorointi ei ollut riittämätöntä, mutta verrokillä oli (**ei, kyllä**).
- Kummallakaan monitorointi ei ollut riittämätöntä (**ei, ei**).

Arviointiraati oli sokkoutettu potilaan elossaolotatuksen suhteen. Tulokseksi saatiin:

Monitorointi riittämätön		Parien määrä
Kuolleet	Elävät	
Kyllä	Kyllä	10
Kyllä	Ei	13
Ei	Kyllä	3
Ei	Ei	9
Yhteensä		35

Suhteelliset osuudet (monitorointi riittämätön): Astmapotilaat: $p_1 = (10 + 13) / 35 = 0,6571$, Verrokkit: $p_2 = (10 + 3) / 35 = 0,3714$. Suhteellisten osuuksien erotus: $p_1 - p_2 =$

$0,2857$ (29 %) ja sen keskivirhe: $SE(p_1 - p_2) = \frac{1}{35} \cdot \sqrt{13 + 3 - \frac{(13 - 3)^2}{35}} = 0,1036$

Newcomben menetelmä			
Parittaisten suhteellisten osuuksien luottamusväli			
1	2	Parien määrä	
Kyllä	Kyllä	r	10
Kyllä	Ei	s	13
Ei	Kyllä	t	3
Ei	Ei	u	9
Parien määrä n=			35
		Suhteelliset osuudet	
Osoittaja		23	13
Nimittäjä		35	35
		Erotus (D)	
Suhteelliset osuudet (p)		0,657	0,371
		0,286	
1- α		0,950	1,960
		$Z_{1-\alpha/2}$	
Erotuksen(1- α)% CL		0,066	0,468

Tutkijat tekivät seuraavan **johtopäätöksen**: Monitorointi on kuolleilla astmapotilailla ollut **29 %**:ssa tapauksia riittämättömämpää kuin verrokeilla ja ero on tilastollisesti merkitsevä. 95 %:n luottamusväli tälle arviolle on **6,6 % - 46,8 %**.

Lukumäärän luottamusväli Poisson- jakauman tapauksessa

Käyttötilanne: esim. Ilmaantuvuuslukujen arviointi

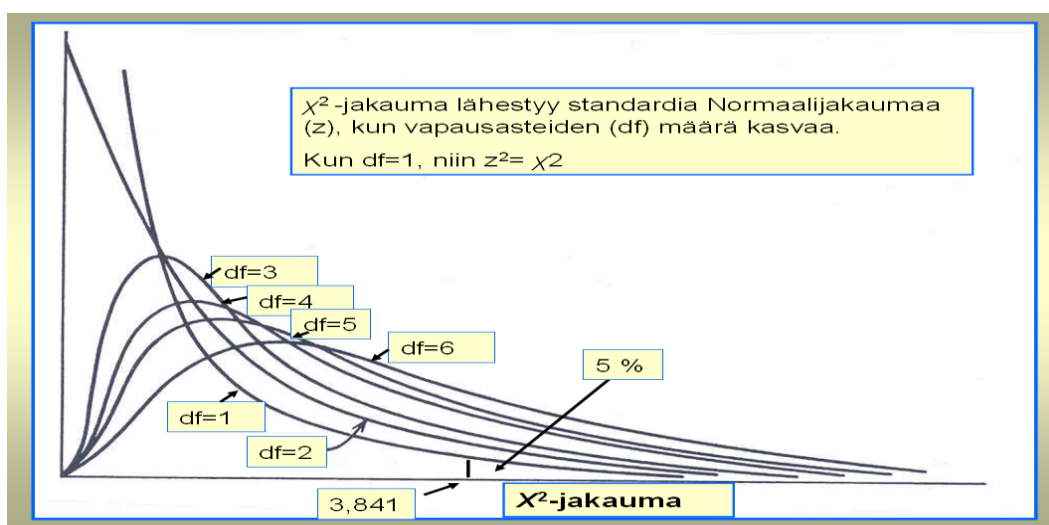
Poisson-jakauma "Poisson distribution" on harvinaisten kaksiarvoisten (kyllä / ei) tapahtumien tutkimisessa ja testaamisessa käytetty jakauma. Jakauman muotoa säätelee vain yksi parametri λ , joten kyseessä on sikäli harvinainen jakauma, että sekä sen keskiarvo ja hajonta ovat samat. Parametri λ on **tapahtumien** (onnettomuuksien lukumäärä aikayksikössä, bakteeripesäkkeinen lukumäärä elatusainemaljalla jne.) **keskimääräinen intensiteetti aika- / pinta-ala- / tilavuusyksikköä (t) kohti**.

Poisson- jakaumaa käytetään yleensä silloin, kun arvioitu tapahtumien todennäköisyys $< 0,10$. Parametrin λ piste-estimaatti on: $\hat{\lambda} = \frac{k}{t}$, missä k on havaittujen tapausten lukumäärä ja λ :n $100 \cdot (1 - \alpha) \%$:n luottamusväli voidaan laskea kaavoilla:

$$\text{alaraja: } \underline{\lambda} = \frac{\chi^2_{1-\alpha/2}(2k)}{2} \quad \text{ja yläraja: } \bar{\lambda} = \frac{\chi^2_{\alpha/2}(2k+2)}{2}$$

Osoittajissa olevat luvut ovat χ^2 - jakauman prosenttipisteitä vapausastein $2k$ ja $2k+2$. Nämä prosenttipisteet voidaan laskea esim. ohjelmilla StaTable ja Excel.

χ^2 - jakauma "chi-squared distribution" on todennäköisyysjakauma, joka saadaan summaamalla df kappaletta standardia Normaalijakaumaa noudattavien suureiden z neliöitä. Parametria df kutsutaan vapausasteiksi ("degrees of freedom") ja se säätelee jakauman muotoa. Jakauma on oikealle vino ja sitä käytetään yleisesti sekä luottamusvälien laskennassa että monissa tilastollisissa testeissä



Poisson- jakauman luottamusvälit voidaan laskea myös esim. ohjelmalla **CIA**: valikoista **“Epidemiological studies”** ▶ **“Standardized ratios”** siten, että kohtaan **“Observed number (O)”** sijoitetaan lukumäärä k ja kohtaan **“Expected number (E)”** luku 1.

Esim.

Asbestialtistus ja keuhkosityö (Armitage ja Berry 1987, s. 133).

Aineistona oli suuri joukko työntekijöitä, jotka olivat altistuneet asbestipölylle. Todettiin $k = 33$ keuhkosityöpäkuolemaa seuranta-aikana. Odotusarvo (E) vastaavassa perusjoukossa oli 20. Odotusarvo voi olla laskettu esimerkiksi kaavalla $E = \sum n_i R_i$, missä R_i on keuhkosityöpäriski perusjoukon (referenssipopulaation) ikäluokassa i , ja n_i on kyseiseen luokkaan kuuluvien henkilöiden määrä tutkimusryhmässä.

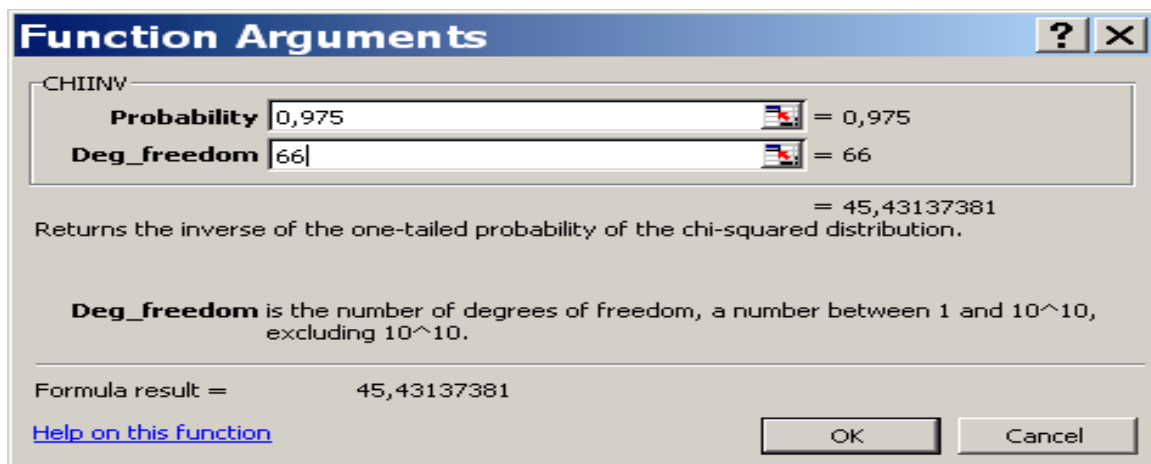
Onko keuhkosityöpäriski lisääntynyt? (Nollahypoteesi $H_0: k = k_0 (= E)$)

Koska $k = 33$, niin vapausasteet ovat $2k = 66$, $2k+2 = 68$, joten luottamusväliksi tulee:

CL₉₅ %: $\underline{\lambda} = 45,431/2 = 22,7$, $\bar{\lambda} = 92,69/2 = 46,3$

Johtopäätös: Koska odotusarvo 20 ei sisälly tähän 95 %:n luottamusväliin, niin asbestille altistuneilla työntekijöillä näyttäisi olevan lisääntynyt keuhkosityöpäkuolemien riski.

Excelillä arvo **45,431** saadaan valikoista: **“Insert function”** ▶ **“Statistical”** valitsemalla funktio **CHIINV** ja sijoittamalla **1- α /2=0,975** lokeroon **“Probability”** ja vapausaste arvo **66** lokeroon **“Deg_freedom”**.



Epidemiologisissa tutkimuksissa edellä esitetyn esimerkin kaltaisissa tilanteissa lasketaan usein myös nk. standardoitu kuolleisuussuhde (SMR, "Standardized Mortality Ratio" tai SIR "Standardized Incidence Ratio"). Tässä tapauksessa: **SMR** = $O/E = 33/20 = 1,65$ ja sen **95 %** luottamusväli vastaavasti **(1,14, 2,32)** ja 99 % väli: (1,00, 2,54).

Nämä välit voidaan laskea osoitteessa:

<http://www.mv.helsinki.fi/home/sarna/stats/PoissonCL.xls>

olevalla Excel-laskimella tai ohjelmalla CIA kohdassa "Standardized ratios".

Poisson jakauman luottamusväli		luottamusväli		SIR	luottamusväli	
Havaittu lukumäärä	33	22,72	46,34	1,65	1,14	2,32
Odotettu lukumäärä	20					
Luottamustaso	95					

The screenshot shows the 'Epidemiological Studies' software window. It has several tabs: 'Incidence Study', 'Unmatched Case Control Study', 'Series of Unmatched Case Control Studies', and 'Matched Case Control Study'. Under 'Unmatched Case Control Study', there are sub-tabs for 'Incidence Rates', 'Standardised Ratios', 'Ratio of Two Standardised Ratios', and 'Standardised Rates'. The 'Standardised Ratios' sub-tab is active. In the 'Standardised Ratio' section, the 'Observed number of cases' is 33 and the 'Expected number of cases' is 20. The 'Calculate' button is pressed. The 'Results' section shows a 'Standardised Ratio' of 1,650 and a '95% Confidence Interval for the Standardised Ratio' from 1,136 to 2,317.

Mikäli asiaa haluttaisiin testata (Testit ja niiden yleisperiaatteet käsitellään seuraavassa luvussa.) niin voitaisiin käyttää esim. normaalijakaumaan perustuvaa likimääräistä testiä:

$$z = \frac{|k - k_0| - 1/2}{\sqrt{k_0}} = \frac{|33 - 20| - 1/2}{\sqrt{20}} = 2,80, \mathbf{P = 0,0026}$$

Johtopäätös: H_0 hylätään ja erehtymisriski on korkeintaan $P = 0,0026$.

Lukumäärien suhteen luottamusväli Poisson-jakauman tapauksessa

Käyttötilanne: esim. vakioitujen ilmaantuvuuslukujen vertailussa.

Oletetaan, että k_A ja k_B ovat havaittuja lukumääriä Poisson- jakaumista, joiden odotusarvot ovat λ_A ja λ_B . Suhteen $R = \lambda_A / \lambda_B$ luottamusväli lasketaan kaavalla:

$$(R_L, R_U) = \left(\frac{\theta_L}{1 - \theta_L}, \frac{\theta_U}{1 - \theta_U} \right), \text{ missä } \theta = \lambda_A / (\lambda_A + \lambda_B) \text{ ja } (\theta_L, \theta_U) \text{ sen binomiaalinen luottamusväli.}$$

Esim. MenA rokotuskoe suomalaisilla 3 kk - 5 v ikäisillä lapsilla vuonna 1975 (Heikki Peltola et al., 1977) Ryhmä A: $n_A = 49295$, sai MenA-rokotetta (Meningococcus group A polysaccharide) Ryhmä B: $n_B = 48977$, sai Hib-rokotetta (Haemophilus influenzae type b polysaccharide) Ryhmä C: $n_C = 31906$, ei rokotettu **Tulos:** Ensimmäisen seurantavuoden aikana todettiin seuraavat tapausmäärät

	A	B	C
k	0	6	13
CL_{95%}	(0, 3.7)	(2.2, 13.1)	(6.9, 22.2)

Taulukossa olevat 95 % luottamusvälit tapahtumien lukumäärälle k voidaan laskea CIA-ohjelmalla kohdasta **"Epidemiological studies"** ► **"Incidence study"** esim. ryhmässä B seuraavasti:

1: Epidemiological Studies

Incidence Study | Unmatched Case Control Study | Series of Unmatched Case Control Studies | Ma

Incidence Rates | Standardised Ratios | Ratio of Two Standardised Ratios

Incidence Rates

Observed number of cases

Number of person-years of observation

per person-years

1 10 100 1000 10000 100000 1000000

Results

Incidence Rate per 1 person-years

95% Confidence Interval for the Incidence Rate

to

Kun tapausmäärät (k) ja niiden luottamusvälien päätepisteet kerrotaan suhteilla $100000/n_A$, $100000/n_B$ ja $100000/n_C$ saadaan ilmaantuvuuksiksi (I) 100000 henkilövuotta kohden ja niiden 95 % luottamusväleiksi:

	A	B	C
I	0	12.3	40.7
CL_{95%}	(0, 7.5)	(4.5, 26.8)	(21.6, 69.6)

Nämä tulokset voidaan laskea CIA:lla esim. Ryhmässä B seuraavasti:

1: Epidemiological Studies

Incidence Study | Unmatched Case Control Study | Series of Unmatched Case Control Studies | Ma

Incidence Rates | Standardised Ratios | Ratio of Two Standardised Ratios

Incidence Rates

Observed number of cases

Number of person-years of observation

per person-years

1 10 100 1000 10000 100000 1000000

Results

Incidence Rate per 100000 person-years

95% Confidence Interval for the Incidence Rate

to

Vakioitujen ilmaantuvuuslukujen vertailu ryhmien B ja C välillä:

$k_B = 6$ ja $k_C = 13$, joten $\hat{\theta} = \frac{k_B}{k_B + k_C} = \frac{6}{19} = 0,316$. Tämän suhteellisen osuuden

binomiaalisiksi luottamusväleiksi tulee CIA:lla: **CL₉₅ %:** $(\hat{\theta}_L, \hat{\theta}_U) = (0,126, 0,565)$ ja

CL₉₉ %: $(\hat{\theta}_L, \hat{\theta}_U) = (0,0895, 0,633)$. Kun nämä luvut sijoitetaan edellä oleviin

kaavoihin saadaan ilmaantuvuuslukujen suhteeksi: $\hat{R} = k_B / k_C = 6 / 13 = 0,4615$

ja luottamusväleiksi: **CL₉₅ %** on $(\hat{R}_L, \hat{R}_U) = \left(\frac{0,126}{1-0,126}, \frac{0,565}{1-0,565} \right) = (0,14, 1,30)$

ja **CL₉₉ %** = $(\hat{R}_L, \hat{R}_U) = (0,10, 1,72)$. Koska ryhmäkoot eroavat, sekä \hat{R} että sen

luottamusväli pitää kertoa suhteella $n_C / n_B = 31906 / 48977$ ja siten saadaan

korjatuksi arvoksi $\hat{R}^* = 0,3007$ ja sen luottamusväleiksi:

CL₉₅ %: $(\hat{R}_L^*, \hat{R}_U^*) = (0,0937, 0,8479)$ ja **CL₉₉ %:** $(\hat{R}_L^*, \hat{R}_U^*) = (0,0651, 1,1205)$.

Nämä luottamusvälit voidaan laskea myös suoraan CIA:lla valikoista: **"Epidemiological studies" ► "Ratio of Two standardised Ratios"**. Luottamustaso desimaalien määrä valitaan kohdasta **"Options"**

Epidemiological Studies		
Incidence Study	Unmatched Case Control Study	Series of Unmatched Case Control Studies
Incidence Rates	Standardised Ratios	Ratio of Two Standardised Ratios
Ratio of Two Standardised Ratios		
	Group 1	Group 2
Observed cases	<input type="text" value="6"/>	<input type="text" value="13"/>
Expected cases	<input type="text" value="48977"/>	<input type="text" value="31906"/>
Results		
Standardised Ratio	<input type="text" value="0,0001"/>	<input type="text" value="0,0004"/>
Ratio of the two Standardised Ratios	<input type="text" value="0,3007"/>	
95% Confidence Interval for the Ratio of the Two Standardised Ratios		
	<input type="text" value="0,0937"/>	to <input type="text" value="0,8479"/>

Tulkinta: Todetaan, että aivokalvontulehdusriski oli rokottamattomassa ryhmässä selvästi suurempi kuin ryhmässä B. Suojaava vaikutus oli noin 70 % - yksikköä $100 \cdot (1 - 0,3007)$ ja lisäksi todetaan, että 95 %:n luottamustasolla väli ei sisällä lukua yksi. 99 %:n tasolla ykkösen tosin sisältyy väliin, mutta silloinkin väli on vahvasti epäsymmetrinen ykkösen suhteen. Testipuolella tämä merkitsisi suunnilleen samaa kuin, että ryhmien välillä on eroa, $P < 0,05$. Luottamusväli antaa kuitenkin paljon enemmän tietoa. Se antaa käsityksen rokotteen suojaavan vaikutuksen suuruudesta toisin kuin P-arvo, joka ilmaisee ainoastaan millä varmuudella nollahypoteesi "ryhmien välillä ei ole eroa" voidaan hylätä.

Huom.

StatXact 8 ohjelmalla saadaan täsmälleen sama korjattu riskisuhde ja sen 95 % luottamusväli kuin yllä. Tarkaksi P-arvoksi se antaa joko 0,0164 tai 0,0202 menetelmästä riippuen.

Huom.

Jos riskisuhde käännetään, saadaan $R_{C/A} = 3,326$, joten rokottamattomilla on noin kolminkertainen riski saada MenA:n aiheuttama aivokalvontulehdus Hib-rokotetta saaneisiin verrattuna.

Edellä esitetyt tulokset voidaan laskea myös **StatXact 8** ohjelmalla valikoista **"Non-parametrics" ▶ "Poisson rates" ▶ "CI on Common Relative Risk"** Luottamustaso voidaan valita esim. 99 % kohdasta **"Options"**. Tarkaksi 2-suuntaiseksi P-arvoksi se antaa 0,0111, 0,0164 tai 0,0202 menetelmästä riippuen.

Number of informative strata:	1	Datan syöttömuoto: (case data)					
Event Variable:	event	event	exposure	n	strata		
Score Variable:	exposure	6	1	48977	1		
Rate Variable:	n	13	0	31906	1		
Stratum Variable:	strata						
Inference:							
		P-Value	P-Value	P-Value	P-Value	P-Value	P-Value
Type	Statistic	Tail	1-Sided	Point Prob.	1-Sided	Tail	2-Sided
Asymptotic						.GE.	0.0149
Exact	6	.LE.	0.0101	0.0075	0.0202	.GE.	0.0164
Exact: Method 2	0.0075					.LE.	0.0111
Parameters:							
		95.00%	95.00%				
		CI Limits	CI Limits				
Type	CMLE	Lower	Upper				
Asymptotic	0.3007	0.1143	0.7910				
Exact	0.3007	0.0937	0.8478				
Mid-P	0.3007	0.1053	0.7803				

Esim. MRC-tutkimus (Lievän hypertension hoitokokeilu).

Lopputulostuottajana oli aivoinfarktien määrä. Aktiivihoitoryhmässä ilmaantui 60 aivoinfarktitapausta ja lumehoitoryhmässä 109 tapausta. Tutkijat ilmoittivat tuloksen tapausmäärien suhteellisena reduktiona, eli $(109 - 60)/109 = 1 - 60/109 = 0.45$ (45 %).

Epidemiological Studies

Incidence Study | Unmatched Case Control Study | Series of Unmatched Case Control Studies | M

Incidence Rates | Standardised Ratios | Ratio of Two Standardised Ratios

Ratio of Two Standardised Ratios

	Group 1	Group 2
Observed cases	60	109
Expected cases	42911	42661

Results	
Standardised Ratio	0,001 0,003
Ratio of the two Standardised Ratios	0,547
95% Confidence Interval for the Ratio of the Two Standardised Ratios	
	0,393 to 0,757
99% Confidence Interval for the Ratio of the Two Standardised Ratios	
	0,354 to 0,833

Näiden lukujen perusteella saadaan **reduktion luottamusväli**: CL95 %: (1-0.757, 1-0.393) =(0.243,0.607) = **(24 %, 61 %)** CL99 %: (0.167, 0.646) = **(17 %, 65 %)**

Multinomijakaumaan perustuva luottamusväli

Multinomijakauma "multinomial distribution" on binomijakauman yleistys tilanteeseen, jolloin lopputulosmuuttuja on useampi kuin kaksiluokkainen.

Käyttötilanne:

Multinomijakaumaan perustuvaa luottamusväliä tarvitaan, kun lopputulosmuuttuja on s-luokkainen ja saa arvot x_1, x_2, \dots, x_s . Merkitään vastaavia koodeja: l_1, l_2, \dots, l_s

Esim.

$s = 3$, $x_1 =$ "parempi", $x_2 =$ "ennallaan", $x_3 =$ "huonompi" ja näitä x:n arvoja vastaavat koodiarvot ovat: $l_1 = +1$, $l_2 = 0$, $l_3 = -1$.

Oletetaan, että aineistossa on n potilasta ja että lopputulosten suhteelliset osuudet ovat: $(p_1, p_2, \dots, p_s) = (n_1/n, n_2/n, \dots, n_s/n)$

Lopputulos on summapistemäärä: $L = n \cdot \sum_{i=1}^s p_i l_i = \sum_{i=1}^s n_i l_i$

L:n multinomijakaumaan perustuva $100 \cdot (1 - \alpha) \%$:n luottamusväli on:

$$L \pm \sqrt{\chi_{\alpha}^2(s-1)} \cdot \sqrt{\sum p_i l_i^2 - (\sum p_i l_i)^2} \cdot \sqrt{n}$$

missä $\chi_{\alpha}^2(s-1)$ -arvot saadaan esim. ohjelmilla StaTable tai Excel.

Esim.

Reumakipujen analgeettihoido (Isomäki et al. 1984)

Tutkimuksessa vertailtiin kymmenen anti-inflammatorisen lääkkeen analgeettista vaikutusta 90:llä nivelreumapotilaalla. Potilaat olivat iältään 21-75-vuotiaita ja kaikilla oli riittävästi kipuja, jotta lääkitykseen oli tarvetta. Potilaista oli miehiä 25 ja naisia 65. Kukin potilas sai kokeen aikana kahta eri lääkettä, kolme päivää kutakin. Tutkimus toteutettiin yksöissokkoperiaatteella siten, että potilas ei tiennyt mitä lääkettä hän milloinkin otti. Kustakin lääkkeestä oli kokeessa mukana ulkonäöltään identtinen lumevalmiste. Potilaiden kipulääkitys oli lopetettu

vähintään päivää ennen koetta. Kullekin potilaalle annettiin satunnaistetussa järjestyksessä yhtä kymmenestä vertailtavasta lääkkeestä (lääke A) ja jonkin toisen lääkkeen lumevastinetta (lääke B) kolmen päivän ajan. Yhden päivän puhdistumisjakson jälkeen potilas sai lääkettä B ja lääkkeen A lumevastinetta kolmen päivän ajan. Kaikkia lääkkeitä verrattiin kaikki kaikkia vastaan ikään kuin kaksinkertaisessa jalkapallosarjassa, joten vertailuita tuli yhteensä 90 ($2 \cdot (10 \cdot 9/2)$), eli 45 seuraavanlaista otteluparia:

Potilas	Periodi 1	Periodi 2	Arvio
i	A (aktiivi) ja B (lume)	B (aktiivi) ja A (lume)	+1, 0 tai -1
j	B (aktiivi) ja A (lume)	A (aktiivi) ja B (lume)	+1, 0 tai -1

Periodi 1 vastaa ikään kuin kotiottelua ja periodi 2 vierasottelua. Kunkin kokeen jälkeen potilaalta kysyttiin, että kummalla tutkimusperiodilla kipujen lievitys oli hänen mielestään ollut tehokkaampaa vai oliko se ollut yhtä tehokasta molemmilla periodeilla. Lopputulosvaihtoehdot kussakin kokeessa koodattiin: +1, 0 tai -1. Mikäli potilas piti jotakin lääkettä toista parempana, niin lopputulos kyseiselle lääkkeelle oli +1. Tulos oli 0, jos potilas piti molempia yhtä hyvinä ja -1, mikäli lääke oli potilaan mielestä toista huonompi. Lääkkeiden saamat pistemäärät laskettiin yhteen ja tätä summaa (L) käytettiin kriteerinä lääkkeiden välisessä vertailussa. L:n teoreettinen vaihteluväli oli siten (+18, -18). Kullekin lääkkeelle laskettiin pistemäärä L ja sen 95 %:n luottamusväli.

Tulos: Tässä tapauksessa $s=3$ ja $n=18$. Lopputulos parhaan lääkkeen osalta oli: $(n_1, n_2, n_3) = (11, 4, 3)$, joten $(p_1, p_2, p_3) = (11/18, 4/18, 3/18) = (0,6111, 0,2222, 0,1667)$ ja $s=3$. Edellä olevien kaavojen perusteella saadaan $L=11 \cdot (+1) + 4 \cdot (0) + 3 \cdot (-1) = +8$ ja luottamusvälin kaavassa esiintyväksi prosenttipisteeksi tulee $\chi_{0,05}^2(2) = 5,991$ (StaTable tai Excel, kun $\alpha=0,05$ on sijoitettu lokeroon "Right" ja vapausasteikohaan $df=s-1=2$).

Pienten aritmeettisten laskutoimitusten jälkeen parhaan pistemäärän luottamusväliksi $CL_{95\%}$ saadaan: **(0,070, 15,93)**. Huonoimman lääkkeen pistemäärä oli **-13** ja luottamusväli vastaavasti: **(-21,4, -4,62)**.

Eksponttijakauman perustuva luottamusväli

Käyttötilanne: Eloanjäämisanalyysissä esiintyvien suureiden arvioinnissa

Eksponttijakauma on yksi tavallisimmista elossaolututkimuksissa käytetyistä todennäköisyysjakaumista. Siihen perustuvaa parametrilla mallia voidaan käyttää, mikäli on perusteltua olettaa, että **riskitiheysfunktio $h(t)$** ("hazard function") on vakio λ yli koko tarkasteltavan aikavälin. Parametrille λ ("hazard rate") voidaan laskea $100(1 - \alpha)$ %:n luottamusväli logaritimuunnoksen avulla. Logaritimuunnosta kannattaa käyttää siksi, että $\log_e \lambda$ noudattaa paremmin normaalijakaumaa kuin λ . $\log_e \lambda$:n Normaalijakauman perustuva luottamusväli on:

$$(\log_e \lambda - z_{1-\alpha/2} \cdot SE(\log_e \lambda), \log_e \lambda + z_{1-\alpha/2} \cdot SE(\log_e \lambda))$$

missä $z_{1-\alpha/2}$ on normaalijakauman prosenttipiste (esim. 95 %:n välissä $z_{0,975}=1,96$) ja

$SE(\log_e \lambda) = \frac{1}{\sqrt{d}}$, missä d on tapahtumien määrä seuranta-aikavälillä. Parametrin λ väli

saadaan ottamalla eksponenttimuunnos ($\exp(\cdot)$), eli logaritmfunktion käänteismuunnos, yllä olevan välin päätepisteistä.

Parametrin λ luottamusvälin perusteella voidaan laskea myös luottamusväli elossaolotodennäköisyyksille kaavalla:

$$S(t) = \exp(-\lambda \cdot t)$$

Esim.

Paksunsuolen syöpätutkimus (McIllmurray ja Turkie 1987). Seurattavia potilaita oli 24, joista 12 henkilöä kuoli neljän vuoden seurannan aikana. kuolleilla yhteenlaskettu seuranta-aika oli 180 kk ja elossa olevilla 251 kk, joten kokonaisseuranta-aika oli 431 kk. Parametrille λ saadaan siten arvio: $\lambda = 12/431 = 0,0278$, joten $\log_e \lambda = -3,5827$, $d=12$, joten $SE(\log_e \lambda) = 0,2887$. Sijoittamalla nämä arvot yllä olevaan kaavaan, saadaan väli $(-4,1485, -3,0169)$ ja ottamalla eksponenttimuunnos välin päätepisteistä saadaan: $\exp(-4,1485) = 0,0158$, $\exp(-3,0169) = 0,0490$ λ :n 95 %:n luottamusväliksi.

Huom. Tämä väli ei ole symmetrinen piste-estimaatin $\lambda = 0,0278$ suhteen.

Jos halutaan tarkastella elossaolotilannetta vuoden kohdalla, $S(t)$:n lausekkeeseen sijoitetaan $t = 12$ ja λ :lle arvot 0,0278, 0,0490, 0,0158. Tällöin $S(12) = 0,7163$ ja 95 %:n luottamusväliksi saadaan **(0,5554, 0.8273)**. Eli elossa olevien odote on noin 72 % ja 95 %:n varmuudella elossa olevien määrä on välillä (56 %, 83 %).

Huom.

Johtuen negatiivisesta eksponentista $S(t)$:n lausekkeessa λ :n alarajan perusteella on laskettu S :n yläraja.

Altistus- ja riskisuhteen luottamusvälit

Käyttötilanne:

Altistus- tai riskisuhteen arviointi tapaus-verrokki- tai ilmaantuvuus- (seuranta) tutkimuksissa.

Huom.

Luottamusvälit antavat arviointipohjaa tässä luvussa tarkasteltaville epidemiologisille suureille ainoastaan huomioimalla otantavaihtelua. Ne eivät kontrolloi millään tavalla valikoitumisesta, virheellisestä verrokkien valinnasta tai aineiston keräämisessä syntyviä harhoja, joita epidemiologisissa tutkimuksissa yleisesti esiintyy.

Tapaus-verrokkiasetelmassa käytetään taudin ja altisteen välisen riippuvuuden mittana **altistussuhdetta** ("odds ratio", OR), josta käytetään myös nimitystä **ristitulosuhde** sen laskentatavasta johtuen. Oletetaan, että tulokset on esitetty seuraavasti:

Tapaus-verrokki-tutkimus	Altiste Tupakointi		Yhteensä
	Kyllä	Ei	
Tauti Tapaus	a	b	a+b
Verrokki	c	d	c+d
Yhteensä	a+c	b+d	n

Tällöin OR määritellään suhteena: $OR = \frac{a \cdot d}{b \cdot c}$

OR:n luottamusväli voidaan laskea monella eri menetelmällä. Ne eroavat toisistaan laskennallisen helppouden ja tarkkuuden suhteen. "Logit-menetelmään" perustuva OR:n $100 \cdot (1-\alpha) \%$:n väli lasketaan $\log_e OR$:n luottamusvälin avulla seuraavasti:

$$\text{Alaraja} : e^{\log_e OR - z_{1-\alpha/2} \cdot SE(\log_e OR)} \quad \text{Yläraja} : e^{\log_e OR + z_{1-\alpha/2} \cdot SE(\log_e OR)}$$

Kaavassa $z_{1-\alpha/2}$ on normaalijakauman prosenttipiste. Esim. 95 %:n välissä $z_{0.975}=1,96$ (Excel tai StaTable) ja $\log_e OR$:n keskivirhe lasketaan kaavalla:

$$SE(\log_e OR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

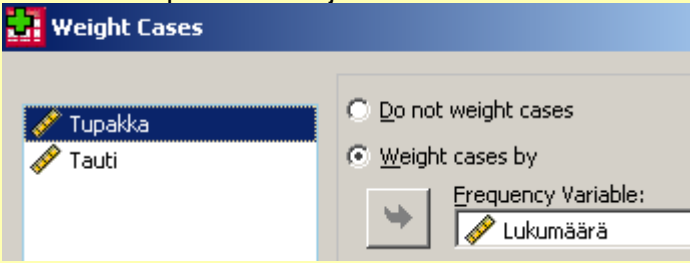
Esim.

Tupakanpolton yhteys keuhkosityöpään. Aineistona 709 sairaalahoidossa olevaa keuhkosityöpäpotilasta ja 709 sairaalahoidossa olevaa verrokkiä. (Doll & Hill, 1954). Tehtävänä on arvioida tupakanpoltosta aiheutuva keuhkosityöpäriski.

		Tupakka		Total
		Kyllä	Ei	
Tauti	Tapaus	688	21	709
	Verrokki	650	59	709
Total		1338	80	1418

Tupakka	Tauti	Lukumäärä
1	1	1
1	2	2
2	1	1
2	2	2

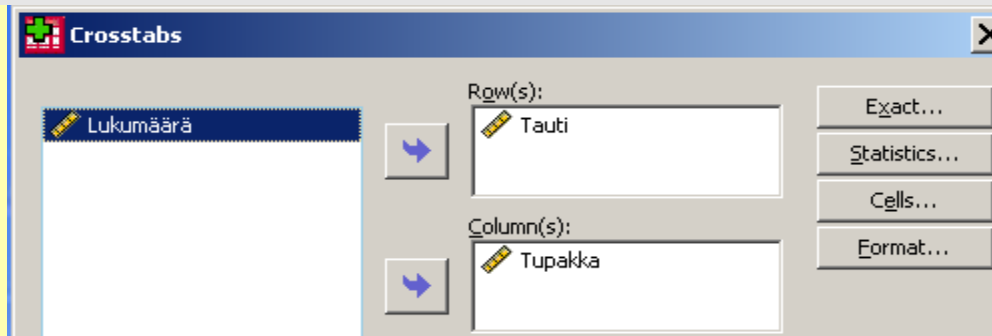
Kohdassa "Data" ► "Weight Cases" lukumäärä määritellään painomuuttujaksi.



Koska alkuperäistä aineistoa ei ole käytettävissä, niin aineisto pitää tallentaa SPSS:n Data Editorille tiivistetyssä muodossa, missä rivi- ja sarakeindikaattoreina ovat tupakka ja tauti ja painomuuttujana on lukumäärä.

SPSS: Valikot: "Analyze" ► "Descriptive Statistics" ► "Crosstabs"

Muuttuja tauti viedään rivimuuttujaksi ja tupakka sarakemuuttujaksi. Kohdasta "Statistics" valitaan "Risk"

**Tulos:****Risk Estimate**

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Tauti (Tapaus / Verrokki)	2,974	1,787	4,949
N of Valid Cases	1418		

CIA:
Tulkinta:

Ristitulosuhteelle (OR) saadaan arvio **2,97** ja sen 95 % luottamusväliksi tulee (1,79, 4,95). Arvo 1 jää selvästi tämän välin ulkopuolelle, mikä on osoitus riskin tilastollisesta merkitysvyydestä.

Mikäli tauti on harvinainen (ilmaantuvuus alle 10 %), OR:n avulla voidaan arvioida myös **riskisuhdetta (RR)**. Luotettavimmin riskisuhde arvioidaan kuitenkin seurantatutkimuksista. Riskisuhteen luottamusväli voidaan laskea seuraavasti:

a) Ilmaantuvuustiedot

	Altistuneet	Altistumattomat	Yhteensä
Tapaukset	a	b	m ₁
Henkilövuodet	n ₁	n ₀	t

b) Kumulatiivinen ilmaantuvuus tai vallitsevuus

	Altistuneet	Altistumattomat	Yhteensä
Tapaukset	a	b	m ₁
Muut	c	d	m ₀
Yhteensä	n ₁	n ₀	t

Riskieron arvio sekä tapauksessa a) että b) lasketaan kaavalla: $RD = \frac{a}{n_1} - \frac{b}{n_0}$

RD:n 100·(1-α) %:n likimääräinen khi-testisuureeseen perustuva luottamusväli on (Miettinen 1974, 1976):

$$\text{Alaraja : } RD \cdot \left(1 - \frac{z_{1-\alpha/2}}{\chi}\right), \quad \text{Yläraja : } RD \cdot \left(1 + \frac{z_{1-\alpha/2}}{\chi}\right)$$

Kaavassa $z_{1-\alpha/2}$ on normaalijakauman prosenttipiste (esim. 95 %:n välissä $z_{0,975}=1,96$ ja 90 %:n välissä $z_{0,95}=1,645$) ja

Riskisuhteen arvio sekä tapauksessa a) että b) lasketaan kaavalla: $RR = \frac{a/n_1}{b/n_0}$

RR:n 100·(1-α) %:n likimääräinen khi-testisuureeseen perustuva luottamusväli on (Miettinen 1974, 1976):

$$\text{Alaraja : } RR^{1 - z_{1-\alpha/2}/\chi}, \quad \text{Yläraja : } RR^{1 + z_{1-\alpha/2}/\chi}$$

tapauksessa a)

$$\chi = \frac{a - m_1 n_1 / t}{\sqrt{m_1 n_1 n_0 / t^2}} \quad (\text{binomiaalinen malli})$$

tapauksessa b)

$$\chi = \frac{a - m_1 n_1 / t}{\sqrt{m_1 m_0 n_1 n_0 / (t^2 (t - 1))}} \quad (\text{hypergeometrinen malli})$$

Esim.

Helikobakteeri-infektion esiintyvyys esikouluikäisillä lapsilla riippuen äidin gastriitti- tai mahahaavahistoriasta (Brenner et al.1998)

		Aidillä mahaava		Yhteensä
		Kyllä	Ei	
Lapsella <i>H. pylori</i> -infektio	Kyllä	6	112	118
	Ei	16	729	745
Yhteensä		22	841	863

$$RR = (6/22) / (112/841) = 2.05,$$

$$\chi = \frac{6 - 118 \cdot 22 / 863}{\sqrt{118 \cdot 745 \cdot 22 \cdot 841 / (863^2 (863 - 1))}} = 1,8797$$

95 %:n luottamusväli:

$$\text{Alaraja} : 2,05^{1 - 1,96/0,5320} = 0,9698,$$

$$\text{Yläraja} : 2,05^{1 + 1,96/0,5320} = 4,3243$$

Nämä välit voidaan laskea Excel-ohjelmalla osoitteessa:
http://www.mv.helsinki.fi/home/sarna/stats/RR_CL.xls)

RR:n luottamusväli					
Lopputulos		Altistus			
		Kyllä	Ei		
Kyllä		6	112	118	
Ei		16	729	745	
		22	841	863	
Exp		3,0081			
SD		1,5917	1,7121		
$\chi =$		1,8797			
RR =		2,0479		Alaraja	Yläraja
		95 %:n luottamusväli		0,9698	4,3243
z		1,9600			

Tulkinta: Tulos antaa viitteitä altisteen ja Hp-statuksen välisestä yhteydestä, joskin RR:n arvo yksi sisältyy niukasti saatuun väliin. CIA ja SPSS laskevat välin eri menetelmällä ja tuloksena on kapeampi väli: (1,01, 4,14).

Viitteet

Armitage P & Berry G. Statistical methods in medical research (2nd edition), Blackwell Scientific Publications, 1987.

Brenner H., Rothenbacher D, Bode G, Adler G. Parental history of gastritis or duodenal ulcer and prevalence of *Helicobacter pylori* infection in preschool children: population based study. BMJ 1998; 316:665.

Cappuccio FP et al. Lack of the effect of oral magnesium on high blood pressure. A double blind study. BMJ 291: 235-8, 1985

CIA 2.1.1: (**C**onfidence **I**nterval **A**nalysis) on hyvä ja helppokäyttöinen erityisohjelma luottamusvälien laskentaan ja satunnaistamislislojen tekemiseen. Ohjelma on oheistuotteena kirjassa: **Statistics with Confidence**, Toim. Altman DG, Machin D, Bryant TN & Gardner MJ. BMJ books 2000. ISBN: 0-7279 -375 -1. Hinta £22.50. http://www.bmjbookshop.com/shop/product_display.asp?SiteLanguage=ENG&ProductID=0727913751 Ohjelmapäivityksiä voi tilata osoitteesta: <http://www.som.soton.ac.uk/cia/>

Curran-Everett D, Taylor S, Kadafar K. Fundamental concepts in statistics: elucidation and illustration. J Appl Physiol 1998; 85:775-86.

Doll R & Hill AB. Mortality of doctors in relation to their smoking habits. A preliminary Report. BMJ 1954;1:1451-5.

Eason J, Markowe HLJ. Controlled investigation of deaths from asthma in hospitals in the North East Thames region. BMJ 1987;294:1255-8.

Isomäki H, Martio J, Kamela K et al. Comparison of analgesic effect of ten nonsteroidal anti-inflammatory drugs. Br J Rheumatol 1984;23:61-5.

Lipid Research Clinics Program. The lipid research clinics coronary prevention trial results. JAMA 1984; 251: 351-74.

McIllmurray MB, Turkie W. Controlled trial of gamma linolenic acid in Dukes' C colorectal cancer BMJ 1987; 294:1260 and BMJ 1987;295:475.

Medical Research Council Working Party. MRC trial of treatment of mild hypertension. BMJ 1985; 291: 97-104.

Miettinen OS. Simple interval estimation of risk ratio. Am J Epidemiol 1974; 100: (abs) 515-6.

Miettinen OS. Estimability and estimation in case-referent studies. Am J Epidemiol 1976; 103: 226-35.

Nam J. Confidence limits for the ratio of two binomial proportions based on likelihood scores: non-iterative method. Biom J 1995; 37: 375-9.

Newcombe RG, Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med 1998; 17: 873-90.

Peltola H, Mäkelä PH, Käyhty H et al. Clinical Efficacy of Meningococcus Group A Capsular Polysaccharide Vaccine in Children Three Months to Five Years of Age. N Eng J Med 1977; 297: 686- 691.

Tukey JW. Explorative Data Analysis. Reading, Mass. Addisin Wesley, 1977.

Wilson EB. Probable inference, law of succesion, and statistical inference. J Am Stat Assoc 1927; 22: 209-12

Tilastollisten testien valinta ja toteuttaminen

Hypoteesien testaaminen

Hypoteesien testaamisessa voidaan erottaa kaksi perustyyppiä:

- yksi testattava suure
- simultaaniset testaukset; monta samanaikaisesti testattavaa suuretta

Hypoteesien testauksessa on kyse valinnasta kahden toisensa poissulkevan hypoteesin välillä:

- **nollahypoteesi** (H_0 , "null hypothesis")
- **vaihtoehtoinen hypoteesi** (H_A , "alternative hypothesis")

Nollahypoteesi (H_0) "null hypothesis" on ennalta määritelty perusväittämä, jonka hyväksymiseen tai kumoamiseen tutkimuksessa pyritään. Esim., että vertailtavien lääkehoitojen A ja B tehon välillä ei ole eroa.

Vaihtoehtoinen hypoteesi (H_A) "alternative hypothesis" on ennalta määritelty vaihtoehtoinen väittämä nollahypoteesille. Esim., että lääkehoitojen teholla on eroa, mutta ei määritellä suuntaa (kaksisuuntainen hypoteesi) tai että hoito A on tehokkaampi kuin hoito B (yksisuuntainen hypoteesi). Tavallisimmin tutkimuksissa käytetään kaksisuuntaista hypoteesia, koska etukäteen ei yleensä voida olla täysin varmoja tuloksen suunnasta.

Tämä päättelyprosessi päättyy ratkaisuun, joka on kaksiarvoinen: a) joko siten, että katsotaan tutkimuksen antavan näyttöä hoitoveroista, jolloin päätetään hyväksyä vaihtoehtoinen hypoteesi tai b) siten, että tutkimuksen perusteella ei saada riittävää näyttöä hoitoveroista, eli nollahypoteesi jää voimaan. H_0 :a ei voida koskaan tilastollisin keinoin todistaa oikeaksi. Oikea johtopäätös päädyttyä siihen, että H_0 jää voimaan, on esimerkiksi "tutkimuksessa ei voitu osoittaa hoitojen A ja B välille tilastollisesti merkitsevää eroa."

H_A voi olla joko **kaksi-** tai **yksisuuntainen**; "on eroa" tai "on ero johonkin tiettyyn suuntaan". Valtaosassa kliinisiä vertailevia tutkimuksia vaihtoehtoinen H_A on kaksisuuntainen, koska etukäteen ei voida olla täysin varmoja, ettei jokin tietty hoito voisi olla esim. lumehoitoa haitallisempikin, ainakin jossakin tutkimuksen kohteeksi valitussa potilasjoukossa. Yksisuuntaista testausta käytetään mm. bioekvivalenssitutkimusten yhteydessä testattaessa siirrettyä nollahypoteesia ("shifted null hypothesis"). Siirron määrää tällöin tutkijan määrittelemä bioekvivalenttisuuden raja-arvo. Yksisuuntaista testausta voidaan käyttää myös silloin, hypoteesin toinen suunta ei ole mielekäs tai testauksen kannalta relevantti. Esim. aiheuttaako runsas kännykän käyttö aivot toiminnan häiriöitä? Joskus yksisuuntaista testausta näkee selvästi käytettävän siitä syystä, että kaksisuuntainen testi ei ole antanut tilastollisesti merkitsevää tulosta ($P < 0,05$). Tämä menettely on luonnollisesti vastoin tieteen sääntöjä. Yksinkertaista ja yleispätevää ohjetta yksi- tai kaksisuuntaisen testin käytöstä ei kuitenkaan voi antaa. Taloudelliset syyt saattavat puoltaa yksisuuntaisen testin käyttöä, koska silloin selvittää paljon pienemmällä potilasmäärällä.

Käytännössä tilastollinen testaus ei anna koskaan täysin varmaa päätöstä kumpaankaan suuntaan, vaan päätöksiin sisältyy virhemahdollisuus. Virheitä on kahta tyyppiä:

α -virhe, hylkäämisvirhe, eli I-lajin virhe: H_0 hylätään, vaikka se todellisuudessa pitää paikkansa. Esim. tehdään johtopäätös, että eri hoitojen välillä on eroa, vaikka sitä todellisuudessa ei ole. Syynä voi olla esimerkiksi sattuma tai väärä testi.

β -virhe, hyväksymisvirhe, eli II-lajin virhe: H_0 jää voimaan, vaikka se onkin väärä ja pitäisi siten hylätä. Esim. tehdään johtopäätös, että eri hoitojen välillä ei ole eroa, vaikka sitä todellisuudessa on. Syynä voi olla riittämätön aineistokoko tai väärä testi.

α -virhe on sama kuin **tilastollinen merkitsevyystaso** ("statistical significance level"). α -virhe voi olla yksi- tai kaksisuuntainen. Aineistokoolla ei ole vaikutusta α -virheen esiintymismahdollisuuteen. Testin voimakkuus on β -virheen funktio $1 - \beta$. Se arvioidaan ensisijaisesti tutkimuksen suunnitteluvaiheessa tai toissijaisesti tutkimustulosten luotettavuutta tarkasteltaessa. β -virhe on aina yksisuuntainen.

Edellä esitetyssä hoitokoe-esimerkissä päätöksentekoon liittyvät virheet voidaan esittää seuraavanlaisena nelikenttänä:

		Todellinen tilanne Hoitoero	
		Kyllä	Ei
Päätös tutkimuksen perusteella	H_0 hylätään ja hyväksytään H_A (hoitoeroa on)	Voima $1 - \beta$	I - lajin virhe α
	H_0 :aa ei voida hylätä (ei hoitoeroa)	II - lajin virhe β	$1 - \alpha$

Kun testattavassa suureessa, esimerkiksi kliinisten vertailtavien tutkimusten lopputulosmuuttujassa, todetaan ero, se voi johtua

- satunnaisvaihtelusta,
- satunnaisvaihtelusta ja hoitoerosta, tai
- satunnaisvaihtelusta ja harhasta

Tilastollisessa tutkimuksessa pyritään tutkimaan voiko todettu ero selittyä satunnaisvaihtelulla. Harhan mahdollisuuksia pyritään välttämään mm. valitsemalla vertailuryhmä oikein.

Monivertailut

Monivertailutestaustilanne syntyy silloin, kun

- päälopputulospäättäjän ohella vertaillaan **useita** muita **lopputulospäättäjia** (jotka on valittu joko ad hoc tai post hoc), joiden testaamiseen ei voimalaskelmissa ole varauduttu
- kun vertaillaan pareittain tuloksia **useina ajankohtina** tai samana ajankohtana **useiden ryhmien** välillä
- tehdään **osaryhmäanalyysijä**
- tehdään **väliallyysijä**.

Monivertailutestaustilanne kasvattaa α -virheen mahdollisuutta ja siksi

monivertailutesteissä ("multiple comparison test") suoritetaan α -virheen korjaus.

Merkitään K :lla vertailujen lukumäärää, α :lla sitä merkitsevyystasoa, jolla tutkimus aiotaan toteuttaa ja α^* :llä todennäköisyyttä virheellisesti hylätä ainakin yksi kaikista mahdollisista nollahypoteeseista. Jos vertailut ovat toisistaan riippumattomia, vallitsee α :n ja α^* :n välillä yhteys seuraavasti:

$$\alpha^* = P(\text{"monivertailussa ainakin yksi merkitsevä tulos vaikka } H_0 \text{ on tosi"}) = 1 - P(\text{"hyväksytään kaikki } H_0\text{:t} \mid \text{ kaikki } H_0\text{:t ovat tosia"}) = 1 - (1 - \alpha)^K.$$

Esim.

K lopputulosmuuttujaa, merkitsevyystaso $\alpha = 0,05$. Sijoittamalla yllä olevaan kaavaan $K = 1, 2, \dots, 10$ saadaan:

K	α^*	K	α^*
1	0,05	6	0,26
2	0,10	7	0,30
3	0,14	8	0,34
4	0,19	9	0,37
5	0,23	10	0,40

Taulukosta voidaan todeta, että α -virheen mahdollisuus kasvaa nopeasti lopputulosmuuttujien lukumäärän lisääntyessä. Kun α on pieni, niin $(1 - \alpha)^K = 1 - K \cdot \alpha$. Sijoittamalla tämä edellä olevaan kaavaan todetaan, että $\alpha^* = K \cdot \alpha$. Jos halutaan säilyttää monivertailutilanteessa alkuperäinen merkitsevyystaso, pitäisi P -arvoja verrata tasoon, joka on jaettu vertailujen lukumäärällä. Tämä nk. **Bonferronin korjaus** (Bonferroni CE, 1936) korjaa liikaa eli on konservatiivinen, koska lopputulosmuuttujat yleensä käytännössä korreloivat keskenään, jolloin edellä oleva α :n ja α^* :n välinen yhteys pidä paikkansa.

Bonferronin korjaus "Bonferroni correction" on monivertailujen yhteydessä käytetty P -arvojen korjausmenettely, jonka tavoitteena on säilyttää alkuperäinen tilastollinen merkitsevyystaso (α -virhe) alun perin suunnitellun suuruisena monivertailuista huolimatta. Tavallisimmin käytetty taso on 0,05 (5 %). Bonferronin korjauksella korjattu P -arvo saadaan kertomalla testin antama P -arvo vertailujen lukumäärällä. Tällä korjauksella on taipumus ylikorjata, koska se olettaa monivertailut toisistaan riippumattomiksi, mitä ne eivät yleensä käytännössä ole.

Ainoa oikea tapa käsitellä monivertailuongelmaa on käyttää sellaisia tilastollisia menetelmiä ja testejä, jotka on suunniteltu tähän tarkoitukseen: toistomittausten ANOVA, Friedmanin testi, Cochranin Q-testi, erilaiset **monivertailutestit**.

Kliinisten tutkimusten suunnitelmissa tulisi ilmoittaa mahdollisimman tarkasti, mitä hypoteeseja on tarkoitus testata ja kuinka testit tai erilaiset vertailut tullaan suorittamaan (**suunnitellut vertailut**, "ad hoc comparisons"). Usein varsinaisten hypoteesien lisäksi tutkimusprosessin analyysivaiheessa kehitellään lisähypoteeseja, valitaan niihin soveltuvat testit ja vertailutavat (**jälkikäteisvertailut**, "post hoc comparisons"). Nämä testaukset tulkitaan kuitenkin suoritettun tutkimuksen kannalta toisarvoisiksi. Lähinnä niillä on merkitystä jatkotutkimusten kehittelyn kannalta.

Parametrinen ja parametriton testi

Tilastollisia testejä on kahta päätyyppiä:

Parametrinen testi ”parametric test” perustuu johonkin teoreettiseen malliin ja sen sisältämien tuntemattomien parametrien arvioimiseen tutkimusaineiston perusteella. Esimerkiksi t-testi on parametrinen testausmenetelmä, joka perustuu Studentin t-jakaumaan.

Parametriton testi ”non-parametric test” ei perustu mihinkään jakaumaoletuksiin, vaan siinä tarvittavat suureet lasketaan suoraan havaintoaineistosta esim. muodostamalla suhteellisia osuuksia, niiden tuloja jne. (esim. Kaplan-Meier -menetelmä) tai muodostamalla järjestyslukuja sekä niiden perusteella johdettuja suureita (esim. Wilcoxon–Mann–Whitney'n testisuure). Parametrittomia menetelmiä käytetään etenkin pienissä ja/ tai hankalissa havaintoaineistoissa, esim. kun aineisto sisältää poikkeavia havaintoarvoja. Niiden voima on lähes sama (noin 95 %) kuin parametrinen menetelmien silloin, kun parametrinen menetelmien käyttöedellytykset ovat voimassa.

Parametrittomien menetelmien käyttöindikaatioita ovat

- mittaukset järjestysasteikolla
- aineisto pieni (alle 20)
- jakaumat monihuippuisia ja/tai epäsymmetrisiä.

P-arvon käyttöön liittyviä rajoitteita ja ongelmia

P-arvon määrittely edellyttää käsitteiden testisuure ja sen otosjakauma määrittelyä. Testisuureella tarkoitetaan mitä tahansa nollahypoteesin testaamiseen käytettyä otossuureta, otoksen havaintoarvojen funktiota. Testisuureen otosjakauma on kaikkien mahdollisten testisuureen saamien arvojen muodostama jakauma. **P-arvo** määritellään seuraavasti:

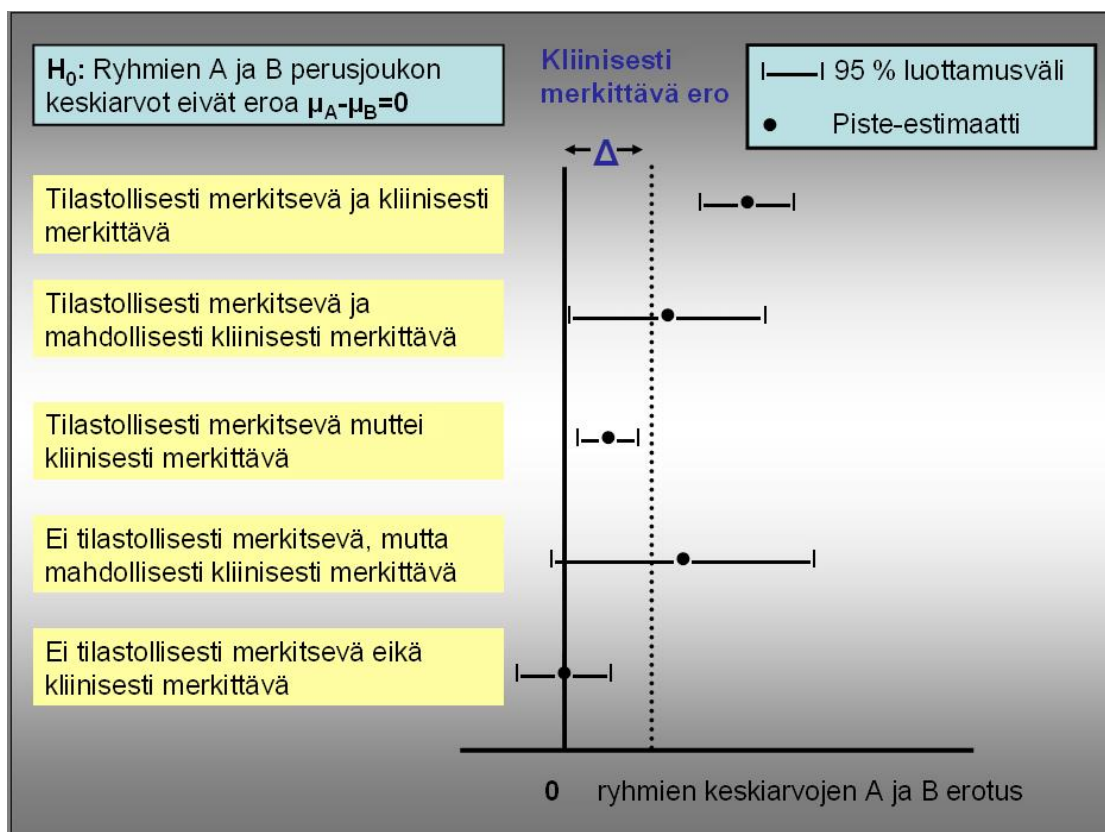
P- arvo ”P- value” on havaintoaineiston ja käytetyn testisuureen otosjakauman perusteella laskettu todennäköisyys saada lopputulos, joka on **vähintään yhtä epätodennäköinen (harvinainen) kuin tutkimuksessa todettu lopputulos edellyttäen, että todellisuudessa nollahypoteesi (H_0) olisi tosi**. P-arvo liittyy ainoastaan H_0 :n testaamiseen, eikä kerro esimerkiksi hoitoerosta mitään, toisin kuin luottamusväli.

P-arvoja on lääketieteellisissä julkaisuissa käytetty ylenpalttisesti ja niiden käyttöön on liittynyt paljon mystiikkaa ja harhakäsityksiä. Koska P-arvo todennäköisyys, niin se voi teoriassa saada mitä tahansa arvoja väliltä [0 -1]. Se ei ole todennäköisyys, että nollahypoteesi on oikein eikä se ole todennäköisyys saada täsmälleen todettu tutkimustulos vaan kuten määritelmästä ilmenee, P-arvoon lasketaan lisäksi mukaan myös todettua tulosta harvinaisemmat hypoteettiset lopputulokset edellyttäen, että nollahypoteesi olisi tosi. Näihin asioihin on viime vuosina kiinnitetty erityistä huomiota tieteellisten lehtien toimituksissa ja tutkijoille suunnatuissa kirjoitusohjeissa.

P-arvo ei suoranaisesti liity aineistokokoon; pieni P-arvo voidaan saavuttaa niin aineistoltaan pienessä kuin suuremmassa tutkimuksessa. Kliinisissä tutkimuksissa mm. hoitoero on ratkaisevassa asemassa. Pienen eron löytäminen edellyttää yleensä isoa Kliinisen biostatistiikan kurssi

aineistoa, sen sijaan suuri ero voidaan todeta pienemmälläkin aineistolla (ks. luku "Aineistokoon arviointi"). Millään tilastollisella testillä ei kuitenkaan yleensä saavuteta maagisena pidettyä tasoa $P < 0,05$, jos ryhmäkoko jää alle kuuden. Oikein pienillä aineistoilla kannattaa yleensä käyttää satunnaistamis- eli permutaatiotestejä (esim. StatXact 8).

Jonkin todetun eron biologinen tai lääketieteellinen merkitsevyys on eri asia kuin tilastollinen merkitsevyys. Siksi artikkeleissa kannattaa käyttää jälkimmäisestä termiä "statistically significant". Kliinisen merkitsevyyden arviointia helpottaa, kun ei tyydytä pelkkään P-arvoon, vaan lasketaan keskeisimmille lopputulosmuuttujille myös luottamusvälit. Vaikka P-arvo ei antaisikaan esim. ryhmien välille tilastollisesti merkitsevää hoitoeroa, voi luottamusväli paljastaa, että käytännön kannalta eroa saattaisi olla, mikä näkyy esim. välin kapeutena ja selvänä epäsymmetrisyytenä nollan suhteen. Isoilla aineistoilla kliinisesti merkityksettömätkin erot saattavat tulla tilastollisesti erittäin merkitseviksi, esim. $P < 0,0001$.



On huomattava, että vaikka tilastollisesti merkitsevän ja ei-merkitsevän tuloksen raja vakiintuneen käytännön mukaan vedetään tasolle 0,05, ei käytäntö perustu mihinkään matemaattiseen tosiseikkaan. Se on täysin keinotekoinen valinta ja aiheuttaa paljon harhakäsityksiä. Esim. tutkija saattaa raportoida P-arvoon 0,051 päätyneen tuloksen tilastollisesti ei-merkitseväksi, kun taas joku toinen tutkija samankaltaisesta tutkimuksesta saamansa tuloksen $P = 0,049$ tilastollisesti merkitseväksi. Jos tutkijat eivät ilmoita P-arvoa, ei lukijalla ole mahdollista arvioida tulosten merkitystä. Koko tilastollisen testauksen periaatteet on täysin väärin ymmärretty, mikäli näillä kahdella tuloksella katsottaisiin olevan jotain eroa käytännön kannalta. On syytä muistaa, että eri ohjelmienkin tuottamat P-arvot samasta testistä voivat vaihdella jonkin verran, ja joskus jopa paljon, riippuen siitä, minkälaista laskenta-algoritmia mikäkin ohjelma käyttää.

Käsitteitä melkein merkitsevä ("almost significant"), merkitsevä ("significant") ja erittäin merkitsevä ("highly significant") ei nykysuositusten mukaan tulisi käyttää. Ne ovat peräisin taulukoiden laadinnasta ja antavat P-arvon luonteesta väärän kuvan. Useimmissa tilanteissa P-arvo kannattaa antaa ns. tarkkana, siten että siinä on korkeintaan neljä desimaalia ja kaksi merkitsevää numeroa, esim. $P = 0,0032$, $P = 0,0002$ tai $P < 0,0001$. Tätä pienemmät P-arvot eivät ole mielekkäitä, koska P-arvo liittyy vain α -virheeseen. Jonkin geneettisen ominaisuuden esiintymistodennäköisyys voisi olla $3,2 \cdot 10^{-6}$ (eli 3,2/1 000 000). P-arvon kohdalla näin pieni arvo ei anna enempää varmuutta H_0 :n oikeellisuudesta kuin mitä arvo 0,0001 antaisi.

P-arvo on laskettu ehdolla, että H_0 on tosi, eli kyseessä on ehdollinen todennäköisyys $P(T | H_0)$, missä T on tutkimuksen lopputulos. Samaan lopputulokseen T voidaan kuitenkin päätyä monien muidenkin hypoteesien kuin H_0 :n vallitessa. Mikäli olisikin niin, että H_0 :n asemesta jokin vaihtoehtoisista hypoteeseista H_A olisikin tosi, niin saataisiin aivan eri P-arvoja. P-arvo vastaa itse asiassa aivan eri kysymykseen, mihin tutkija toivoisi ja usein luulee sen vastaavan. Tutkija tavoittelee yleensä todennäköisyyttä $P(H_0 | T)$, eli että onko H_0 tosi sen tiedon perusteella, mitä tutkimuksen lopputulos (T) tuotti. Tähän P-arvo ei vastaa. Todennäköisyydet $P(T | H_0)$ ja $P(H_0 | T)$ voivat olla hyvinkin erilaisia. Andersen (1984) esittää kirjassaan esimerkin: T=kuume, H_0 =umpisuolen tulehdus, $P(H_0 | T)$ voisi olla 0,01 ja $P(T | H_0)$ luokkaa 0,50, joten melkoinen ero,

Ehdollinen todennäköisyys "conditional probability" $P(A | B)$ ilmaisee tapahtuman A esiintymismahdollisuuden, mikäli ehto B on voimassa (tapahtuma B on tapahtunut); esim., jonkin sairauden ikäluokkakohtaiset esiintyvyydet, Tällöin tapahtuma A on sairauden esiintyminen ja tiettyyn ikäluokkaan kuuluminen on ehto B.

Ehdolliset todennäköisyydet $P(H_0 | T)$ ja $P(T | H_0)$ voidaan liittää toisiinsa kaavalla:

$$P(H_0 | T) = P(H_0) \cdot P(T | H_0) / P(T)$$

Tämä on nk., **Bayesin kaava**. $P(H_0 | T)$ on todennäköisyys, että H_0 pitää paikkaansa ehdolla, että tutkimuksessa on saatu lopputulos T. Tämä todennäköisyys on suoraan verrannollinen nollahypoteesin a priori todennäköisyyteen $P(H_0)$ ja P-arvoon. A priori todennäköisyys (ennakkotodennäköisyys) $P(H_0)$ perustuu tutkimuksen ulkopuoliseen ennakkotietoon.

Ennakkotodennäköisyys "a priori probability" on todennäköisyys, joka perustuu aikaisempaan tietoon. Esim. diagnostisissa testeissä todennäköisyys, joka kertoo kuinka paljon diagnosoitavaa tautia esiintyy tutkimuksen kohteena olevassa perusjoukossa (väestössä). Tätä ennakkokäsitystä korjataan kunkin diagnosoitavan potilaan kohdalla tekemällä diagnostisia testejä ja laskemalla niiden perusteella tutkittavan taudin jälkikäteis- (*a posteriori*) todennäköisyys, esim. Bayesin kaavaa käyttäen.

Bayesiläinen lähestymistapa "Bayesian approach" tarkoittaa menettelyä, missä a priori todennäköisyysjakaumien muodossa esitettyä ennakkotietoa tai -uskomusta tarkennetaan tutkimuksessa kerätyn tiedon avulla. Tieto, jolla ennakkotietoa korjataan, esitetään uskottavuus("likelihood")funktion muodossa. Näin saadaan nk. a posteriorinen todennäköisyysjakauma, jonka avulla voidaan arvioida tutkittavaa asiaa, esim. hoitoeron suuruutta vertailevissa tutkimuksissa.

Bayesiläiset menetelmät eivät toistaiseksi ole saavuttaneet suurta suosiota kliinisissä tutkimuksissa. $P(H_0)$:n arviointia on pidetty liian subjektiivisena. Toisaalta esim. aikaisemmista vastaavista tutkimuksista saatavan a priori tiedon hyödyntämättä jättäminenkin, kuten P-arvon laskemisessa tapahtuu, ei ole korrektaa. Osasyynä näiden

menetelmien vähäiselle käytölle on ollut helppokäyttöisten ohjelmistojen puuttuminen. Tilanne on tässä suhteessa kuitenkin korjaantumaan päin.

Tilastollisten testien valintaperusteet

Tilastollisten testien valintapäätös on yleensä melko monitahoinen asia. Testi tulisi valita siten, että se on riittävän herkkä toteamaan eroja jos niitä on olemassa, mutta toisaalta se ei saisi olla myöskään liian optimistinen. Ainakin alla esitettyihin asioihin pitää ottaa kantaa sopivaa testiä mietittäessä:

<p>Aineiston rakenne / testausasetelma</p> <ul style="list-style-type: none"> • Riippumattomat "independent" otokset/ryhmät • Riippuvat "related", "dependent" otokset/ryhmät • Kaltaistetut "matched" otokset/ryhmät
<p>Mitta-asteikko</p> <ul style="list-style-type: none"> • Välimatka/suhdeasteikko • Järjestysasteikko • Laatueroasteikko • Binomiaalinen asteikko
<p>Otosten/ryhmien lukumäärä</p> <ul style="list-style-type: none"> • Yksi • Kaksi • Yli kaksi
<p>Testauksen tarkoitus</p> <ul style="list-style-type: none"> • Heterogeenisuus • Trendi • Assosiaatio
<p>Aineiston koko</p> <ul style="list-style-type: none"> • "Kohtalainen" (yli 30 ryhmää kohden) • "Pienehkö" (10- 30 ryhmää kohden) • "Pieni" (alle 10 ryhmää kohden)
<p>Jakauma</p> <ul style="list-style-type: none"> • Normaali- / Log-normaali- • Binomi- • Poisson-

Monissa tilastopaketeissa on sisäänrakennettuna ohjelmamoduuleja "statistical adviser", "statistical navigator", "tutor", "guide", "wizard", jotka avustavat tutkijaa testi- tai menetelmävalinnoissa. Nämä ohjelmien valikot tai niiden esittämät kysymykset perustuvat mm. yllä esitettyihin valintaperusteisiin. Lisäksi usein niissä kerrotaan testien ja menetelmien käyttöedellytyksistä.

Seuraavissa taulukoissa on esitetty kliinisten tutkijoiden tavallisimmin käyttämiä perustestejä asteikkotyypeittäin eri testausasetelmissa.

1 muuttuja, 1 ryhmä, 1 tai 2 ajankohtaa tai kaltaistetut ryhmät:

Asteikko	Testausasetelma	Testi
Välimatka tai suhde	Otoksen tunnusluvun tai malliin liittyvän parametrin testaaminen	z-testi, t-testi
	Muutos keskiarvotasolla, kaksi ajankohtaa tai kaltaistetut parit ja isohko ryhmä	Parittainen t-testi
	Muutos (tai ero) mediaanitasolla 2 ajankohtaa (tai kaltaistetut parit)	Wilcoxonin parittainen testi*
	Heterogeenisuus, kaltaistetut parit	Fisherin satunnaistamistesti
	Yhteensopivuus, Normaalisuus	Kolmogorov–Smirnovin testi, Shapiro–Wilkin W-testi
Järjestys	Satunnaisuus	”Runs”-testi
	Muutos	Wilcoxonin parittainen testi*
	Yhteensopivuus	Kolmogorov–Smirnovin testi, Khii ² -yhteensopivuustesti
Laatuero	Tunnusluvun testaus (suhteellinen osuus, lukumäärä jne...), yhteensopivuus	Binomitesti, Poisson-testi, Khii ² -yhteensopivuustesti
	Muutosten heterogeenisuus (2 x 2)-taulukko	McNemarin testi
	Muutosten heterogeenisuus (r x r)-taulukko	Yleistetty McNemarin testi (”Marginal homogeneity”-testi)
	Heterogeenisuus, kaltaistetut parit	Bowkerin testi

*Tunnetaan myös ”parittaisena rank - sum”-, ”Wilcoxon-Pratt”- ja ”Wilcoxon signed rank”-testinä

1 muuttuja, 2 toisistaan riippumatonta ryhmää:

Asteikko	Testausasetelma	Testi
Välimatka tai suhde	Heterogeenisuus (pienet ryhmät)	Fisher-Pitmanin permutaatiotesti (satunnaistamistesti)
	Keskiarvojen ero (suuret ryhmät)	Riippumattomien ryhmien t-testi
Järjestys	Mediaanien ero	Wilcoxon–Mann–Whitneyn testi (”rank sum”-testi / U-testi)
	Suhteellisten osuuksien trendit: (r x c)-taulukot tai ositetut (r x c)-taulukot,	”Linear by linear”- testi, Cochran–Armitagen trenditesti, ”Normal scores”-testi, Wilcoxonin ”rank-sum”-testi
	Heterogeenisuus, rajatut havainnot	Logrank-testi, Wilcoxon–Gehanin testi
Laatuero	Heterogeenisuus (2x2)-taulukko, pienet ryhmät	Fisherin eksakti testi
	Heterogeenisuus (2x2)-taulukko, suuret ryhmät	Pearsonin Khii ² -testi
	Heterogeenisuus (r x c)-taulukko, pienet ryhmät	Fisher–Freeman–Haltonin testi, (yleistetty Fisherin eksakti testi)
	Heterogeenisuus (r x c)-taulukko, suuret ryhmät	LR-testi (uskottavuussuhdetesti) Pearsonin Khii ² -testi

	Heterogeenisuus ositettu (r x c)-taulukko (sekoittava tekijä mukana)	Mantel–Haenszelin testi
--	--	-------------------------

1 muuttuja, > 2 riippumatonta ryhmää, yksi luokitteleva tekijä:

Asteikko	Testausasetelma	Testi
Välimatka tai suhde	Keskiarvojen erot, suurehkot ryhmät	Yksisuuntainen varianssianalyysi (Anova), F-testi
	Keskiarvojen erot, pienet ryhmät	Satunnaistamistesti
	Trendit keskiarvoissa	Anova, kontrastit
Järjestys	Heterogeenisuus, mediaanien erot	Kruskal–Wallisin parametrin varianssianalyysi
	Trendit mediaaneissa, monotonisuus	Jonckheere–Terpstran trenditesti, "Linear by linear"- testi
	Trendit, rajattuja havaintoja	Tarone–Waren trenditesti
Laatuero	Suhteellisten osuuksien homogeenisuus tai jakaumien yhteensopivuus	G ² -testi, LR-testi, Khii ² -testi
	Kolmiulotteiset kontingenssi-taulukot (r x r x r)	Log-lineaariset mallit

1 muuttuja, > 2 riippumatonta ryhmää, kaksi luokittelevaa tekijää:

Asteikko	Testausasetelma	Testi
Välimatka tai suhde	Heterogeenisuus, keskiarvojen erot	Kaksisuuntainen varianssianalyysi ("two-way Anova"), F-testi
Järjestys	Heterogeenisuus, mediaanien erot	Kaksisuuntainen parametrin varianssianalyysi
Laatuero	Heterogeenisuus, spesifit hypoteesit, Assosiaatiot	Log-lineaariset mallit

1 muuttuja, > 2 toisistaan riippuvaa ryhmää:

Asteikko	Testausasetelma	Testi
Välimatka tai suhde	Heterogeenisuus, toistot (esim. eri ajankohtina), suuret ryhmät	Toistomittauksien varianssianalyysi ("Analysis with repeated measures"), F-testi
	Heterogeenisuus, toistot (esim. eri ajankohtina), pienet ryhmät	Satunnaistamistesti
	Heterogeenisuus, toistot (esim. eri ajankohtina), pienet ryhmät, kaltaistus	Pitman–Welchin testi

Järjestys	Heterogeenisuus, mediaanien erot, toistot, kaltaistetut henkilöt	Friedmanin kaksisuuntainen varianssianalyysi
	Trendit mediaaneissa, monotonisuus	Pagen trenditesti
Laatuero	Suhteellisten osuuksien heterogeenisuus	Cochranin Q-testi

Assosiaation (riippuvuuden) mitat:

Asteikko	Mitattava asia	Mitta
Välimatka tai suhde	Riippuvuus, muuttujat x, y (kaksiulotteinen normaalijakauma)	Pearsonin korrelaatiokerroin
Järjestys	Riippuvuus, muuttujat x, y	Spearmanin järjestyskorrelaatiokerroin
	Osittaiskorrelaatiomahdollisuus	Kendallin Tau
	Epäsymmetrinen riippuvuus, (r x s)-taulukko x:n ja y:n suhteen, x=ennustaja, y=ennustettava muuttuja	Somersin D, Gamma-kerroin
	Yhtäpitävyys, esim. useiden arvioijien välillä	Kendallin W (konkordanssikerroin)
Laatuero	Yhtäpitävyys, esim. samat koehenkilöt ja kaksi testiä tai havainnoijaa	Cohenin kappa Painotettu kappa
	Riippuvuus (r x c)-taulukot	Goodman–Kruskalin assosiaatiomitat

Tilastolliset testisuureet voidaan laskentatavan mukaan luokitella asymptoottisiin ja eksakteihin testeihin.

Asymptoottinen testi "asymptotic test" on testi, joka perustuu approksimaatioon, eli likimäämisarviointiin, Normaalijakaumalla tai jollain muulla todennäköisyysjakaumalla siten, että käytetty arvio tarkentuu, kun aineistokokoo n kasvaa. Synonyymi asymptoottiselle testille on suurten otosten testi "large sample test"

Eksakti testi "exact test" on oletettuun otosjakaumaan perustuva tilastollinen testi, jonka antama p-arvo on tarkka edellyttäen, että oletettu otosjakauma pitää paikkansa, esim. hypergeometriseen jakaumaan perustuva Fisherin eksakti testi. Pienillä aineistoilla kannattaa käyttää eksakteja testejä, sillä niiden antamat p-arvot ovat luotettavampia kuin asymptoottisilla testeillä saadut.

Normaalisuuden testaaminen

Monissa lääketieteellisissä aineistoissa muuttujien jakauman normaalisuuden tarkistaminen on aiheellista, koska hyvin usein normaalisuusoletus ei ole voimassa. Tällöin on syytä välttää sellaisia menetelmiä ja testejä, jotka voimakkaasti perustuvat

normaalisuuteen. Esimerkiksi paljon käytettyä t-testiä luonnehditaan yleensä melko robustiksi testiksi, mikä tarkoittaa, ettei se ole kovin herkkä poikkeamille niistä oletuksista, joihin testi perustuu. Poikkeavat havaintoarvot aiheuttavat kuitenkin herkästi t-testin nimittäjässä olevan keskivirheen kasvamisen, jolloin testistä tulee konservatiivinen, ylivarovainen.

Kaksi tavallisimmin käytettyä testiä normaalisuuden tarkistamiseen ovat **Shapiro-Wilkin W** ja **Kolmogorov–Smirnovin** testit.

Shapiro-Wilkin W-testi testaa havaitun jakauman poikkeavuutta määritellystä todennäköisyysjakaumasta, tavallisimmin normaali- tai eksponenttijakaumasta.

Kolmogorov–Smirnovin testi on parametrin yhteensopivuustesti minkä tahansa kahden jakauman välillä. Testi perustuu maksimaaliseen absoluuttiseen eroon vertailtavien kumulatiivisten jakaumien välillä. Kolmogorov–Smirnovin testi on yleensä voimaltaan heikompi kuin Shapiro-Wilkin testi.

Näistä jälkimmäistä voidaan käyttää yhteensopivuustestinä paitsi normaalijakauman niin myös monien muidenkin todennäköisyysjakaumien, kuten esim. tasaisen, Poisson ja binomijakauman suhteen. Testit antavat P-arvon, joka tulkitaan siten, että pieni P-arvo ($< 0,05$) merkitsee poikkeamaa testattavasta jakaumasta, esim. normaalijakaumasta. Shapiro-Wilkin testin parametrin arvioidaan automaattisesti havaintoaineistosta, joten testin käyttö on helppoa. Testin voima on todettu empiirisissä kokeissa hyväksi, mm. monille vinoille jakaumille, joita kliinisissä tutkimuksissa yleisesti esiintyy.

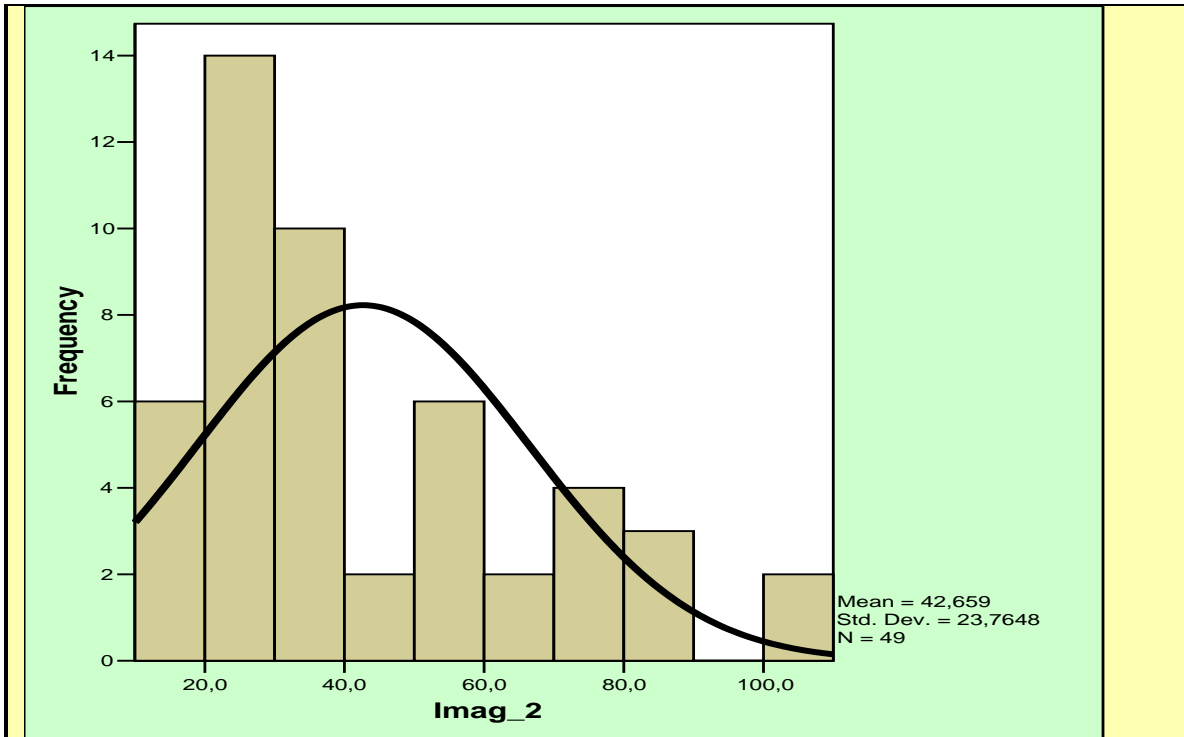
Esim. Tietokanta TRIAL. Ongelma: Onko muuttujan "Imag_2" jakauma normaalinen? **SPSS**-ohjelmalla saadaan seuraavat tulokset:

SPSS: Valikot: **"Analyze"** ► **"Descriptive Statistics"** ► **"Explore"**. Kohtaan **"Dependent list"** vietään tutkittava muuttuja **"Imag_2"**. Kohdasta **"Plots"** klikataan lokerot **"Normality plots with tests"** ja **"Histogram"**.

Tulos:

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Imag_2	,183	49	,000	,896	49	,000

a. Lilliefors Significance Correction

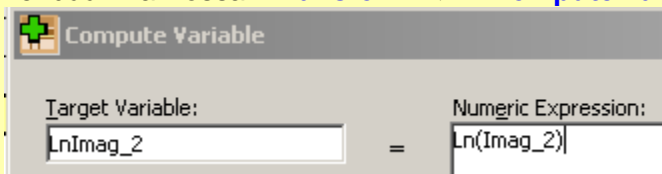


Normaalijakauman referenssikäyrä histogrammiin saadaan valikoista "Graphs" ▶ "Histogram", kun klikataan lokero "Display normal curve"

Tulkinta:

Molemmat testit antavat tuloksen $P < 0,05$, joten muuttujan Imag_2 jakauma poikkeaa Normaalijakaumasta. Histogrammin perusteella Imag_2:n jakauma muistuttaa Log-normaalijakaumaa.

Tehdään valikossa "Transform" ▶ "Compute variable" muunnos:



Tulos:

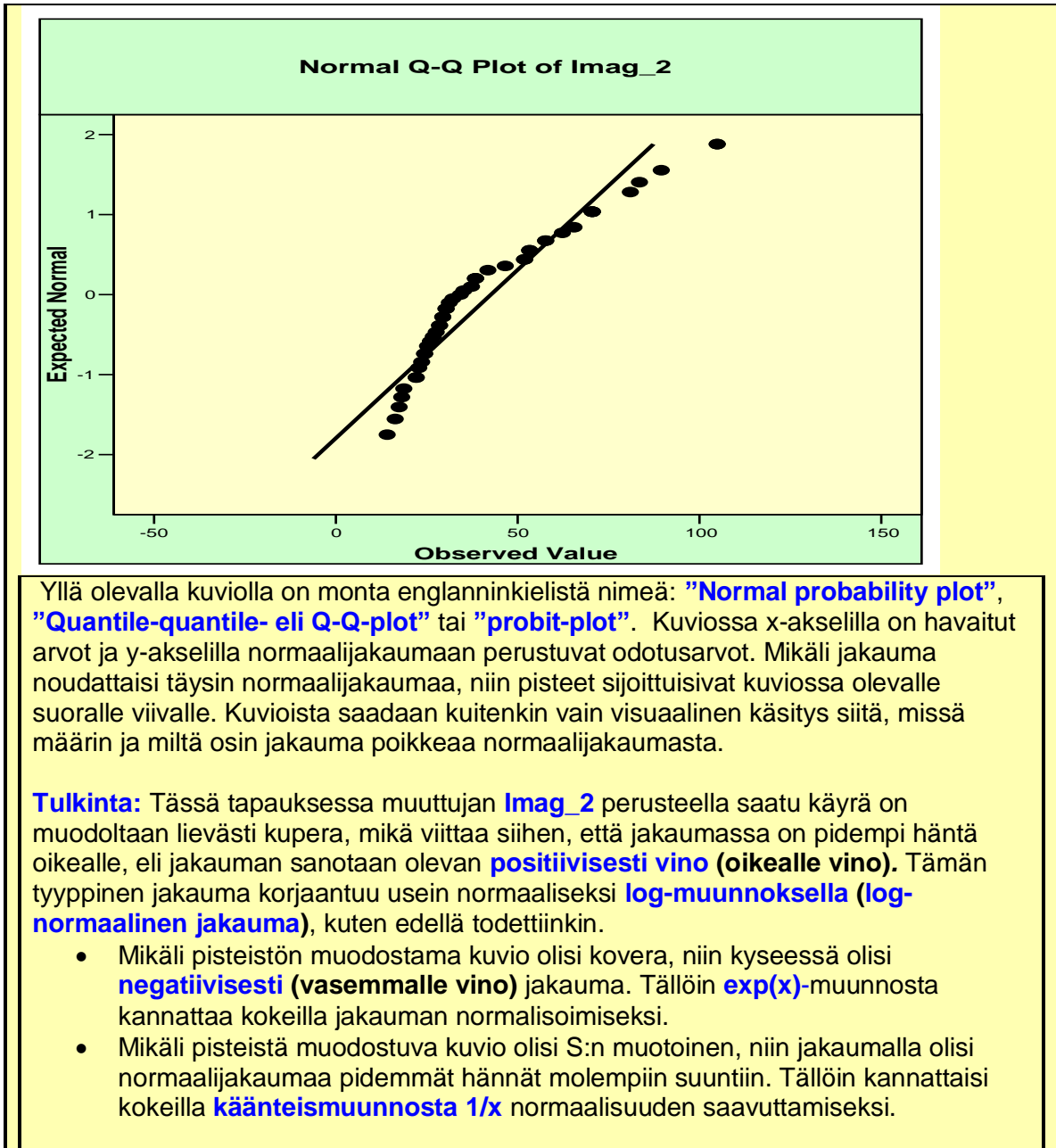
Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
LnImag_2	,091	49	,200*	,978	49	,473

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Tulkinta: Koska molemmissa testeissä $P > 0,05$, niin todetaan, että logaritmuunnos korjasi muuttujan "Imag_2" vinouden, joten muuttujalle "LnImag_2" Normaalisuusoletus on voimassa.



Jakauman vinoutta ja huipukkuutta voidaan kvantitatiivisesti verrata normaalijakaumaan **vinouuskertoimella** ("coefficient of skewness") g_1 ja **huipukkuuskertoimella** g_2 ("coefficient of kurtosis").

Mikäli a) $g_1 > 0$, b) $g_1 = 0$ tai c) $g_1 < 0$, niin jakauma on:

- positiivisesti vino (vino oikealle)
- symmetrinen
- negatiivisesti vino (vino vasemmalle).

Mikäli a) $g_2 > 0$, b) $g_2 = 0$ tai c) $g_2 < 0$, niin jakauma on:

- huipukkaampi kuin normaalijakauma ("**leptokurtic**"), eli jakaumassa on **enemmän** havaintoja sekä lähellä keskiarvoa että jakauman hännissä kuin normaalijakaumassa
- yhtä huipukas kuin normaalijakauma ("**mesokurtic**")

- c) litteämpi kuin normaalijakauma (**“platykurtic”**), eli jakaumassa **vähemmän** havaintoja sekä lähellä keskiarvoa että jakauman hännissä kuin normaalijakaumassa

SPSS:llä nämä suuret g_1 ja g_2 saadaan valikoista: **”Analyze” ▶ ”Descriptive Statistics” ▶ ”Explore”**.

Descriptives		
	Statistic	Std. Error
Skewness	1,018	,340
Kurtosis	,307	,668

Tarvittaessa voidaan myös laskea normaalijakaumaan perustuvat testit: $z = g_1 / SE(g_1)$ ja $z = g_2 / SE(g_2)$ jakauman poikkeavuudelle normaalijakaumasta. Kyseisessä esimerkkitapauksessa $g_1 = 1,018$, $SE(g_1) = 0,340$, $z = 2,994$, $P = 0,0028$ (StaTable) ja $g_2 = 0,307$, $SE(g_2) = 0,668$, $z = 0,4596$ ja $P = 0,6458$, joten muuttujan **Imag_2** jakauma on tilastollisesti merkitsevästi vinompi kuin normaalijakauma, mutta ei huipukkaampi.

Käytännössä tutkijoiden aineistot ovat usein niin pieniä, että normaalisuustestit eivät ole kovin voimakkaita ja siten jakaumien normaalisuudesta ei ole helppo varmistua. Silloin kannattaa usein harkita parametrittomia testivaihtoehtoja lähes oletusarvoisesti parametrusten asemesta, vaikka useimmat parametriset testit ovatkin nk. robusteja, ts. eivät ole herkkiä poikkeavuudelle normaalisuudesta. Parametriset testit ovat vain hivenen voimakkaampia kuin vastaavat parametrittomat silloin, kun jakauma on täysin normaali.

Varianssien homogeenisuustestit

Joissakin testeissä edellytetään vertailtavien ryhmien varianssien homogeenisuutta. Tätä voidaan tutkia mm. seuraavin testein:

Hartleyn F-max-testi

Perustuu kahden varianssin suhteeseen

Bartlettin testi

Sekä Hartleyn että Bartlettin testit ovat herkkiä poikkeamille normaalisuudesta. Hartleyn testi on lisäksi herkkä ryhmien erisuuruudelle.

Box'n testi

Testi soveltuu erityisesti isoille aineistoille

Levenen testi

Box ja Levene ovat melko robusteja testejä. Normaalijakaumillekin Levenen testi on lähes yhtä hyvä kuin Hartleyn tai Bartlettin testi.

Esim.

Tietokanta TRIAL. Ongelma: Onko muuttujan Imag_2 varianssi sama sekä miehillä että naisilla?

SPSS:llä varianssien homogeenisuus voidaan testata **Levenen** testillä valikoista: **"Analyze" ► "Compare means" ► "Independent Samples t test"** asettamalla testattava muuttuja **"Imag_2"** kohtaan **"Test Variable"** ja kohtaan **"Grouping Variable"** ryhmämuuttuja **"Gender"**

Tulos:

		Levene's Test for Equality of Variances	
		F	Sig.
Imag_2	Equal variances assumed	,223	,639

Tulkinta: Koska tämä arvo on selvästi suurempi kuin yleisesti kriittisenä arvona pidetty 0,05, niin varianssien homogeenisuusoletus jää voimaan.

Permutaatiotesti riippuville otoksille

Permutaatio- eli satunnaistamistestit "randomization tests" ovat niin sanottuja eksakteja testejä, joiden otosjakauma nollihypoteesin vallitessa ja siten myös laskettavat P-arvot saadaan tutkijan keräämästä havaintoaineistosta sopivasti permutoimalla, esimerkiksi toistuvasti jakamalla potilaat uudelleen eri hoitoryhmiin. Normaalijakaumaoletusta ei tarvita. Permutaatiotestejä voidaan käyttää jatkuville, järjestys- ja nominaaliasteikollisille muuttujille. Tavallisimmin niitä käytetään pienillä aineistoilla. Suuriin aineistoihin sovellettaessa käytetään Monte Carlo simulaatiota jolloin saatava P-arvo ei ole yksikäsitteinen vaan tietty väli.

Käyttötilanne:

- Kaksi otosta, esimerkiksi samoista henkilöistä kahtena eri ajankohtana
- Aineistokoko on alle 10 (voidaan laskea suuremmillekin aineistoille)
- Mittaukset välimatka-asteikolla (normaalisuusoletusta ei tarvita)

Esim.

FEV₁ (l/sek) ennen ja jälkeen bronkodilataattorin käytön astmaatikoilla (Campbell MJ & Machin D, 1999, s. 155) Aineisto: viisi koehenkilöä.

patno	FEV_ennen	FEV_jälkeen	Erotus
1	1,5	1,7	,2
2	1,7	1,9	,2
3	2,1	2,2	,1
4	1,6	1,9	,3
5	2,4	2,4	,0

Tuloksia eri testeillä **SPSS** ohjelmalla:

	Paired Samples Test	Pair 1
--	---------------------	--------

		FEV_ennen - FEV_jälkeen
Paired Differences	Mean	-1,600
	Std. Deviation	,1140
t		-3,138
df		4
Sig. (2-tailed)		,035

Wilcoxon Signed Ranks Test	FEV_jälkeen - FEV_ennen
Z	-1,841(a)
Asymp. Sig. (2-tailed)	,066
Exact Sig. (2-tailed)	,125
Exact Sig. (1-tailed)	,063
Point Probability	,063

a Based on negative ranks.

Test Statistics(b,c)			FEV_jälkeen - FEV_ennen
Z			-1,841(a)
Monte Carlo Sig. (2-tailed)	Sig.		,123
	99% Confidence Interval	Lower Bound	,104
		Upper Bound	,141
	Monte Carlo Sig. (1-tailed)	Sig.	
99% Confidence Interval		Lower Bound	,046
		Upper Bound	,074

(a) Based on negative ranks. (b) Wilcoxon Signed Ranks Test
(c) Based on 2000 sampled tables with starting seed 2000000.

Tulkinta: Näin pienellä aineistokoolla asymptoottiset testit (esim. t-testi) antavat yleensä ylioptimistisia tuloksia, liian pieniä P-arvoja. Aineistokokoa pitäisi tässä tapauksessa kasvattaa noin **10** koehenkilöön, mikäli standardoitu muutos on todellisuudessa kuten yllä **1,08** (muutoksen keskiarvo / muutoksen SD), voimaksi vaaditaan **90 %**, ja kaksipuoliseksi merkitsevyydestasoksi **0,05**.

Parittainen t-testi

Parittainen t-testi "Paired t-test/Mached pair t-test" on erityismuoto t-testistä ja sitä käytetään kahden parittaisen ryhmän, esim. kaltaistettujen pariin tai samojen henkilöiden toistomittausparien muodostaman erotusmuuttujan jakauman sijainnin testaamiseen.

Käyttötilanne:

Halutaan verrata keskiarvoja kahdessa toisistaan riippuvassa ryhmässä

Oletetaan, että (x_{1i}, x_{2i}) , $i = 1, \dots, n$ ovat parittaisia havaintoarvoja muuttujasta x joko a) parikaltaistetuissa ryhmissä 1 ja 2 tai b) saman ryhmän toistettuja mittausravvoja,

esimerkiksi kahtena ajankohtana. Lasketaan erotukset $d_i = x_{2i} - x_{1i}$, $i = 1, \dots, n$ ja erotusmuuttujan d keskiarvo \bar{d} , keskihajonta SD_d ja keskivirhe $SE_d = SD / \sqrt{n}$.

Laskentakaava: \bar{d} / SE_d

Tämä suure noudattaa Studentin t-jakaumaa vapausastein $df = n - 1$. Studentin t-jakauma on muodoltaan hivenen litteämpi kuin normaalijakauma, ts. sillä on pidemmät hännät sekä vasemmalle että oikealle. Jakauman litteyden määrää vapausaste df . Kun df kasvaa t-jakauma lähestyy normaalijakaumaa. Kun n on suuruusluokkaa 30, jakaumat ovat lähes yhtenevät.

Olettamukset parittaisen t-testin käytölle:

1. Differenssimuuttujan d jakauma tulee olla likimain normaalinen
2. Havaintoarvojen ryhmän/toiston sisällä tulee olla toisistaan riippumattomia.

Huom.

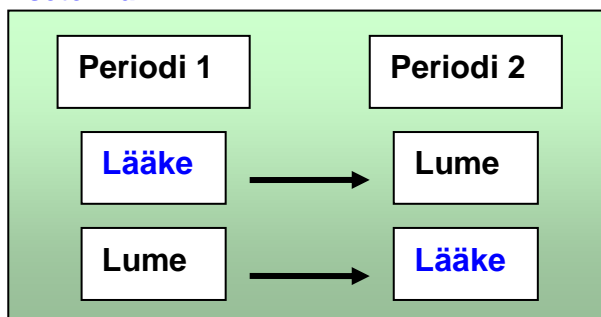
Olettamus 1 ei tarkoita, että muuttujan x jakauman tarvitsisi olla normaalinen ryhmissä 1 ja 2. Riittää, kun d noudattaa likimain normaalijakaumaa. t-testi ei ole kovin herkkä poikkeamalle normaalisuudesta, eli testi on tässä suhteessa ns. "robustinen". Kovin pienillä aineistoilla (alle 10) poikkeavilla havaintoarvoilla voi olla suuri vaikutus t-testin tulokseen ja t-testistä tulee konservatiivinen (varovainen), koska testisuureen nimittäjässä oleva keskivirhe kasvaa. Lisäksi normaalisuudesta on vaikea varmistua ja silloin kannattaa käyttää t-testin parametritonta vaihtoehtoa, Wilcoxonin parittaista testiä.

Olettamus kaksi on testin kannalta olennainen. Testi ei sovellu toisistaan riippuville havaintoarvoille. Riippuvuus voisi syntyä esim. seuraavasti: 30 henkilöltä on mitattu verenpaine kolme kertaa ja t-testi lasketaan ikään kuin havaintoja olisi 90. Toistomittausten huomioon ottaminen edellyttää erityismenetelmiä.

Esim.

Kliininen koe rauhoittavan lääkkeen käytöstä neuroottisilla potilailla (Armitage & Berry 1987 s. 105, 411). Kyseessä oli kymmenellä potilaalla suoritettu lumehoitokontrolloitu ristikkäistutkimus, ts. puolet potilaista satunnaistettiin aloittamaan lääkehoidolla ja toinen puoli lumehoidolla. Puhdistumisjakson jälkeen roolit vaihtuivat. Hoitoaika kummallakin periodilla oli yksi viikko. Lopputulosmuuttujana tarkasteltiin ahdistuneisuuspistemäärää, jonka skaala oli (0-30). Onko lääke- ja lumehoidolla saatujen tulosten välillä tilastollisesti merkitsevää eroa?

Asetelma:



Tulokset:

Potnro	Lääke	Lume	Erotus (Lääke- Lume) (d _i)
1	19	22	-3
2	11	18	-7
3	14	17	-3
4	17	19	-2
5	23	22	1
6	11	12	-1
7	15	14	1
8	19	11	8
9	11	19	-8
10	8	7	1

Testataan aluksi differenssimuuttujan d_i normaalisuusoletuksen paikkansapitävyys.

SPSS: valikot: **“Analyze”** ► **“Descriptive Statistics”** ► **“Explore”**.

Kohtaan **“Dependent list”** vietään tutkittava muuttuja **“Erotus”**. Kohdasta **“Plots”** klikataan lokerot **“Normality plots with tests”** ja **“Histogram”**

Tulos:**Tests of Normality**

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Erotus	,206	10	,200*	,936	10	,512

a. Lilliefors Significance Correction *. This is a lower bound of the true significance.

Tulkinta: Molemmilla testeillä $P > 0,05$, joten normaalisuusoletus jää voimaan.

Huom.

Näin pienellä aineistokoolla differenssimuuttujan normaalisuusoletuksen varmistaminen voi olla ongelmallista. Yksikin poikkeava havaintoarvo saattaa kasvattaa keskivirhettä ja tehdä t-testistä konservatiivisen, eli antaa liian suuren P-arvon.

SPSS: Parittainen t-testi löytyy valikoista: **“Analyze”** ► **“Compare means”**

► **“Paired-Samples T Test”** viemällä lokeroon **“Paired variables”** muuttujat **“Lääke”** ja **“Lume”**.

Tulos:**Paired Samples Statistics**

	Mean	N	Std. Deviation	Std. Error Mean

Pair 1	Lääke	14,80	10	4,686	1,482	
	Lume	16,10	10	4,954	1,567	
Paired Samples Correlations						
		N	Correlation	Sig.		
Pair 1	Lääke & Lume	10	,556	,095		
Paired Samples Test						
					Pair 1	
					Lääke - Lume	
Paired Differences	Mean				-1,300	
	Std. Deviation				4,547	
	Std. Error Mean				1,438	
	95% Confidence Interval of the Difference	Lower				-4,553
		Upper				1,953
t				-,904		
df				9		
Sig. (2-tailed)				,390		
<p>Tulkinta: Koska Sig. (2-tailed) eli P-arvo $>0,05$ niin yhdistyneisyyspistemäärässä lääke- ja lumehoitoperiodien välillä ei ole tilastollisesti merkitsevää eroa.</p>						

Wilcoxonin parittainen testi

Wilcoxonin parittainen testi "Wilcoxon's signed rank/matched pair test" on parametriton, jakaumasta riippumaton, testi. Käytetään erityisesti silloin, kun testattava muuttuja on järjestysasteikollinen tai t-testin asemesta, kun jatkuvan muuttujan jakauma ei ole normaalin.

Käyttötilanne: Samat henkilöt kahtena eri ajankohtana tai kaltaistetut parit

Testi perustuu "+" ja "-" merkkisten järjestyslukujen summiin T_+ ja T_- . Jos aineiston lukupareissa on erotuksia, joiden tulos on nolla, laskevat monet tilasto-ohjelmat Wilcoxonin testistä Prattin (1959) esittämän modifikaation. Testin lisäksi kannattaa yleensä laskea mediaaniero (esim. hoitoero) ja sen 95 % luottamusväli.

Järjestyslukujen muodostaminen

Havaintoaineiston lukupareista $(x_1, y_1), \dots, (x_n, y_n)$ lasketaan erotukset $d_i = x_i - y_i$, jotka lajitellaan suuruusjärjestykseen $|d_i|$:n mukaan.

Silloin kun aineistossa esiintyy erotuksia, joiden arvo $d_i = 0$, on järjestyslukujen muodostamiseksi kaksi tapaa:

- Tavanomaisessa Wilcoxonin testissä tapaukset $d_i = 0$ jätetään pois järjestyslukuja muodostettaessa. Esimerkiksi SPSS:ssä menetellään näin.
- Toinen menettelytapa on antaa niille kaikille järjestysluku 0 ja seuraavat järjestysluvut alkavat arvosta (nollien määrä) + 1 (Prattin modifikaatio, ks. Rahe, 1974)

Mikäli aineistossa esiintyy samoja erotuksia ("ties"), niin järjestysluku lasketaan seuraavasti: Jos $|d_{(j)}| = |d_{(j+1)}| = \dots = |d_{(j+k)}|$, niin järjestysluku on $\frac{(j) + (j+k)}{2}$.

Jos merkitään nollassa poikkeavien erotusten määrää n' :lla, niin tällöin $T_+ + T_- = \frac{n'(n'+1)}{2}$.

Testissä voidaan käyttää joko summaa T_+ tai T_- . Pienillä n' -arvoilla ($n' \leq 25$) T_+ (tai T_-) noudattavat likimain normaalijakaumaa siten, että:

P-arvo katsotaan taulukoista (esim. Altman 1991, s.531). Suurilla n' -arvoilla ($n' > 25$) T_+ (tai T_-) noudattavat likimain normaalijakaumaa siten, että:

$$E(T_+) = \frac{n'(n'+1)}{4}, \quad \text{Var}(T_+) = \frac{n'(n'+1)(2n'+1)}{24} - \frac{\sum_{j=1}^g (t_j^3 - t_j)}{2}$$

Summatermi on samoista erotuksista ("ties") johtuva korjaustekijä, missä g on samojen erotusten muodostamien ryhmien määrä ja t_j on erotusten määrä tällaisessa ryhmässä j . **Asymptottinen testisuure** on:

$$z = \frac{|T_+ - E(T_+)| - \frac{1}{2}}{\sqrt{\text{Var}(T_+)}} , \text{ osoittajassa oleva } \frac{1}{2} \text{ on jatkuvuuskorjaus}$$

Esim.

AIDS:in hoitoon käytetyn AZT hoidon vaikutus seerumin antigeenitasoon. (Makuch RD & Parks WP 1988)

Potnro	Pre_AZT	Post_AZT	Erotus
01	149	0	-149
02	0	51	51
03	0	0	0
04	259	385	126
05	106	0	-106
06	255	235	-20
07	0	0	0
08	52	0	-52
09	340	48	-292
10	0	0	0
11	180	77	-103

12	0	0	0
13	84	0	-84
14	89	0	-89
15	212	53	-159
16	554	150	-404
17	500	0	-500
18	424	165	-259
19	112	98	-14
20	2600	0	-2600

Ratkaisu perinteisellä Wilcoxonin testillä: 20 potilasta, 4 nollaa, joten $n' = 16$

Järjestysluku	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$ d_i $	14	20	51	52	84	89	103	106	126	149	159	292	295	404	500	2600

$T_+ = 12$, $n' \cdot (n' + 1) / 2 = 136$, ($T_- = 136 - 12 = 124$), $E(T_+) = n' \cdot (n' + 1) / 4 = 68$,
 $Var(T_+) = n' \cdot (n' + 1) \cdot (2n' + 1) / 24 = 374$ (Ei samoja erotuksia) ja $SE(T_+) = 19,34$
z-testi: $(|T_+ - E(T_+)| - \frac{1}{2}) / SE(T_+) = (|12 - 68| - \frac{1}{2}) / 19,34 = 2,87$ ja $P = 0,0041$.

SPSS: Testi saadaan valikoista "Analyze" ► "Nonparametric Tests" ► "2 Related Samples" viemällä lokeroon "Test Pairs" muuttujat "Pre_AZT" ja "Post_AZT".
 Kohdassa "Test Type" klikataan "Wilcoxon" ja kohdasta "Exact" valitaan "Exact"
Tulos:

Ranks				
		N	Mean Rank	Sum of Ranks
Post_AZT - Pre_AZT	Negative Ranks	14 ^a	8,86	124,00
	Positive Ranks	2 ^b	6,00	12,00
	Ties	4 ^c		
	Total	20		

a. Post_AZT < Pre_AZT

b. Post_AZT > Pre_AZT

c. Post_AZT = Pre_AZT

Test Statistics ^b	
	Post_AZT - Pre_AZT
Z	-2,896 ^a
Asymp. Sig. (2-tailed)	,004
Exact Sig. (2-tailed)	,002
Exact Sig. (1-tailed)	,001

Point Probability	,000
--------------------------	-------------

a. Based on positive ranks.
b. Wilcoxon Signed Ranks Test

Huom.
SPSS tulostaa P-arvot kolmella desimaalilla, mutta jos haluaa arvot tarkemmin, niin arvon saa näkyviin tuplaklikkaamalla tulosteessa kyseistä kohtaa tai kopiaimalla taulukon esim. Exceliin. Esim. tässä eksakti P neljällä desimaalilla on =0,0021

Huom. SPSS laskee z-testin ilman jatkuvuuskorjausta ½.

Esim.**Kliininen koe rauhoittavan lääkkeen käytöstä neuroottisilla potilailla**

(Armitage & Berry 1987 s. 105, 411). Kyseessä oli kymmenellä potilaalla suoritettu lumehoitokontrolloitu ristikkäistutkimus, ts. puolet potilaista satunnaistettiin aloittamaan lääkehoidolla ja toinen puoli lumehoidolla. Puhdistumisjakson jälkeen roolit vaihtuivat. Hoitoaika kummallakin periodilla oli yksi viikko. Lopputulosmuuttujana tarkasteltiin ahdistuneisuuspistemäärää, jonka skaala oli (0-30).

Tulokset:

Potnro	Lääke	Lume	Erotus (Lääke-Lume) (d_i)
1	19	22	-3
2	11	18	-7
3	14	17	-3
4	17	19	-2
5	23	22	1
6	11	12	-1
7	15	14	1
8	19	11	8
9	11	19	-8
10	8	7	1

$ d_i $	1	1	1	1	2	3	3	7	8	8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
järjestysluku	2,5				5	6,5		8	9,5	

Tulos edellä esitetyillä kaavoilla:

$T_+ = 2,5 + 2,5 + 2,5 + 9,5 = 17$, $n \cdot (n+1)/2 = 55$, $E(T_+) = 10 \cdot 11 / 4 = 27,5$,
 $Var(T_+) = (2 \cdot 10 \cdot 11 \cdot 21 - (4^3 - 4) - (2^3 - 2) - (2^3 - 2)) / 48 = 96,25 - 1,5 = 94,75$ (1,5 on samoista erotuksista aiheutuva korjaus), $SE(T_+) = \sqrt{94,75} = 9,734$, $z = (10,5 - 1/2) / 9,734 = 1,0277$ ja $P = 0,3041$

$$\text{(Parittainen t-testi: } t = \frac{\bar{d}}{SE(d)} = -1,3 / 1,438 = -0,904, P = 0,390, df = 9)$$

SPSS: Valikot **"Analyze" ► "Nonparametric Tests" ► "2 Related Samples"**. Lokeroon **"Test Pairs"** viedään muuttujat **"Lääke"** ja **"Lume"**. Kohdassa **"Test Type"** klikataan **"Wilcoxon"** ja kohdasta **"Exact"** valitaan **"Exact"**

Tulos:

Ranks				
		N	Mean Rank	Sum of Ranks
Lume - Lääke	Negative Ranks	4 ^a	4,25	17,00
	Positive Ranks	6 ^b	6,33	38,00
	Ties	0 ^c		
	Total	10		

a. Lume < Lääke, b. Lume > Lääke, c. Lume = Lääke

Test Statistics ^b	
	Lume - Lääke
Z	-1,079 ^a
Asymp. Sig. (2-tailed)	,281
Exact Sig. (2-tailed)	,3223
Exact Sig. (1-tailed)	,161
Point Probability	,023

Huom. SPSS laskee asymptoottisen testin ilman jatkuvuuskorjausta ja siksi satu P-arvo on liian optimistinen. StatXact 8:llä laskettu eksakti testi antoi saman tuloksen **P=0,3223** (2-suuntainen).

a. Based on negative ranks. b. Wilcoxon Signed Ranks Test

McNemarin testi

McNemarin testi "McNemar's test" on parittaisten aineistojen luokiteltujen muuttujien riippuvuustesti. Esim. Yksi potilasjoukko, jolle on suoritettu jokin interventio ja halutaan tutkia tapahtuuko intervention johdosta muutosta mittarissa, joka on joko laatueroasteikollinen tai luokiteltu muuttuja.

Käyttötilanne:

Lopputulos on kaksiarvoinen (**Kyllä**, **Ei**) ja kyseessä on jokin seuraavista parittaisista asetelmista:

- Samat henkilöt mitattu kahtena eri ajankohtana ja tarkastellaan muutosta lopputulosmuuttujassa
- Ristikkäistutkimusasetelma
- Kaltaistettu tapaus-verrokkiasetelma

Parittaisessa aineistossa lopputulosvaihtoehtoja on siten neljä: (**Kyllä**, **Kyllä**), (**Kyllä**, **Ei**), (**Ei**, **Kyllä**) ja (**Ei**, **Ei**). Olkoon **e**, **f**, **g** ja **h** näihin vaihtoehtoihin liittyvät havaintoaineiston perusteella lasketut lukumäärät. Ristikkäistutkimuksessa lopputulos taulukkomuotoon tiivistettynä on siten:

Ristikkäistutkimus	Vaste hoitoon A		Yhteensä
	Kyllä	Ei	
Vaste hoitoon B Kyllä	e	f	e+f
Ei	g	h	g+h
Yhteensä	e+g	f+h	n

ja kaltaistetussa tapaus-verrokkitutkimuksessa vastaavasti:

Kaltaistettu tapaus-verrokkitutkimus	Verrokki Altistus		Yhteensä
	Kyllä	Ei	
Tapaukset Altistus Kyllä	e	f	e+f
Ei	g	h	g+h
Yhteensä	e+g	f+h	n

Approksimattivinen testisuure:

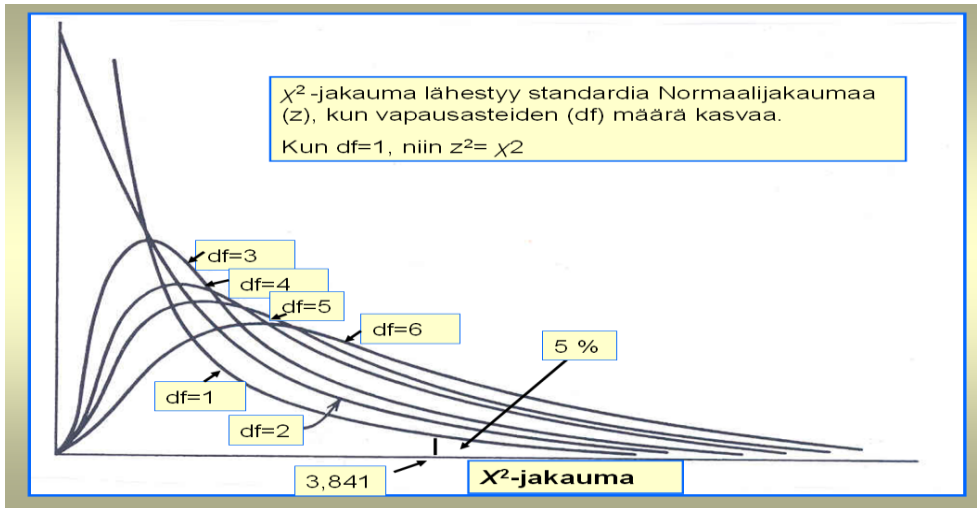
$$\chi^2 = \frac{(f - g)^2}{f + g}$$

Testisuure Yatesin jatkuvuuskorjauksella:

$$\chi_c^2 = \frac{(|f - g| - 1)^2}{f + g}$$

Testisuure noudattaa kummassakin tapauksessa likimain χ^2 -jakaumaa vapausastein 1.

χ^2 - testisuureen jakauma standardia Normaalijakaumaa, kun vapausastemäärä kasvaa.



Esim.

Kaltaistettu tapaus-verrokkitutkimus kivessyövistä (Brown LM, Pottern LM, Hoover RN 1987). Altisteena oli kivesten laskeutumattomuus syntyessä (kyllä -ei)

Kivessyöpä		Verrokki Altistus		Yhteensä
		Kyllä	Ei	
Tapaukset	Altistus Kyllä	4	11	15
	Ei	3	241	244
Yhteensä		7	252	259

Hypoteesit:

H₀: "Kivesten laskeutumattomuus synnytyksessä **ei altista** kivessyövälle"
 H₁: "Kivesten laskeutumattomuus synnytyksessä **altistaa** kivessyövälle"

Testit: $\chi^2 = \frac{(11-3)^2}{11+3} = 4,57, P = 0,033, \chi_c^2 = \frac{(|11-3|-1)^2}{11+3} = 3,50, P = 0,061$

Tarkka testaus, eli nk. **Eksakti McNemarin testi** voidaan suorittaa binomijakauman

perusteella kaavalla: $P = \frac{(f+g)!}{f!g!} \left(\frac{1}{2}\right)^{f+g}$

Lasketaan saadun lopputuloksen (yllä oleva taulukko) ja nollahypoteesin vallitessa sitä harvinaisempien lopputulosten todennäköisyydet seuraavista taulukoista:

- (i)

4	11
3	241

$$P = \frac{(14)!}{11!3!} \left(\frac{1}{2}\right)^{14} = 0,02222$$
- (ii)

4	12
2	241

$$P = \frac{(14)!}{12!2!} \left(\frac{1}{2}\right)^{14} = 0,00555$$
- (iii)

4	13
1	241

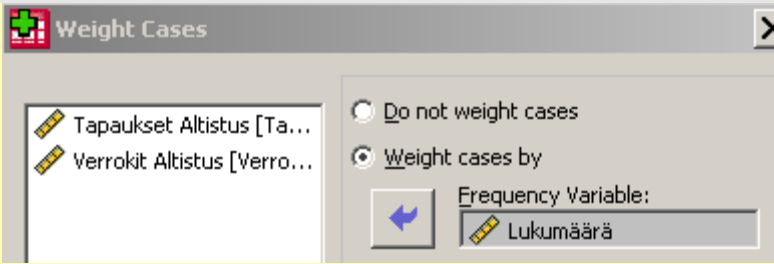
$$P = \frac{(14)!}{13!1!} \left(\frac{1}{2}\right)^{14} = 0,00085$$
- (iv)

4	14
0	241

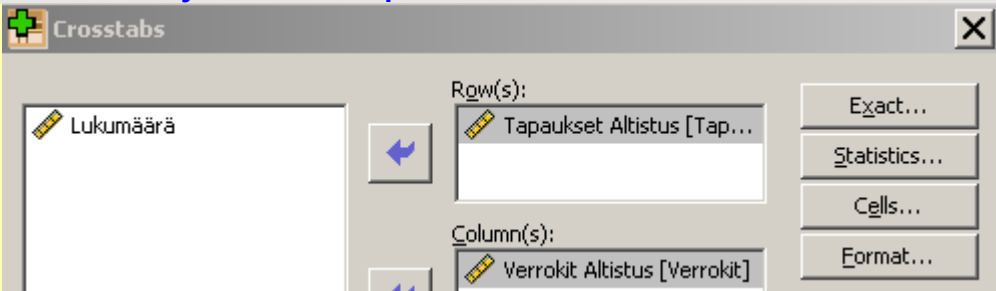
$$P = \frac{(14)!}{14!0!} \left(\frac{1}{2}\right)^{14} = 0,00006$$

Laskemalla nämä yhteen saadaan $P = 0,02868$. Kaksisuuntaisessa testissä tämä arvo kerrotaan kahdella ja siten ns. tarkaksi P-arvoksi saadaan $0,0574$. Vertaamalla approksimatiivisiin arvoihin todetaan, että tämä on melko lähellä jatkuvuuskorjattua arvoa.

SPSS: Kohdassa "Data" ► "Weight Cases" painomuuttujaksi määritellään Lukumäärä.



Valikot: **"Analyze" ► "Descriptive Statistics" ► "Crosstabs"**



Kohdasta **"Exact"** klikataan **"Exact"** ja kohdasta **"Statistics"** valitaan testiksi **"McNemar"**.

Chi-Square Tests

	Value	Exact Sig. (2-sided)
McNemar Test		,057^a
N of Valid Cases	259	

a. Binomial distribution used.

Tulkinta: H_0 jää voimaan; ts. kyseisen tutkimuksen perusteella ei saada riittävästi näyttöä altisteen yhteydestä syöpään sairastumiseen, koska ero tapausten ja verrokkien välillä altistumisessa ei ole tilastollisesti merkitsevä ($P=0,0574$).

McNemar–Bowkerin testi

Käyttötilanne:

- Parikaltaistetut aineistot. Mittaukset laatueroasteikolla
- Lopputulosmuuttujassa yli kaksi luokkaa (arvoa)

Huom. Testistä käytetään myös nimeä "Marginal Homogeneity test (StatXact 8).

Esim.

Kahden kuvantamismenetelmän, CT:n ja MRI:n vertailu alaselän välilevyn pullistuman ja juurikompression toteutamisessa. Aineistona 64 perättäistä avohoitopotilasta, jotka oli lähetetty Amsterdamin yliopistollisen sairaalan neurologian osastolle. Lähetediagnoosi oli LRS ("Lumbosacral Radicular Syndrome"). Potilaille epäiltiin välilevyn pullistumaa nikamavälissä L3-L4, L4-L5 tai L5-S1. Konservatiivinen hoito ei ollut tehonnut näihin potilaisiin. Heiltä tutkittiin kahden havainnoitsijan toimesta 171 välilevyä ja 456 hermojuurta sekä CT:llä että MRI:llä ja haluttiin testata antavatko menetelmät eri määrän eriäviä lopputuloksia.

Lähde: Rijn JC, Klemetso N, Reitsma JB et al. Observer variation in evaluation of lumbar herniated discs and root compression: spiral CT compared with MRI. The British Journal of Radiology 2006; 79: 372-377.

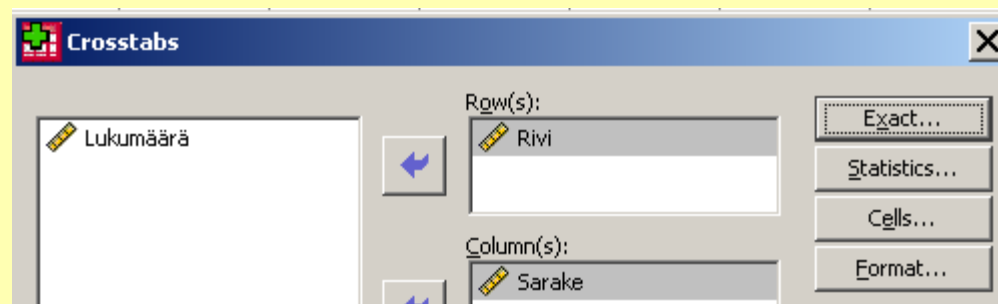
Välilevyn pullistuma HNP ("Herniated Nuclei Pulposus")				
CT	MRI			
	HNP	Eri mieltä	Ei HNP	
HNP	43	4	3	50
Eri mieltä	6	2	12	20
Ei HNP	3	6	92	101
	52	12	107	171

Selitykset:

"HNP" = "molemmat havainnoitsijat ovat samaa mieltä löydöksen HNP olemassaolosta".

"Ei HNP" = "molemmat havainnoitsijat ovat samaa mieltä löydöksen HNP puuttumisesta".

SPSS: Kohdassa "Data" ► "Weight Cases" painomuuttujaksi määritellään **Lukumäärä**. Valikot "Analyze" ► "Descriptive Statistics" ► "Crosstabs" Määritellään taulukon rivit ja sarakkeet ja kohdasta ja kohdasta "Statistics" valitaan testiksi "McNemar".

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
McNemar-Bowker Test	2,400	3	,494
N of Valid Cases	171		

Ohjelma ei laske eksaktia testiä muuta kuin 2x2-taulukolle.

Tulkinta: Menetelmien välillä ei ole tilastollisesti merkitsevää eroa, **P=0,494**.

Juurikompressio RT ("Root Compression")					
		MRI			
		Eri mieltä			
CT		RC	Ei RC		
RC		32	1	2	35
Eri mieltä		12	8	20	40
Ei RC		1	13	367	381
		45	22	389	

Selitykset:

"RC" = "molemmat havainnoitsijat ovat samaa mieltä löydöksen RC olemassaolosta".

"Ei RC" = "molemmat havainnoitsijat ovat samaa mieltä löydöksen RC puuttumisesta".

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
McNemar-Bowker Test	11,126	3	,011
N of Valid Cases	456		

Tulkinta: Menetelmien välillä on tilastollisesti merkitsevää ero, **P=0,011**. Menetelmä CT tuotti enemmän erimielisiä päätelmiä kuin MRI.

Esim.

Jonkin taudin ja ABO-veriryhmän välinen yhteys. Tapaus-verrokkiasetus. Luvut taulukossa ovat parien lukumääriä.

	Verrokkit				Total
	A	B	AB	O	
Tapaukset A	74	14	6	66	160
B	2	4	2	4	12
AB	10	12	2	12	36
O	18	8	3	64	93
Total	104	38	13	146	301

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)

McNemar-Bowker Test	51,305	6	,000
N of Valid Cases	301		

Tulkinta: Koska $P < 0,0001$, niin taudin ja ABO-veriryhmän välillä on tilastollinen yhteys. O- ja B veriryhmää on verrokeilla selvästi enemmän.

LR-testi siirtymätodennäköisyystaulukoille

Käyttötilanne:

Kyseessä on yleistetyn McNemarin testin yleistys, siten että halutaan verrata kahdessa eri ryhmässä tapahtuvia muutoksia keskenään. Tarkastelun kohteena oleva muuttuja x on, kuten McNemarin testin tapauksessakin, joko laatuero- tai järjestysasteikollinen saaden arvot x_1, \dots, x_r ja kummassakin ryhmässä on suoritettu parittaiset mittaukset esim. kahtena eri ajankohtana ja tuloksena on saatu seuraavanlaiset ristiintaulukoinnit. Taulukon luvut ovat lukumääriä siten, että esim. $n_{11}^{(k)}$ on niiden parien lukumäärä ryhmässä k ($k=1$ tai 2), jotka ovat kumpanakin ajankohtana saaneet muuttuja-arvon x_1 , jne...

		Ajankohta (t+1)				Σ
		x_1	. . .	x_r		
Ajankohta (t)	x_1	$n_{11}^{(k)}$	$n_{12}^{(k)}$	$n_{1r}^{(k)}$	$n_1^{(k)}$	
	x_2	$n_{21}^{(k)}$	$n_{22}^{(k)}$	
	
	x_r	$n_{r1}^{(k)}$. . .	$n_{rr}^{(k)}$	$n_r^{(k)}$	

Testissä on kyse siitä ovatko muutokset erilaiset vertailtavissa ryhmissä, ts. eroavatko päädiagonaalin alapuolella ja yläpuolella olevat luvut, esim. $n_{21}^{(1)}$ ja $n_{12}^{(1)}$ toisistaan enemmän/vähemmän kuin vastaavat luvut ryhmässä 2, eli $n_{21}^{(2)}$ ja $n_{12}^{(2)}$, jne. . . Testisuurena voidaan käyttää uskottavuussuhde, eli LR-testiä.

Uskottavuussuhde (LR), "likelihood ratio" määrittelee kuinka todennäköinen jokin tapahtuma A on tietyn ehdon B ollessa voimassa verrattuna siihen, että ehto B ei ole voimassa. Kaavana: $LR = P(A | B) / P(A | \bar{B})$
Esim. Millä todennäköisyydellä vatsahaavapotilaalla on helicobakteeri+ verrattuna muihin ylävatsaoireita poteviin potilaisiin.

Oletetaan, tarkasteltava muuttuja on r-luokkainen ja ryhmiä on g kpl.

Testisuure:

$$-2 \cdot \log \lambda = -2 \cdot \sum_{k=1}^g \sum_{i=1}^r \sum_{j=1}^r n_{ij}^{(k)} \left[\log_e \frac{\sum_{k=1}^g n_{ij}^{(k)}}{\sum_{k=1}^g n_i^{(k)}} - \log_e \frac{n_{ij}^{(k)}}{n_i^{(k)}} \right]$$

Tämä testisuure noudattaa likimain χ^2 -jakaumaa vapausastein $r \cdot (r-1) \cdot (g-1)$. Testiä ei valitettavasti löydy ainakaan StatXact 8:stä eikä SPSS 15:sta, mutta testi on kohtalaisen helppo ohjelmoida Exceliin. Vastaavanlaisen testin voinee suorittaa SAS:n Catmod-proseduurilla (Luokiteltujen muuttujien toistomittaus-Anova).

Esim.

x = kivun määrä: x₁ = "ei", x₂ = "lievä", x₃ = "kohtalainen", x₄ = "vaikea".
 Lääkkeet: A ja B. Ajankohtat: (t) = ennen hoitoa, (t+1) = hoidon jälkeen

Lääke A		Ajankohta (t+1)				
	X	X ₁	X ₂	X ₃	X ₄	Σ
Ajankohta (t)	X ₁	0	1	0	0	1
	X ₂	3	9	0	0	12
	X ₃	1	2	10	2	15
	X ₄	0	0	0	2	2

Lääke B		Ajankohta (t+1)				
	X	X ₁	X ₂	X ₃	X ₄	Σ
Ajankohta (t)	X ₁	0	0	0	0	0
	X ₂	5	4	1	0	10
	X ₃	8	5	4	0	17
	X ₄	3	0	0	0	3

LR-testi siirtymätodennäköisyydsmatriiseille:

Lääke A

X	X ₁	X ₂	X ₃	X ₄	S	X	X ₁	X ₂	X ₃	X ₄	S	X	X ₁	X ₂	X ₃	X ₄	S
X ₁	0	1	0	0	1	X ₁	0,00	1,00	0,00	0,00	1	X ₁	0,00	1,00	0,00	0,00	1
X ₂	3	9	0	0	12	X ₂	0,25	0,75	0,00	0,00	1	X ₂	0,36	0,59	0,05	0,00	1
X ₃	1	2	10	2	15	X ₃	0,07	0,13	0,67	0,13	1	X ₃	0,28	0,22	0,44	0,06	1
X ₄	0	0	0	2	2	X ₄	0,00	0,00	0,00	1,00	1	X ₄	0,60	0,00	0,00	0,40	1
S	4	12	10	4		S	0,32	1,88	0,67	1,13		S	1,24	1,81	0,48	0,46	

Lääke B

X	X ₁	X ₂	X ₃	X ₄	S	X	X ₁	X ₂	X ₃	X ₄	S	X	X ₁	X ₂	X ₃	X ₄	S
X ₁	0	0	0	0	0	X ₁	0,00	0,00	0,00	0,00	0	X ₁	0,00	1,00	0,00	0,00	1
X ₂	5	4	1	0	10	X ₂	0,50	0,40	0,10	0,00	1	X ₂	0,36	0,59	0,05	0,00	1
X ₃	8	5	4	0	17	X ₃	0,47	0,29	0,24	0,00	1	X ₃	0,28	0,22	0,44	0,06	1
X ₄	3	0	0	0	3	X ₄	1,00	0,00	0,00	0,00	1	X ₄	0,60	0,00	0,00	0,40	1
S	16	9	5	0		S	1,97	0,69	0,34	0,00		S	1,24	1,81	0,48	0,46	

Summattavat termit:

0	0	0	0
1,1241	-2,1457	0	0
1,4395	0,9902	-4,2121	-1,5154
0	0	0	-1,8326

0	0	0	0
-1,592	1,5608	-0,7885	0
-4,118	-1,4803	2,48096	0
-1,532	0	0	0

Testitulos: $-2 \cdot \log \lambda = 23,243$, $df = 4 \cdot 3 \cdot 1 = 12$, **P = 0,0257**. Tämä arvo saadaan esim. Excelillä seuraavasti: CHIDIST(23,243,12)

Excel-ohjelmia: <http://www.mv.helsinki.fi/home/sarna/stats/> on LR-testien laskemisohjelmat 2x2, 3x3, 4x4 ja 5x5-taulukoille.

Riippumattomien ryhmien t-testi

Riippumattomien ryhmien t-testi "independent sample t-test" on erityismuoto t-testistä. Sitä käytetään kahden toisistaan riippumattoman ryhmän, keskiarvojen vertailuun. Testattavan suureen oletetaan noudattavan normaalijakaumaa. Testistä on olemassa kaksi perusmuotoa, riippuen siitä ovatko vertailtavien ryhmien varianssit homogeeniset vai heterogeeniset. Jälkimmäiseen tapaukseen tilastopaketeista löytyy monia testin modifikaatioita. Ne perustuvat erilaisiin vapausasteiden muuntamistapoihin, joilla testisuure saadaan noudattamaan normaalijakaumaa.

Käyttötilanne:

Halutaan verrata keskiarvoja kahden toisistaan riippumattoman ryhmän, esim. hoito- ja kontrolliryhmän välillä. Riippumattomuus tarkoittaa, että vertailtavissa ryhmissä ei saa olla esim. samoja koehenkilöitä.

Oletetaan, että (x_1, \dots, x_{n_1}) ja (x_1, \dots, x_{n_2}) ovat havaintoarvoja jatkuvasta muuttujasta x kahdessa toisistaan riippumattomassa ryhmässä 1 ja 2. Lasketaan ryhmien keskiarvot \bar{x}_1 ja \bar{x}_2 ja keskihajonnat SD_1 ja SD_2 . Studentin t-testi lasketaan kaavoilla:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)},$$

$$SE(\bar{x}_1 - \bar{x}_2) = SD_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$SD_p = \sqrt{\frac{(n_1 - 1) \cdot SD_1^2 + (n_2 - 1) \cdot SD_2^2}{n_1 + n_2 - 2}}$$

SD_p on ryhmien 1 ja 2 yhdistetty hajonta.

Olettamukset t-testin käytölle:

1. Muuttujan x jakauma tulee likimain normaalin.
2. Havaintoarvojen vertailtavien ryhmien sisällä tulee olla toisistaan riippumattomia.
3. Varianssit ovat homogeeniset.

Huom.

t-testi ei ole kovin herkkä poikkeamalle normaalisuudesta, eli testi on tässä suhteessa ns. "robustinen". Kovin pienillä aineistoilla (ryhmäkoot alle 10) poikkeavilla havaintoarvoilla voi olla suuri vaikutus t-testin tulokseen ja t-testistä tulee konservatiivinen, koska testisuureen nimittäjässä oleva keskivirhe kasvaa. Lisäksi normaalisuudesta on vaikea varmistua ja silloin kannattaa käyttää t-testin parametritonta vaihtoehtoa, Wilcoxonin–Mann–Whitneyn testiä.

Olettamus kaksi on testin kannalta olennainen. Testi ei sovellu toisistaan riippuville havaintoarvoille. Riippuvuus voisi syntyä esim. siten, että vertailtavissa ryhmissä on samoja henkilöitä tai vertailtavien ryhmien sisällä on toistomittauksia samoista henkilöistä. Toistomittauksen huomioon ottaminen edellyttää erityismenetelmiä.

Varianssien homogeenisuus voidaan testata Levenen testillä. Mikäli testin antama P-arvo on alle 0,05, homogeenisuusoletus hylätään. Tähän tilanteeseen on olemassa useita t-testin modifikaatioita, jotka tavallisimmat tilastopakettien laskevat. Usein tässä tilanteessa kannattaa kuitenkin käyttää Mann–Whitneyn U-testiä.

Esim.

Tiedosto: Trial.sav. **Ongelma:** Halutaan testata onko hoitoryhmien Q ja R välillä eroa muutosmuuttujan $Imag_d = Imag_S - Imag_2$ keskiarvoissa.

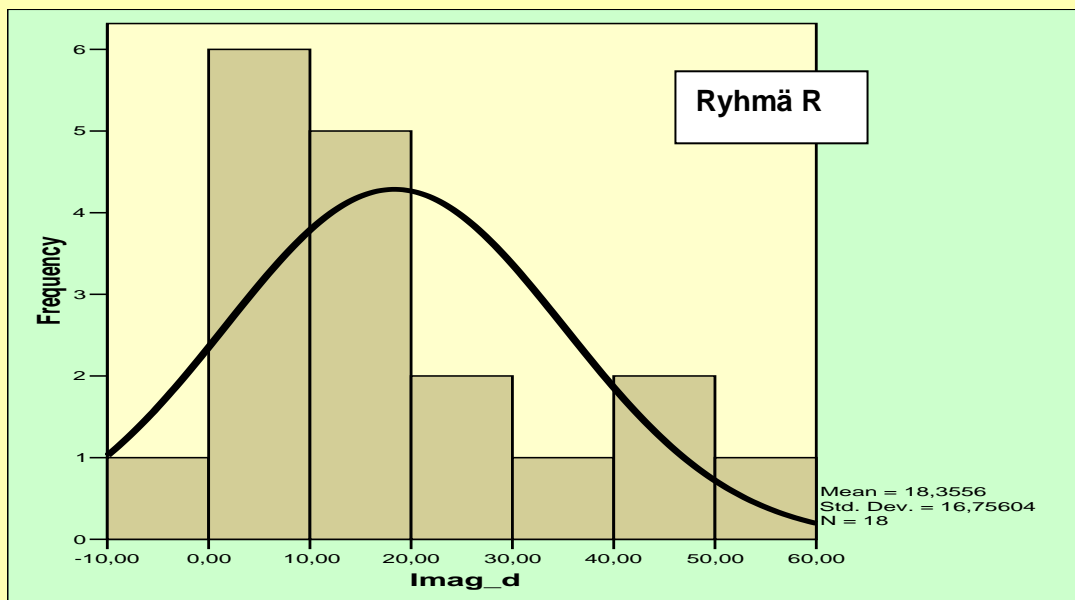
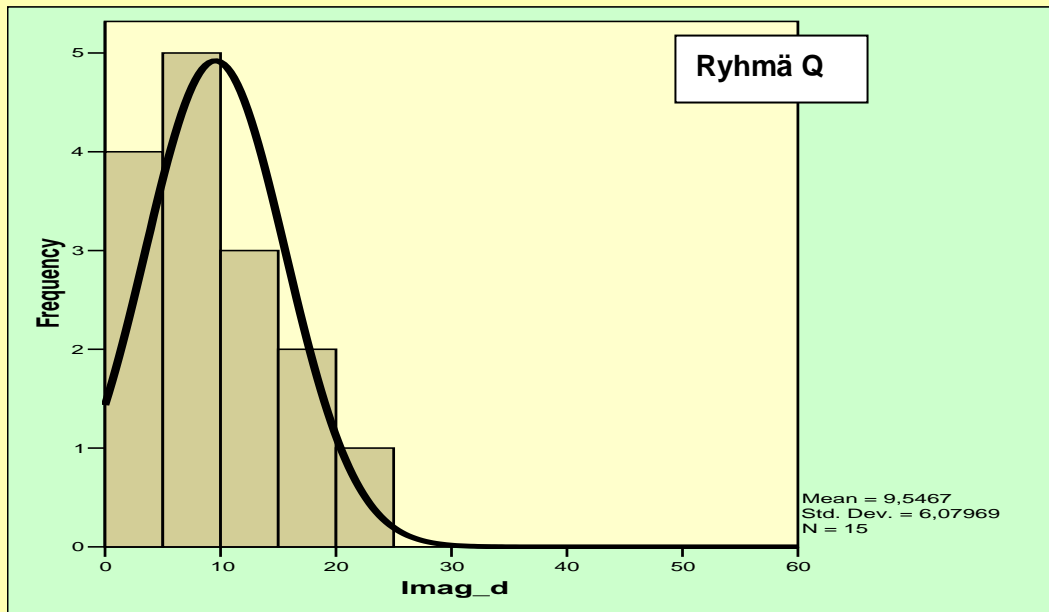
SPSS: Valikot: "Analyze" ► "Compare means" ► "Independent-Samples t-test" asettamalla "Test variable"= $Imag_d$ ja "Grouping variable" =Therapy. Vertailtaviksi rytmiksi määritellään Q (=2) ja R (=3).

Tulos:

Group Statistics					
	Therapy	N	Mean	Std. Deviation	Std. Error Mean
Imag_d	Group Q	15	9,5467	6,07969	1,56977
	Group R	18	18,3556	16,75604	3,94944

Independent Samples Test			
		Imag_d	
		Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	9,325	
	Sig.	,005	
t-test for Equality of Means	t	-1,929	-2,073
	df	31	22,125
	Sig. (2-tailed)	,063	,050

	Mean Difference	-8,80889	-8,80889
	Std. Error Difference	4,56711	4,24997
95% Confidence Interval of the Difference	Lower	-18,12358	-17,61990
	Upper	,50580	,00212



Tulkinta: Koska Levenen testiin liittyvä F-testi antaa tulokseksi $P < 0,05$, niin varianssien yhtä suuruus oletus hylätään ja tuloksia pitää tarkastella kohdasta **"Equal variances not assumed"**. Keskiarvoeroksi tulee -8,81 ja P-arvoksi **0,05**, joka on täsmälleen sama kuin mitä yleisesti pidetään tilastollisen merkitsevyyden rajana. **95 %:**n luottamusvälikin, **(-17,620, 0,002)**, pitää juuri ja juuri nollan sisällään. Tarkastelemalla ryhmien Q ja R jakaumia todetaan kuitenkin t-testin käytön kannalta

ongelma; jakaumat poikkeavat hajonnoiltaan ja muodoiltaan toisistaan ja erityisesti R:n jakauma on vino oikealle, joten t-testin käytölle asetetut oletukset 1 ja 3 eivät ole voimassa.

Tulos logaritmuunnoksen jälkeen:

Huom.

Ryhmissä Q ja R ei ole negatiivisia arvoja, joten logaritmuunnos on mahdollinen. Muussa tapauksessa arvoihin pitäisi lisätä vakio ennen logaritmuunnosta, esim. 10, jos ryhmä P olisi mukana vertailussa, koska siinä ryhmässä pienin arvo on -9.8),

Independent Samples Test

		LnImag_d	
		Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	2,468	
	Sig.	,126	
t-test for Equality of Means	t	-,804	-,845
	df	31	27,337
	Sig. (2-tailed)	,427	,405
	Mean Difference	-,29348	-,29348
	Std. Error Difference	,36485	,34717
	95% Confidence Interval of the Difference		
	Lower	-1,03761	-1,00541
	Upper	,45065	,41844

Tulkinta: Koska Levenen testiin liittyvä F-testi antaa tulokseksi $P > 0,05$, niin varianssien yhtä suuruus oletus on voimassa ja tuloksia pitää tarkastella kohdasta "**Equal variances assumed**". Saatua P-arvo $0,427$ ei enää ole läheskään tilastollisesti merkitsevä ja **95 %**:n luottamusvälikin pitää selvästi nollan sisällään.

Opetus: t-testi voi antaa harhaisen tuloksen, jos vertailtavien ryhmien jakaumat ovat erilaisia.

Mann–Whitneyn U-testi

Mann–Whitneyn U-testi “Mann–Whitney's U test” on parametriton (jakaumasta riippumaton) testi, joka on vaihtoehto riippumattomien ryhmien t-testille testaamaan onko jakaumien sijainti sama molemmissa ryhmissä. Käytetään erityisesti silloin, kun testattava muuttuja on järjestysasteikollinen tai t-testin asemesta, kun jatkuvan muuttujan jakauma ei ole normaalin. Testi on algebrallisesti sama kuin Wilcoxonin järjestyslukujen summatesti, eli antaa saman P-arvon. Tästä syystä testistä käytetään myös nimitystä Wilcoxon–Mann–Whitneyn testi.

Mann–Whitneyn U-testin voima on likimain 95,5 % verrattuna t-testiin silloin kun t-testin edellyttämä normalisuusoletus on voimassa.

Käyttötilanne:

Halutaan testata ovatko kaksi toisistaan riippumatta valittua satunnaisotosta x_{1i} , $i = 1, \dots, n_1$ ja x_{2i} , $i = 1, \dots, n_2$ peräisin perusjoukoista, joiden jakaumat ovat identtiset tarkastelun kohteena olevan muuttujan suhteen?

Huom.

- U-testi ei testaa pelkästään mediaanien eroa, vaan jakaumien sijaintia. Vertailtavilla ryhmillä voi olla samat mediaanit ja silti voidaan saada merkitsevä P-arvo U-testillä.
- Jos ryhmillä on sama jakauma, niin siirtymä jakauman sijainnissa siirtää keskiarvoja ja mediaaneja yhtä paljon ja siten mediaanien ero on sama kuin keskiarvojen ero. Näin ollen U-testi testaa myös keskiarvojen eroa.
- Erot hajonnoissa voivat olla yhtä merkityksellisiä U-testissä kuin erot mediaaneissa.
- Jos muuttujan arvot muunnetaan järjestyslukuiksi ja tehdään riippumattomien otosten t-testi U-testin asemesta, niin saadut P-arvot ovat likimain samoja.
- Logaritimuunnoksen käytöllä ei ole vaikutusta U-testin tulokseen, toisin kuin Wilcoxonin parittaisessa testissä.

Testillä on seuraavia **ominaisuuksia**:

- Ei oletusta normalisuudesta.
- Alkuperäisiä arvoja käytetään vain järjestyslukujen muodostamiseen, joiden avulla testisuureen arvo lasketaan.
- Testin voimakkuus on vain hieman heikompi kuin t-testin, kun sille asetetut oletukset ovat voimassa.

Merkitään: U_{12} on niiden parien lukumäärä, joissa $x_{1i} < x_{2i}$
 U_{21} on niiden parien lukumäärä, joissa $x_{1i} > x_{2i}$

Niistä pareista, joissa $x_{1i} = x_{2i}$ lasketaan puolet kumpaankin summaan, U_{12} ja U_{21} . Kumpaa tahansa summista U_{12} tai U_{21} voidaan käyttää testisuurena U , jolle pätee seuraavanlainen tulkinta: $U / (n_1 \cdot n_2)$ on niiden parien suhteellinen osuus, joille $n_{1i} < n_{2i}$.

Luvuille U_{12} ja U_{21} pätee seuraava: $0 \leq U_{12}, U_{21} \leq n_1 \cdot n_2$

$$E(U_{12}) = n_1 \cdot n_2 / 2$$

Esim.

Tiedosto: Trial.sav. Ongelma: Halutaan testata onko aktiivihoitoryhmien Q ja R välillä eroa muutosmuuttujan $Imag_d = Imag_S - Imag_2$ mediaaneissa.

SPSS: Valikot **"Analyze" ► "Nonparametric tests" ► "2 Independent-Samples"**. Testiksi valitaan **"Mann-Whitney U"** ja kohtaan **"Test variable"** viedään **"Imag_d"** ja kohtaan **"Grouping variable"** viedään **"Therapy"**. Kohdasta **"Exact"** klikataan **"Exact"**

Tulos:

Ranks				
	Therapy	N	Mean Rank	Sum of Ranks
Imag_d	Group Q	15	14,47	217,00
	Group R	18	19,11	344,00
	Total	33		

Test Statistics ^b	
	Imag_d
Mann-Whitney U	97,000
Wilcoxon W	217,000
Z	-1,374
Asymp. Sig. (2-tailed)	,169
Exact Sig. [2*(1-tailed Sig.)]	,178 ^a
Exact Sig. (2-tailed)	,178
Exact Sig. (1-tailed)	,089
Point Probability	,006

a. Not corrected for ties.

b. Grouping Variable: Therapy

Tests of Normality							
Therapy		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
	Group Q	,131	15	,200*	,961	15	,714
	Group R	,169	18	,185	,893	18	,043

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Tulkinta: Toisin kuin t-testi Mann-Whitneyn U-testi ei anna hoitoryhmien välille tilastollisesti merkitsevää eroa, $P > 0,05$. Kun vertailtavien ryhmien jakaumat

poikkeavat huomattavasti toisistaan tai, kun aineistot ovat niin pieniä, että on vaikea tarkistaa t-testin oletusten paikkansapitävyyttä, niin kannattaa valita Mann–Whitneyn U-testi t-testin asemesta.

Mediaaniero ja sen 95 % luottamusväliä SPSS ei valitettavasti laske eikä jostain syystä ilmoita myöskään Mann–Whitneyn-testin tulosten yhteydessä ryhmäkohtaisia mediaaneja. Sen sijaan ryhmien mediaanien eron SPSS laskee. Se ei kuitenkaan ole täsmälleen sama kuin mediaaniero, jonka laskemisproseduuri on kuvattu kohdassa mediaanieron luottamusväli. Käytännössä mediaanien ero on kuitenkin yleensä lähes sama kuin mediaaniero. Ryhmäkohtaiset mediaanit SPSS:llä saadaan seuraavasti:

SPSS: Valikot: "Analyze" ► "Compare Means" ► "Means". Viedään kohtaan "Dependent List" tässä tapauksessa muuttuja "Imag_d" ja kohtaan "Independent List" muuttuja "Therapy". Kohdasta "Options" klikataan "Grouped Median"

Tulos:

Report	
Grouped Median	
Therapy	Imag_d
Group Q	9,0000
Group R	13,5500
Total	7,1500

Huom.

Mikäli n_1 tai n_2 ovat alle 20, niin P-arvo kannattaa laskea eksakteilla algoritmeilla (esim. SPSS 17 tai StatXact 8) tai katsoa taulukoista. Kun aineistokoot ylittävät 20, voidaan käyttää asymptoottista testisuureta:

$$z = (U - m) / s, \text{ missä } m = n_1 \cdot n_2 / 2 \text{ ja } s = \sqrt{\frac{(n_1 + n_2 + 1) \cdot m}{6}}$$

Wilcoxonin järjestyslukujen summatesti

Testi on algebrallisesti sama kuin Mann–Whitneyn U-testi.

Merkitään: T_1 on arvojen x_{1i} järjestyslukujen summa yhdistetyssä järjestetyssä aineistossa ja T_2 on arvojen x_{2i} järjestyslukujen summa yhdistetyssä järjestetyssä aineistossa. Samoille lukuarvoille järjestyslukuksi tulee järjestyslukujen keskiarvo (kuten Wilcoxonin parittaisessa testissä).

T_1 :lle pätee seuraava epäyhtälö: $n_1 \cdot (n_1 + 1)/2 \leq T_1 \leq n_1 \cdot n_2 + n_1 \cdot (n_1 + 1)/2$

T_1 :n odotusarvo voidaan laskea kaavalla: $E(T_1) = n_1 \cdot (n_1 + n_2 + 1)/2$

Voidaan näyttää, että:

$$U_{12} + U_{21} = n_1 \cdot n_2 \text{ ja } T_1 + T_2 = (n_1 + n_2) \cdot (n_1 + n_2 + 1) / 2$$

Tästä seuraa, että U-testi ja Wilcoxonin testi ovat algebrallisesti yhtäpitäviä, ja tuottavat täysin saman testituloksen (P-arvon).

$$U_{21} = n_1 \cdot n_2 + n_2 \cdot (n_2 + 1) / 2 - T_2 \text{ ja } U_{12} = n_1 \cdot n_2 + n_1 \cdot (n_1 + 1) / 2 - T_1$$

Kun lukupareissa **ei ole** samoja lukuarvoja, niin $\text{Var}(U_{12}) = \text{Var}(T_1) = n_1 \cdot n_2 \cdot (n + 1) / 12$, missä $n = n_1 + n_2$. Kun lukupareissa **on** samoja lukuarvoja, niin

$$\text{Var}(U_{12}) = \text{Var}(T_1) = \frac{n_1 \cdot n_2}{12 \cdot n \cdot (n + 1)} \cdot \left[n^3 - n - \sum_t (t^3 - t) \right]$$

U- tai T-suureen keskivirheeksi saadaan siten $\text{SE}(U) = \sqrt{\text{Var}(U)} = \text{SE}(T)$

Huom.

ROC-käyrän ("Receiver Operating Characteristic") pinta-ala (A) ja sen keskivirhe voidaan laskea U-testisuureen avulla seuraavasti: $A = 100 \cdot \left(1 - \frac{U}{n_1 \cdot n_2}\right) \%$ ja

keskivirhe saadaan kaavasta: $\text{SE}(A) = \sqrt{\frac{\text{Var}(U)}{n_1 \cdot n_2}}$

ROC-käyrä on (x, y)-koordinaatistoon piirretty käyrä, missä x-akselina on väärän negatiivisen tuloksen todennäköisyys (1-spesifisyys) ja y-akselina on oikean positiivisen tuloksen todennäköisyys (sensitiivisyys), eli ROC-käyrän y- x-akselit ovat positiivisen uskottavuussuhteen **LR+** osoittaja ja nimittäjä..

Esim.

Hypoteettinen aineisto: Etsitään tietokonetomografiakuvista jotain tiettyä läydöstä. Kuvista 58 on normaaleita ja 51:ssä on jokin löydös. Arviointi tehdään järjestysasteikolla 1 – 5. Arvioitsijan (tai arviointiryhmän) tulos on seuraava:

	Arviointiskaala					Σ
	Ehdottoman normaali	→			Ehdottoman epänormaali	
Todellinen status	1	2	3	4	5	
Normaali	33	6	6	11	2	58
Epänormaali	3	2	2	11	33	51
Σ	36	8	8	22	35	109
Järjestysluku	18.5	40.5	48.5	63.5	92.0	

Tässä esimerkissä normaalit ja epänormaalit kuvat muodostavat toisistaan riippumattomat ryhmät ja koska asteikkona on järjestysasteikko, niin testiksi soveltuu Wilcoxonin järjestyslukujen summatesti. Sen avulla voidaan testata ovatko lopputuloksena olevat frekvenssijakaumat sijainniltaan erilaiset. Nollahypoteesi on: "jakaumat eivät poikkea toisistaan". Se tarkoittaa, että arvioitsija tunnistaa löydöksiä sattumanvaraisesti. Vaihtoehtoinen hypoteesi on: "Jakaumien suhteen on eroa".

Järjestysluku on muodostettu yhdistetyn aineiston perusteella. Esim. arvoja 1 on 36 kpl, joten järjestyslukuksi tulee arvojen 1, 2, ..., 36 keskiarvo, eli 18,5. Vastaavasti arvoja 2 on 8 kpl, joten niitä vastaavat järjestysluvut ovat 37, ..., 44 ja keskiarvo on 40,5. Siten esimerkiksi $T_1 = 33 \cdot 18,5 + 6 \cdot 40,5 + 6 \cdot 48,5 + 11 \cdot 63,5 + 2 \cdot 92,0 = 2027$.

StatXact 8-ohjelmalla saadaan eksaktilla testivaihtoehdolla seuraavat tulokset:

$T_1 = 2027$, $E(T_1) = 3190$, $SE(T_1) = 158,1$, $T_2 = 3968$, $z = -7,356$, $U = 316,0$

$A = 1 - 316 / (58 \cdot 51) = 0,893$ (89,3 %), $SE(A) = \sqrt{\frac{158,1^2}{58 \cdot 51}} = 0,0291$ (2,91 %)

Mikäli käytettävissä ei ole alkuperäistä aineistoa, vaan edellä olevan taulukon kaltainen tiivistetty muoto, niin SPSS-tilastopakettia varten tiedot tallennetaan muotoon:

<i>Todellinen status^a</i>	<i>Arvio</i>	<i>Lukumäärä</i>
1	1	33
1	2	6
1	3	6
1	4	11
1	5	2
2	1	3
2	2	2
2	3	2
2	4	11
2	5	33

(a) 1= Normaali, 2=Epänormaali.

missä muuttuja "Totuus" ilmaisee taulukon rivit, "Arvio" sarakkeet ja "Lukumäärä" lokerofrekvenssit. **SPSS**-ohjelmalla **ROC-käyrä** ja siihen liittyvät suureet ja testit voidaan tässä tapauksessa suorittaa seuraavasti:

Määritellään havaintojen painotus valikoista **"Data"** ► **"Weight cases"** asettamalla painomuuttujaksi **"Lukumäärä"**. Sen jälkeen otetaan valikot **"Analyze"** ► **"ROC curve"** ja tehdään valinnat:

ROC Curve

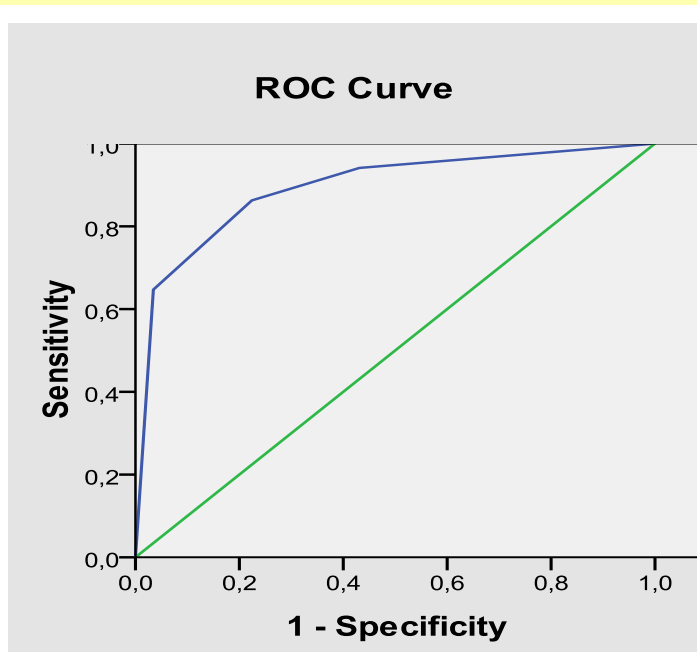
Test Variable: Arvio

State Variable: Todellinen status [Totuus]

Value of State Variable: 2

Display

- ROC Curve
- With Diagonal reference line
- Standard error and confidence interval
- Coordinate points of the ROC Curve



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s): Arvio

Area	Std. Error ^a	Asymptotic Sig. ^b	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,893	,032	,000	,830	,956

The test result variable(s): Arvio has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Tulkinta: Käyrän alle jäävä pinta-ala on 0,893. Nollahypoteesi (=arviointi sattumanvaraista) hylätään ($P < 0,001$), eli CT-kuvien arviointi toimii.

Coordinates of the Curve, Test Result Variable(s): Arvio

<i>Positive if Greater Than or Equal To(a)</i>	<i>Sensitivity</i>	<i>1 - Specificity</i>	<i>Sensitivity+ Specificity</i>
,00	1,000	1,000	1,000
1,50	,941	,431	1,510
2,50	,902	,328	1,574
3,50	,863	,224	1,639
4,50	,647	,034	1,613
6,00	,000	,000	1,000

The test result variable(s): Arvio has at least one tie between the positive actual state group and the negative actual state group. (a) The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Tulkinta: Koordinaattien perusteella voi määrittää optimaalisen katkaisukohtan. Mikäli se määritellään kohtana, jossa sensitiivisyyden ja spesifisyyden summa on maksimissaan, niin katkaisukohtan arvoksi tulee **3,50**.

χ^2 -nelikenttätesti

Käyttötilanne:

- 1) Vertailevissa kliinisissä tutkimuksissa** ("comparative study design").
Oletetaan, että vertaillaan kahta toisistaan riippumatonta ryhmää, jotka ovat otoksia kahdesta perusjoukosta 1 ja 2. Riippumattomuus tarkoittaa, että ryhmät muodostuvat eri havaintoyksiköistä; esim. eri henkilöistä. Oletetaan, että vastemuuttuja (lopputulosuuttuja) on kaksiarvoinen (kyllä, ei).
Tutkimuskysymys on: Ovatko "kyllä"-vastausten suhteelliset osuudet perusjoukoissa 1 ja 2 samat (H_0 , nollahypoteesi) vai onko niiden välillä eroa (H_A , vaihtoehtoinen hypoteesi)?
- 2) Poikittaistutkimusasetelmassa** ("cross-sectional study design").
Oletetaan, että kyseessä on yksi $n:n$ suuruinen otos tutkittavasta perusjoukosta ja kukin otokseen kuuluva henkilö (tai muu havaintoyksikkö) on luokiteltu kahden kaksiarvoisen muuttujan A ja B mukaisesti (kyllä, ei). Tutkimuskysymys on: Ovatko kaksiarvoiset muuttujat A ja B toisistaan riippumattomia (H_0 , nollahypoteesi) vai onko niiden välillä riippuvuutta (H_A , vaihtoehtoinen hypoteesi)?

Oletetaan, että tutkimustulos joko asetelman 1) tai 2) perusteella on taulukoitu muotoon:

		Vaste (lopputulokset)		Yhteensä
		Kyllä	Ei	
Ryhmä	1	a	b	a+b
	2	c	d	c+d
Yhteensä		a+c	b+d	n

		Muuttuja B		Yhteensä
		Kyllä	Ei	
Muuttuja A	Kyllä	a	b	a+b
	Ei	c	d	c+d
Yhteensä		a+c	b+d	n

Huom. Lukujen a , b , c ja d tulee olla **lukumääriä** eikä esimerkiksi prosenttilukuja.

Asetelmassa 1 ryhmien reunasummat, otoskoot, $a+b$ ja $c+d$ ovat kiinteitä lukuja, tutkijan valitsemia, mutta vastemuuttujan (kyllä, ei)-vastausten määrät, reunasummat $a+c$ ja $b+d$ määräytyvät satunnaisesti, ”kyllä”-vastausten vallitsevuuden perusjoukossa perusteella. Asetelmassa 2 sekä rivi että sarakereunasummat määräytyvät satunnaisesti, muuttujien A ja B ”kyllä”-vastausten vallitsevuuksien perusjoukossa perusteella. On olemassa myös asetelmia 1 ja 2 huomattavasti harvinaisempi asetelma, jossa saadaan yllä olevien taulukoiden muotoinen tulos. Siinä sekä rivi- että sarakereunasummat ovat kiinteitä, tutkijan määräämiä. Asetelmaa käytetään mm. kokeellisissa tutkimuksissa. Kaikkiin näihin asetelmiin suurissa otoksissa soveltuu sama alun perin K. Pearsonin (v.1900) ehdottama asymptoottinen testisuure:

Testisuure:

$$\chi^2 = \frac{(a \cdot d - b \cdot c)^2 \cdot n}{(a + c) \cdot (b + d) \cdot (a + b) \cdot (c + d)}$$

Testisuure noudattaa χ^2 -jakaumaa vapausastein 1.

Nelikentän lokerofrekvensseille voidaan laskea myös **odotusarvot** olettaen, että ryhmän ja lopputuloksen välillä ei olisi mitään riippuvuutta. Merkitään nelikentän lukujen a , b , c , d **odotusarvoja** $E(a)$, $E(b)$, $E(c)$ tai $E(d)$. Riippumattomuusoletuksen perusteella nämä lasketaan seuraavasti: **(rivisumma) · (sarakesumma) / (kokonaissumma)**

Esim. $E(a) = (a + b) \cdot (a + c) / n$. Tilastollisissa paketeissa on optio näiden odotusarvojen laskemiseksi.

Yleisesti odotusarvo voidaan määritellä seuraavasti:

Odotusarvo “expected value”

Esim. $(\text{aineistokoko}) \cdot (\text{jonkin tapahtuman todennäköisyys})$

χ^2 -testisuure toimii huonosti ”pienillä” aineistoilla, ts. kun $E(a)$, $E(b)$, $E(c)$ tai $E(d)$ ovat pieniä. R.A. Fisherin (v. 1922) ehdottama nyrkkisääntö käsitteelle ”pieni” tarkoittaa sitä, että nelikentän pienin odotettu frekvenssi on pienempi kuin 5. Tämä arvo 5 on ilmeisesti melko mielivaltaisesti valittu. E.S. Pearson ehdotti (v.1947) versiota, missä K. Pearsonin testisuureen kaavassa n korvataan $(n-1)$:llä, suure kerrotaan $(n-1)/n$:llä. Tätä versiota kutsutaan **$(n-1)\chi^2$ -testiksi**.

2x2 taulukoiden analysoimiseksi on esitetty lukuisia eri versioita. **Yatesin** (v.1934) ehdottama nk. ”**jatkuvuuskorjattu**” testiversio on seuraavanlainen:

$$\chi_c^2 = \frac{(|a \cdot d - b \cdot c| - \frac{n}{2})^2 \cdot n}{(a + c) \cdot (b + d) \cdot (a + b) \cdot (c + d)}$$

Tämä testiversio on konservatiivinen, eli korjaa useimmiten huomattavasti liikaa (Campbell I, 2007). Useimmat tilastopakettit tulostavat nelikenttien yhteydessä myös asymptoottisen **uskottavuussuhdetestin** (”likelihood ratio test”) joka lasketaan tässä tapauksessa seuraavasti:

$$LR = 2 \cdot (a \cdot \ln \frac{a}{E(a)} + b \cdot \ln \frac{b}{E(b)} + c \cdot \ln \frac{c}{E(c)} + d \cdot \ln \frac{d}{E(d)})$$

LR-testi on pienillä lokerofrekvensseillä jonkin verran luotettavampi kuin perusversio χ^2 -testistä. Tilastopakettien tulostamalla trenditestillä (esim. ”Linear-by-linear test”) ei tässä yhteydessä ole merkitystä, koska rivi- ja sarakemuuttujat ovat kaksiarvoisia. 2x2 taulukoiden analysoimiseksi on lisäksi ehdotettu lukuisia nk. ”eksakteja” testejä, esim **Fisher–Irwinin** testi, jota kutsutaan myös **Fisherin tarkaksi testiksi**.

Suositus 2x2 taulukoiden analysoimiseksi: (Campbell I, 2007):

- Kun ”pienin odotettu frekvenssi >1” käytä **(n-1) χ^2 -testiä**
- Kun ehto ei ole voimassa käytä **Fisherin tarkkaa** testiä siten, että kaksisuuntainen P-arvo lasketaan Irwinin (v. 1935) esittämällä tavalla ynnäämällä todennäköisyyksiä testisuureen molemmista hännistä (ks. seuraava luku).

χ^2 -testisuure voidaan muuntaa (n-1) χ^2 -testisuureeksi esim. Excel-ohjelmalla:

[http://www.mv.helsinki.fi/home/sarna/stats/\(n-1\)Khi2-testi.xls](http://www.mv.helsinki.fi/home/sarna/stats/(n-1)Khi2-testi.xls)

Esim.

Lasolin myrkyllisyyskoe rotilla (Oy Alko Ab). Rotat satunnaistettiin kahteen ryhmään. Ensimmäinen ryhmä sai 6-10 g/kg puhdasta etanolia 25 % liuoksena. Toinen ryhmä sai yhtä paljon etanolia sisältävää Lasolia. Lopputulosmuuttuja oli kuolema seuranta-aikana. Onko ryhmien välillä tilastollisesti merkitsevästi eroa?

		Lopputulos		Total	SPSS:		
		Kuoli	Ei		Ryhmä	Lopputulos	Lukumäärä
Ryhmä	Etanoli	14	6	20	1	1	14
	Lasol	7	13	20	1	2	6
Total		21	19	40	2	1	7
					2	2	13

SPSS: Oletetaan, että käytettävissä ei ole alkuperäistä aineistoa vaan ainoastaan nelikentän luvut. Määritellään havaintojen painotus valikoista ”**Data**” ► ”**Weight cases**” asettamalla painomuuttujaksi ”**Lukumäärä**”. Sen jälkeen otetaan valikot: ”**Analyze**” ► ”**Descriptive Statistics**” ► ”**Crosstabs**”. Määritellään taulukon rivit ja sarakkeet ja klikataan kohdasta ”**Statistics**” ”**Chi-square**”.

Tulos:

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	4,912 ^a	1	,027		
Continuity Correction ^b	3,609	1	,057		
Likelihood Ratio	5,019	1	,025		
Fisher's Exact Test				,056	,028
Linear-by-Linear Association	4,789	1	,029		
N of Valid Cases	40				

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 9,50.

b. Computed only for a 2x2 table

Tulkinta:
Tässä tapauksessa pienin odotettu frekvenssi on selvästi suurempi kuin 1, eli 9,5, joten voidaan käyttää Pearsonin asymptoottista 2-puolista testiä. Ryhmien välillä on tilastollisesti merkitsevä ero **P=0,027**. Mahdollisesti tutkijoiden yllätykseksi Lasol ryhmä menestyi paremmin!

(n-1)χ ² -testi			
χ ² -arvo	n	(n-1)χ ² -arvo	P-arvo
4,912	40	4,789	0,0286

Pienin odotettu frekvenssi on tässä tapauksessa paljon suurempi kuin 1, joten Pearsonin χ²-testin perusversio ja modifioitu versio antavat käytännössä saman tuloksen.

Ryhmän ja lopputuloksen välistä riippuvuutta nelikentässä voidaan mitata useilla suureilla. Eräs tavallisimmin käytetyistä on **ristitulosuhde ("odds ratio")**, joka lasketaan kaavalla:

$$OR = \frac{a \cdot d}{b \cdot c}$$

Ristitulosuhde "Odds ratio" (OR) on kahden kaksiarvoisen muuttujan välinen riippuvuuden mitta. OR=1 merkitsee, ettei ole riippuvuutta. Mittaa käytetään etenkin tapaus-verrokki-tutkimusasetelmassa taudin ja altisteen välistä yhteyttä arvioitaessa: OR = ("odds" altistuneille) / ("odds" altistumattomille). Käsitteestä käytetään myös nimityksiä kerroinsuhde, vedonlyöntikertoimien suhde tai vetosuhde.

SPSS: Kohdassa "**Crosstabs Statistics**" klikataan "**Chi-square**" ja "**Risk**"

Risk Estimate			
	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Ryhmä (Etanoli / Lasol)	4,333	1,150	16,323

Suhteelliset osuudet ja niiden luottamusvälit

Vertailtavien ryhmien ”kyllä”-vastausten suhteellisten osuuksien $p_1 = a/(a + b)$ ja $p_2 = c/(c + d)$ erotukselle voidaan laskea normaaliijakaumaan perustuva likimääräinen $100(1-\alpha)\%$:n luottamusväli seuraavasti:

$$(p_1 - p_2) \pm z_{1-\alpha/2} \cdot \sqrt{\frac{p_1 \cdot (1 - p_1)}{a + b} + \frac{p_2 \cdot (1 - p_2)}{c + d}}$$

Suhteellisten osuuksien erotusta ja sen luottamusväliä SPSS ei laske, mutta CIA:lla saadaan seuraava tulos:

Proportions and their differences			
Single Sample	Unpaired Samples	Paired Samples	
Unpaired Samples			
	Sample 1	Sample 2	Total
Number with feature	14	7	21
Number without feature	6	13	19
Sample Size	20	20	40
Which calculation method?			
<input checked="" type="radio"/> Recommended (Wilson) <input type="radio"/> Traditional			
Proportions	0,700	0,350	Difference 0,350
95% Confidence Interval for the difference			
Recommended (Wilson) Method	0,042	to 0,579	
Standard Error of difference		0,148	Normal Value 1,960

Prosenttilukujen ero on **35 %**, ja sen luottamusväli $CL_{95\%}$: **(4,2 %, 57,9 %)**. Todetaan, että nolla jää selvästi välin ulkopuolelle, joten tulkinta on sama kuin χ^2 -testin perusteella.

Fisherin tarkka nelikenttätesti

Käyttötilanne:

- Oletetaan, että vertaillaan kahta toisistaan riippumatonta ryhmää. Riippumattomuus tarkoittaa, että ryhmät muodostuvat havaintoyksiköistä; esim. eri henkilöistä.
- Kaksiarvoinen lopputulosmuuttuja (kyllä, ei).
- **Jokin nelikentän lukujen odotusarvoista ≤ 1**
- Testi perustuu hypergeometriseen jakaumaan, mikä merkitsee, että reunasummat oletetaan kiinteiksi (olettaus on epärealistinen joissakin tutkimusasetelmissa)

		Lopputulokset		Yhteensä
		Kyllä	Ei	
Ryhmä	1	a	b	a+b
	2	c	d	c+d
Yhteensä		a+c	b+d	n

Yllä olevaan nelikenttään liittyvä todennäköisyys saadaan **hypergeometriseen jakaumaan** perustuvalla kaavalla:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!}$$

P-arvo saadaan siten, että lasketaan todetun lopputuloksen todennäköisyys p_0 ja sitäkin harvinaisempien lopputulosten (1-tai 2-suuntainen) todennäköisyydet p_i edellyttäen, että H_0 olisi tosi ja summataan $P = \sum p_i$.

Esim. Lopputulokset

8	2	10	$p_0 = 0,02361$
1	5	6	
9	7		

Muut mahdolliset lopputulokset reunasummien ollessa kiinteät, eli 10, 6, 9 ja 7:

a) "Ylempi häntä", Ainoastaan yksi taulukko on mahdollinen, eli

9	1	10	$p_U = 0,00087$
0	6	6	
9	7		

b) "Alempi häntä", viisi mahdollista taulukkoa, eli:

7	3	10	$p_L = 0,15734$
2	4	6	
9	7		

c)

6	4	10	$p_L = 0,36713$
3	3	6	
9	7		

d)

5	5	10	$p_L = 0,33042$
4	2	6	
9	7		

e)

4	6	10	$p_L = 0,11014$
5	1	6	
9	7		

f)

3	7	10	$p_L = 0,01049$
6	0	6	
9	7		

Yksisuuntainen P-arvo lasketaan siten, että lopputuloksena saadun nelikentän todennäköisyyteen $p_0=0,02361$ lisätään H_0 :n ollessa voimassa harvinaisemman tuloksen, eli a):n todennäköisyys $p_U=0,00087$ ja näin saadaan testin tulos: **P=0,02448**.

Kaksisuuntaisessa testauksessa tähän P-arvoon pitää vielä lisätä toisesta "hännästä" lopputulokseen f) liittyvä todennäköisyys 0,01049, joka siis on H_0 :n ollessa voimassa harvinaisempi ($p_L < p_0$) kuin todettu lopputulos. (Tämä on alun perin Irwinin v.1935 esittämä laskentatapa ja siksi tätä Fisherin testin versiota kutsutaan myös **Fisher-Irwinin testiksi**). Näin saadaan kaksisuuntaiseksi P-arvoksi **0,03497**, joka on selvästi pienempi kuin 2*(yksisuuntainen P-arvo) eli 0,04896.

SPSS: Valikot: "Analyze" ► "Descriptive Statistics" ► "Crosstabs". Määritellään taulukon rivit ja sarakkeet ja klikataan kohdasta "Statistics" "Chi-square".

Tulos	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Fisher's Exact Test	,035	,024

3 cells (75,0%) have expected count less than 5. The minimum expected count is 2,63.

Huom.

- Eräät tilastopakettit antavat Fisherin testin P-arvon joko vain a) yksisuuntaisena (eivät tosin välttämättä ilmoita sitä!) tai b) laskevat sen kaksisuuntaisen P-arvon kertomalla yksisuuntaisen P-arvon kahdella. Tämä laskutapa antaa Fisher-Irwinin laskentasääntöä konservatiivisemmän tuloksen, elleivät jommat kummat reunasummista ole samat, eli $(a + b) = (c + d)$ tai $(a + c) = (b + d)$.
- Osa tilastopaketeista (esim. StatXact 8) laskee myös nk. "mid-P-arvon". Mid-P on vähemmän konservatiivinen, ts. voimakkaampi kuin Fisherin testin perusversio. Siinä laskutavassa summaukseen, jolla lopullinen P-arvo saadaan, todetun lopputuloksen P-arvosta otetaan mukaan vain puolikas. Yllä olevassa esimerkissä summaukseen otetaan $p_0/2$ eikä p_0 .

Fisherin tarkka testi (r x c)-taulukoille

Käyttötilanne:

- heterogeenisuuden testaaminen
- tarkasteltava muuttuja laatueroasteikollinen
- kaksi vertailtavaa ryhmää

Esim. Onko mahahaavan ja ABO-veriryhmän välillä tilastollisesti merkitsevä yhteys? (Varis K, Salmi H, ym., julkaisematon)

	ABO				Total
	O	A	B	AB	
Aktiivinen mahahaava	18	26	1	1	46
Muut ylävatsaoireiset	8	24	13	6	51
Total	26	50	14	7	97

SPSS:

Ryhmä	ABO	Lukumäärä
1	1	18
1	2	26
1	3	1
1	4	1
2	1	8
2	2	24
2	3	13
2	4	6

SPSS: Määritellään havaintojen painotus valikoista **"Data" ► "Weight cases"** asettamalla painomuuttujaksi **"Lukumäärä"**. Sen jälkeen otetaan valikot: **"Analyze" ► "Descriptive Statistics" ► "Crosstabs"** Määritellään taulukon rivit ja sarakkeet ja klikataan kohdasta **"Statistics" "Chi-square"** ja kohdasta **"Exact" "Exact"**.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	17,572^a	3	,0005	,0003		
Likelihood Ratio	19,935	3	,0002	,0003		
Fisher's Exact Test	18,101			,0003		
N of Valid Cases	97					

a. 2 cells (25,0%) have expected count less than 5. The minimum expected count is 3,32.

b. The standardized statistic is 3,854.

Tulkinta:

ABO-veriryhmäjakaumissa on eroa vertailtavien ryhmien välillä, ts ABO-veriryhmän ja ryhmän välillä on tilastollisesti merkitsevä yhteys.

Binomiaalisten suhteellisten osuuksien trenditesti**Käyttötilanne:**

- Aineisto muodostuu C:stä ryhmästä, lopputuloksena ovat binomiaaliset suhteelliset osuudet $p_1=k_1/n_1, \dots, p_c=k_c/n_c$ ja halutaan testata näiden välistä trendiä. Esim. annos-vaste tutkimuksissa C eri annosta, vaste kaksiarvoinen (+/-) ja halutaan testata onko annoksen ja vasteen välillä monotoninen yhteys.
- Aineisto muodostuu kahdesta ryhmästä, lopputulosmuuttuja on C-luokkainen järjestysasteikollinen muuttuja ja halutaan testata onko lopputuloksen suhteen monotoninen (nouseva/laskeva) trendi ryhmien välillä.

Huom.

Binomiaalisuus tarkoittaa sitä, että lopputuloksen kumpikaan vaihto-ehdoista ei saa olla liian harvinainen (alle 10 %), muuten pitäisi käyttää Poisson-jakaumaan perustuvaa trenditestiä.

Yllä oleviin testaustilanteisiin soveltuvia testejä on useita, esim.:

1. **χ^2 - trenditesti** (Armitage & Berry (1987) s. 372, Altman (1991) s.261)
Testissä lasketaan kaksi χ^2 -suuretta: χ^2_{het} ja χ^2_{trend} , joista ensimmäinen testaa suhteellisten osuuksien heterogeenisuutta ja jälkimmäinen trendiä. Näiden kahden suureen erotuksen perusteella voidaan testata poikkeamaa lineaarisesta trendistä.
2. **Cochran-Armitagen trenditesti** (Ohjelma: esim. StatXact 8)
3. **"Linear by linear association"-testi** (Ohjelma: esim. SPSS: "Crosstabs").
Tämä on sama testi kuin Mantel-Haenszelin χ^2 -testi ja antaa (2xC)-taulukolle saman tuloksen kuin Cochran-Armitagen binomiaalisten suhteellisten osuuksien trenditesti. "Linear by linear"-testi voidaan laskea myös kahteen suuntaan järjestetyille (RxR)-taulukolle.

Huom.

- Trenditesteissä luokille voidaan antaa muitakin lukuarvoja ("scores") kuin 1,2,...,C. Nämä lukuarvot pitää kuitenkin päättää ennen kuin testi suoritetaan.
- Useissa tilanteissa myös Mann-Whitneyn U-testi antaa lähes saman P-arvon kuin trenditestit (2xC)-taulukoille.

Esim. Lannerangan degeneratiiviset muutokset rautatieläisillä ("toppamiehet")
Ikätrendin testaaminen. (Romo M, ym.1980)

		<i>Muutos</i>		<i>Yhteensä</i>
		<i>+</i>	<i>-</i>	
<i>Ikäluokka</i>	<i><35</i>	5	32	37
	<i>35- 44</i>	17	34	51
	<i>45 ja yli</i>	24	22	46
	<i>Yhteensä</i>	46	88	134

SPSS:

Muutos	Ikäluokka	Lukumäärä
1	1	5
1	2	17
1	3	24
2	1	32
2	2	34
2	3	22

Eri trenditesteillä saadaan seuraavat tulokset:

1. $\chi^2_{\text{het}} = 13,631$, $df = 2$, **P = 0,0011**, $\chi^2_{\text{trend}} = 13,628$, $df = 1$, **P = 0,0002**,

$\chi_R^2 = \chi_{het}^2 - \chi_{trend}^2 = 0,003$, **P = 0,9985**. Todetaan, että trenditesti χ_{trend}^2 antaa pienemmän P-arvon kuin heterogeenisuudesta χ_{het}^2 (Pearsonin χ^2 -testi) ja että poikkeama lineaarisesta trendistä on erittäin pieni.

2. Cochran-Armitagen trenditesti binomiaalisille suhteellisille osuuksille (StatXact 8) antaa standardoiduksi poikkeamaksi **3,678** ja asymptoottiseksi kaksisuuntaiseksi P-arvoksi **0,0002** ja vastaavaksi eksaktiksi P-arvoksi **0,0003**.

3. "Linear by linear"-testi (SPSS) on (2x2)-taulukolle täsmälleen sama kuin Cochran-Armitagen trenditesti. Neliöjuuri "Linear by linear"-testisuureesta on sama kuin standardoitu poikkeama.

SPSS: Määritellään havaintojen painotus valikoista **"Data" ► "Weight cases"** asettamalla painomuuttujaksi **"Lukumäärä"**. Sen jälkeen otetaan valikot: **"Analyze" ► "Descriptive statistics" ► "Crosstabs"**. Asetetaan lokeroihin **"Row"** ja **"Column"** indikaattorimuuttujat "Ikäluokka" ja "Muutos" ja kohdasta **"Statistics"** valitaan testiksi **"Chi-square"** ja kohdasta **"Exact"** klikataan **"Exact"**.

Tulos:

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	13,631 ^a	2	,0011	,0010		
Linear-by-Linear Association	13,526 ^b	1	,0002	,0003	,0001	,00009
N of Valid Cases	134					

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 12,70.

b. The standardized statistic is -3,678.

Tulkinta:

Lannelsen degeneratiiviset muutokset lisääntyvät iän mukaan lähes lineaarisesti tässä aineistossa, koska $\chi_{het}^2 = 13,631$, $df = 2$, **P = 0,0011** ja $\chi_{trend}^2 = 13,526$, $df = 1$. **P = 0,0002** joten $\chi_R^2 = \chi_{het}^2 - \chi_{trend}^2 = 0,105$, **P = 0,9488**

Esim.

Alaselän kiputilan hoitokoe. Hypoteettinen esimerkki. Kaksi hoitoryhmää, joissa kummassakin 15 potilasta. Hoito A: "lepo" ja hoito B: "ohjattu liikunta". Lopputulos: subjektiivinen arvio 5-luokkaisella asteikolla.

		Hoito		Yhteensä
		A	B	
Tulos	Selvästi parempi	2	4	6
	Parempi	4	8	12
	Ennallaan	5	2	7
	Huonompi	3	1	4
	Selvästi huonompi	1	0	1

Yhteensä	15	15	30
----------	----	----	----

Eri testeillä saadaan seuraavat tulokset:

1. $\chi^2_{\text{trend}} = 4,337$, $df = 1$, **$P=0,03729$** . $\chi^2_{\text{het}} = 5,286$, $df = 4$, $P = 0,25919$, $\chi^2_R = 0,948$, **$P = 0,81383$** , joten poikkeama lineaarisesta trendistä ei ole merkitsevä
2. Cochran–Armitagen trenditesti ja ”Linear by linear”-testi antavat kaksisuuntaiseksi asymptoottiseksi P-arvoksi **0,0406** ja eksaktiksi P-arvoksi **0,0579**.
3. Mann–Whitney:n testi (StatXact 8) antaa asymptoottiseksi P-arvoksi **0,0414** ja lähes saman eksaktiksi, eli **$P = 0,0411$** .

Mantel–Haenszelin nelikenttätesti

Mantel–Haenszelin testi “Mantel–Haenszel’s test” on yksi yleisimmin käytetyistä testeistä, etenkin epidemiologisissa tutkimuksissa, kun halutaan kontrolloida sekoittavien tekijöiden vaikutusta. Testimenetelmällä voidaan laskea yhdistetty arvio suhteelliselle riskille useista (2 x 2) tai testin yleisemmässä muodossa (2 x k)-taulukoista, $k > 2$. Yhdistetty arvio on painotettu keskiarvo eri ositekohtaisten taulukoiden suhteellisen riskin arvioista. Testiä käytetään etenkin tapaus-verrokkitutkimuksissa ja meta-analyyseissä, jolloin yhdistetyn arvion laskemisessa käytetyt painokertoimet ovat yleensä suoraan verrannollisia yksittäisten ositekohtaisten arvioiden varmuuteen tai tarkkuuteen, eli kääntäen verrannollinen varianssiin. Suuremmat ositteet tai tutkimukset antavat siten suuremman painon yhdistettyyn arvioon kuin pienemmät.

Mantel–Haenszelin arvio “Mantel–Haenszel’s estimate” on Mantel–Haenszelin menetelmällä useista tutkimuksista tai ositteista laskettu yhdistetty arvio suhteelliselle riskille.

Käyttötilanne:

- nelikenttien yhdistäminen joko
 - a) samaa ongelmaa tarkastelevista eri tutkimuksista tai
 - b) saman tutkimuksen eri ositteista
- sekoittavien tekijöiden (“confounders”) hallinta

Oletetaan, että sekoittava tekijä **z** on laatuero-, järjestys- tai luokiteltu välimatka-asteikollinen ja saa arvot: 1, ..., k.

Tutkitaan **x**:n ja **y**:n välistä riippuvuussuhdetta, missä esim.

x = hoitoryhmää ilmaiseva muuttuja,

y = lopputulosta ilmaiseva muuttuja (kyllä, ei)

Halutaan kontrolloida **z**:n vaikutus analysoitaessa **x**:n ja **y**:n välistä riippuvuussuhdetta. Oletetaan, että aineisto on jaettu sekoittavan tekijän mukaisesti **k**:hon ositteeseen ja että kussakin ositteessa on laskettu seuraava nelikenttä:

		Lopputulokset (y)		Yhteensä
		Kyllä	Ei	
(z) Osite j	(x) Ryhmä A	a _j	b _j	M _{1j}
	B	c _j	d _j	M _{0j}
Yhteensä		N _{1j}	N _{0j}	T _j

Lasketaan seuraavat ositekohtaiset suuret:

$$OR_j = \frac{a_j \cdot d_j}{b_j \cdot c_j}, E(a_j) = \frac{M_{1j} \cdot N_{0j}}{T_j}, \text{ missä } i = 0,1 \text{ ja } j = 1, \dots, k.$$

Oletetaan M_{1j} , M_{0j} , N_{1j} ja N_{0j} kiinteiksi. Hypergeometrisesta jakaumasta seuraa, että

$$\text{Var}(a_j) = \frac{M_{1j} \cdot M_{0j} \cdot N_{1j} \cdot N_{0j}}{T_j^2 \cdot (T_j - 1)}$$

Ositekohtainen testisuure lasketaan kaavalla:

$$\chi_{1,j}^2 = \frac{(|a_j - E(a_j)| - \frac{1}{2})^2}{\text{Var}(a_j)}$$

Yli kaikkien ositteiden lasketuksi yhdistetyksi arvioksi ristitulosuhteelle, odotusarvolle ja varianssille saadaan:

$$OR_k = \frac{\sum_{j=1}^k a_j \cdot d_j / T_j}{\sum_{j=1}^k b_j \cdot c_j / T_j}, E(a_k) = \sum_{j=1}^k E(a_j), \text{Var}(a_k) = \sum_{j=1}^k \text{Var}(a_j)$$

Näiden arvioiden perusteella saadaan Mantel-Haenszelin nelikenttätestiksi:

$$\chi_{1,M-H}^2 = \frac{(\sum_{j=1}^k a_j - \sum_{j=1}^k E(a_j)) - \frac{1}{2}}{\sum_{j=1}^k \text{Var}(a_j)}$$

Huom.

- $\chi_{1,M-H} = z$ eli normaalijakauman prosenttipiste.
- Testin kumulatiivisuudesta johtuen odotetut frekvenssit yksittäisissä nelikentissä voivat olla pieniä (< 5).

Esim.

Kaksi hoitoa A ja B, sekoittavana tekijänä ikä, lopputulos parantuminen

a) Tulos ilman sekoittavan tekijän kontrollointia:

		Lopputulos		Yhteensä
		Kyllä	Ei	
Ryhmä	A	33	28	61
	B	20	40	60
Yhteensä		53	68	121

Tulos: Chi-Square Tests	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	5,298(b)	1	,021	,028
Continuity Correction(a)	4,488	1	,034	
Likelihood Ratio	5,342	1	,021	,028
Fisher's Exact Test				,028

(a) Computed only for a 2x2 table

(b) 0 cells (,0%) have expected count less than 5. The minimum expected count is 26,28.

Tulkinta: Hoidolla ja parantumisella on tilastollisesti merkitsevä yhteys; $P = 0,028$, kun ikää ei huomioida.

b) Sekoittavan tekijän kontrollointi osittamalla:

		Lopputulos		Yhteensä	OR
		Kyllä	Ei		
Nuoret	A	14	6	20	2,00
	B	7	6	13	
Keski-ikäiset	A	12	7	19	2,40
	B	5	7	12	
Vanhat	A	7	15	22	1,58
	B	8	27	35	
Yhteensä		53	68	121	

SPSS:

Osite	Ryhmä	Lopputulos	Lukumäärä
1	1	1	14
1	1	2	6
1	2	1	7
1	2	2	6
2	1	1	12
2	1	2	7
2	2	1	5
2	2	2	7
3	1	1	7
3	1	2	15
3	2	1	8
3	2	2	27

SPSS: Määritellään havaintojen painotus valikoista **"Data"** ► **"Weight cases"** asettamalla painomuuttujaksi **"Lukumäärä"**. Sen jälkeen otetaan valikot: **"Analyze"** ► **"Descriptive statistics"** ► **"Crosstabs"**. Asetetaan lokeroihin **"Row"**, **"Column"** ja **"Layer"** muuttujat **"Ryhmä"**, **"Lopputulos"** ja **"Osio"** Kohdasta **"Statistics"** valitaan testiksi **"Chi-square"** ja **"Cochran's and Mantel-Haenszel statistics"**. Kohdasta **"Exact"** klikataan **"Exact"**.

Chi-Square Tests

Osio		Value	df	Exact Sig. (2-sided)
Nuoret	Pearson Chi-Square	,888 ^a	1	,465
Keski-ikäiset	Pearson Chi-Square	1,372 ^d	1	,288
Vanhat	Pearson Chi-Square	,559 ^f	1	,542

a. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 4,73.

b. Computed only for a 2x2 table

d. 0 cells (,0%) have expected count less than 5. The minimum expected count is 5,42.

f. 0 cells (,0%) have expected count less than 5. The minimum expected count is 5,79.

Tulkinta: Mikään ositekohtaisista P-arvoista ei ole tilastollisesti merkitsevää

Tests of Homogeneity of the Odds Ratio

	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	,196	2	,907
Tarone's	,196	2	,907

Tulkinta: Ositekohtaiset OR:t eivät poikkea toisistaan tilastollisesti merkitsevästi. Voidaan laskea yhdistetty OR.

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	2,636	1	,104
Mantel-Haenszel	1,976	1	,160

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Tulkinta: Hoidot A ja B ovat epätasapainossa eri ikäryhmissä (reunajakaumat ovat erilaiset), ts. ikä assosioituu hoitoja ilmaisevaan muuttujaan. Lisäksi ositekohtaisista OR:stä todetaan, että ikä assosioituu myös hoidon lopputulokseen. Näin ollen ikä

täyttää sekoittavan tekijän vaatimukset. Iän kontrolloinnin jälkeen testitulos ei ole enää tilastollisesti merkitsevä, sillä $\chi^2_{1,M-H} = 1,976$, $df = 1$, **P = 0,160**
Hoidon A parempi lopputulos selittyy siten suurelta osin iällä.

Mantel-Haenszel Common Odds Ratio Estimate

Estimate		1,902	
ln(Estimate)		,643	
Std. Error of ln(Estimate)		,398	
Asymp. Sig. (2-sided)		,106	
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound Upper Bound	,872 4,148
	ln(Common Odds Ratio)	Lower Bound Upper Bound	-,137 1,423

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1,000 assumption. So is the natural log of the estimate.

Tulkinta: Yhdistetyksi ristitulosuhteeksi saadaan **OR_k = 1,902**. Yhdistäminen ei ole perusteltua elleivät ositekohtaiset ristitulosuhteet ole keskenään homogeenisia. Tässä tapauksessa homogeenisuusvaatimus täyttyi (**P=0,907** Breslow ja Dayn testi). Yhdistetyn riskisuhteen **95 % eksaktiksi luottamusväliksi** saadaan StatXact 8:llä (**0,8064, 4,367**).

Mikäli lopputulos on k-luokkainen, voidaan testaukseen käyttää edellä kuvattua nelikenttätestin yleistystä, Mantel-Haenszelin testiä (2 x k) -taulukkoille. Huolimatta tämän testin yleisyydestä ja käyttökelpoisuudesta se yllättäen puuttuu monista tavallisimmista tilastopaketeista, kuten SPSS.

Mantel-Haenszelin testiä voidaan käyttää monissa muissakin kuin edellä mainituissa tutkimusasetelmissä, esimerkiksi tapaus-verrokkitutkimuksissa.

Kruskal–Wallisin yksisuuntainen varianssianalyysi

Kruskal–Wallisin testi, “**Kruskal–Wallis test**” Parametriton merkitsevyydestä, jolla verrataan useamman kuin kahden ryhmän jakaumien sijaintia. Käytetään erityisesti järjestysasteikollisille muuttujille, mutta myös vaihtoehtona yksisuuntaiselle varianssianalyysille (Anova), silloin, kun tämän oletukset eivät täyty. Se on yleistys Mann–Whitneyn U-testistä. Kahden ryhmän tapauksessa molemmat testit antavat saman tuloksen.

Käyttötilanne:

- Onko k riippumatonta otosta peräisin samasta perusjoukosta (tai identtisistä perusjoukoista), joilla on sama mediaani.

- testi perustuu järjestyslukuihin
- testi edellyttää vähintään järjestysasteikkoa
- ryhmä (otos-) koot n_i voivat olla erisuuria

Testisuure:

$$H = \frac{12}{N \cdot (N + 1)} \cdot \sum_{i=1}^k n_i \bar{R}_i^2 - 3 \cdot (N + 1)$$

Kaavassa $N = \sum_{i=1}^k n_i$ ja $\bar{R}_i = \frac{\sum_{j=1}^{n_i} R_j}{n_i}$ on järjestyslukujen R_j keskiarvo ryhmässä i .

Mikäli kyseessä on pieni aineisto (ryhmäkoot 5 tai alle) P-arvo katsotaan testisuurelle H laadituista taulukoista. Muissa tapauksissa H :n merkitsevyyden arvioimiseen voidaan käyttää χ^2 -jakaumaa vapausastein $(k - 1)$.

Kun järjestetyssä aineistossa ryhmästä riippumatta esiintyy samoja lukuarvoja ("ties"), pitää testisuuretta H korjata ottamalla nämä tasatulokset huomioon. Testisuure H pitää

jakaa tekijällä: $1 - \frac{\sum_{i=1}^g t_i^3 - t_i}{N^3 - N}$, missä g on niiden lukuarvoryhmien määrä, joissa

tasatuloksia esiintyy ja t_i on tällaisissa ryhmissä olevien lukujen määrä.

Esim. Gastryl- esimerkkiaineisto 1 (Lähde: MedStat manuaali sivut 18.1- 18.2).

"Gastryl"-nimisen lääkkeen imeytymiskoe. Tabletteja oli päällystemateriaaliltaan kolmea tyyppiä A, B ja C. Tutkijat halusivat testata riippuko seerumin konsentraatio tablettityypistä? Koehenkilöitä oli yhteensä 24. Kuhunkin ryhmistä A, B ja C satunnaistettiin 8 koehenkilöä. Aineisto on fiktiivinen.

Potnro	Tyyppi	Konsentraatio	Järjestysluku R_i	Summa $\sum R_i$	Keskiarvo \bar{R}_i
1	A	70	8	61	7,63
2	A	63	6		
3	A	42	2		
4	A	83	9,5		
5	A	20	1		
6	A	97	15,5		
7	A	46	3,5		
8	A	97	15,5		
9	B	49	5	107,5	13,44
10	B	89	13		
11	B	67	7		
12	B	133	23		
13	B	118	20		
14	B	46	3,5		
15	B	140	24		
16	B	85	12		
17	C	83	9,5	17,5	
18	C	132	22		
19	C	99	17,5		

20	C	109	19	
21	C	99	17,5	
22	C	84	11	
23	C	90	14	
24	C	122	21	131,5 16,44

Taulukon järjestysluvut on tehty SPSS:n "Transform" valikossa seuraavasti:

Rank Cases

Variable(s):
Konsentraatio

By:

Assign Rank 1 to
 Smallest value
 Largest value

Display summary tables

Rank Types... Ties...

Rank Cases: Ties

Rank Assigned to Ties
 Mean
 Low
 High
 Sequential ranks to unique values

$$\text{Testisuure: } H = \frac{12}{24 \cdot 25} \cdot (8 \cdot 7.63^2 + 8 \cdot 13.43^2 + 8 \cdot 16.43^2) - 3 \cdot (24 + 1) = 6,42$$

Korjattu testisuure: Gastryl-esimerkkiaineistossa $g = 4$ (tasatulokset 46, 83, 97 ja 99) ja kaikissa on kaksi lukua, joten kaikki $t_i = 2$. Koska $N = 24$, on korjaus siten: $1 - 4 \cdot (2^3 - 2) / (24^3 - 24) = 0,998$, joten **korjattu $H = 6,43$** . Koska vapausasteet ovat $k - 1 = 2$, saadaan χ^2 -jakauman perusteella (StaTable) **$P = 0,040$** (kaksipuolinen)

SPSS: Valikot "Analyze" ► "Nonparametric tests" ► "K Independent-Samples".

Kohdassa "Test Type" testiksi valitaan "Kruskal–Wallis H" ja kohtaan "Test variable" viedään muuttuja "Konsentraatio" ja kohtaan "Grouping variable" muuttuja "Tyyppi" ja määritellään sen koodiarvojen vaihtelualue 1-3.

Tulos:

Ranks			
	Tyyppi	N	Mean Rank
Konsentraatio	A	8	7,63
	B	8	13,44
	C	8	16,44
	Total	24	

Test Statistics^{a,b}

	Konsentraatio
Chi-Square	6,435
df	2
Asymp. Sig.	,040

a. Kruskal Wallis Test b. Grouping Variable: Tyyppi

Tulkinta: Tablettityyppien välillä on tilastollisesti merkitsevä ero

Dunnin testi Kruskal–Wallisin testin yhteydessä

Mikäli Kruskal–Wallisin testi on antanut tilastollisesti merkitsevän eron eri ryhmien välillä, niin seuraavaksi tutkijalla yleensä herää kysymys: Mitkä ryhmät eroavat toisistaan? Tällöin syntyy monivertailutilanne ja α -tasoa pitää korjata vertailujen lukumäärällä. Testinä voidaan käyttää Bonferronin korjaukseen perustuvaa Dunnin-monivertailutestiä (Dunn 1964).

Käyttötilanne:

Ryhmiä halutaan verrata joko:

a) pareittain keskenään, jolloin vertailuiden määrä on $k \cdot (k - 1) / 2$ ja **Bonferroni** korjattu merkitsevyys taso on siten:

$$\alpha^* = \frac{\alpha}{k \cdot (k - 1) / 2}$$

b) ryhmiä kontrolli (tai lume) ryhmään, vertailuiden määrä on $k - 1$ ja **Bonferroni** korjattu merkitsevyys taso on siten:

$$\alpha^* = \frac{\alpha}{k - 1}$$

Monet tilastopakettit (esim. SPSS) eivät jostain syystä laske Kruskal–Wallisin testin yhteydessä Dunnin monivertailutestiä. SPSS tulostaa kuitenkin järjestyslukujen keskiarvot \bar{R}_i kullekin ryhmälle, joiden avulla voidaan laskea esim. Excelillä Dunnin testiin liittyvät P-arvot. Laskenta tehdään seuraavasti:

1. Lasketaan ensin

$$z_{\alpha^*} = \frac{|\bar{R}_i - \bar{R}_j|}{\sqrt{\frac{N \cdot (N + 1)}{12} \cdot \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

Kaavassa $\alpha^* = \frac{\alpha}{k \cdot (k - 1) / 2}$, $N = \sum_{i=1}^k n_i$ ja z_{α^*} on normaalijakauman prosenttipiste.

2. z_{α^*} -arvoa vastaava P-arvo saadaan Excelillä kaavalla **2*(1-NORMDIST(z_{α^*}))**

3. Näin saatu P-arvo kerrotaan vertailuiden lukumäärällä, $k \cdot (k - 1) / 2$, jos verrataan pareittain kaikkia ryhmiä keskenään tai $k - 1$, jos ryhmiä verrataan kontrolliryhmään.

Esim. Gastryl-tutkimus. Monivertailut:

Tabletti- tyyppi	\bar{R}_i	Tabletti- tyyppi	\bar{R}_i	$ \bar{R}_i - \bar{R}_j $	P-arvo
A	7,63	B	13,44	5,81	0,3005
A	7,63	C	16,44	8,81	0,0380
B	13,44	C	16,44	3,00	1.0000

Tulkinta: Tablettityyppien A ja C välillä on tilastollisesti merkitsevä ero.

Esim.

Ryhmiä A ja C välinen monivertailukorjattu P-arvo saadaan Excel-ohjelmalla <http://www.kttl.helsinki.fi/sama/stats/DunnKruskal-Wallis.xls> seuraavasti:

Dunnin monivertailutesti Kruskal–Wallisin testille						
k	N	n_i	n_j	mean R_i	mean R_j	Abs(mean R_i - mean R_j)
3	24	8	8	7,63	16,44	8,8125
$k*(k-1)/2$	$N*(N+1)/12$	$1/n_i+1/n_j$	$\sqrt{B_4*C_4}$			z(alfa*)
3	50	0,250	3,536			2,4926
						p(alfa*)
						0,0127
						P-arvo
Kaikki parittaiset vertailut:						0,0380
Vertailu kontrolliryhmään:						0,0254
Selitykset:						
k=ryhmien lukumäärä						
N=havaintoyksiköiden kokonaislukumäärä						
n_i =havaintoyksiköiden lukumäärä ryhmässä i						
n_j =havaintoyksiköiden lukumäärä ryhmässä j						
mean R_i =järjestyslukujen keskiarvo ryhmässä i						
mean R_j =järjestyslukujen keskiarvo ryhmässä j						
(esim. SPSS laskee nämä)						

Jonckheere–Terpstra trenditesti

Jonckheere–Terpstran testillä etsitään tietyn tyyppistä poikkeamaa rivien ja sarakkeiden riippumattomuudesta ($r \times r$)-ristiintaulukosta, jossa rivi- ja sarakeluokilla on luonnollinen järjestys. Esim. oletetaan, että rivit ovat hoitoryhmiä kasvavan lääkeannoksen mukaisesti ja sarakkeet edustavat vastetta, joka on järjestysasteikollinen. Ks. myös "Linear by linear"-testi.

Jonckheere–Terpstra on Kruskal–Wallisin trenditesti. Se testaa mediaanien monotonista järjestystä (nousevaa/laskevaa). Testisuureen J laskenta tapahtuu seuraavasti: jos vertailtavia ryhmiä on k kappaletta, niin ensin lasketaan kaikki $k*(k-1)/2$ kappaletta Mann–Whitneyn U-testejä ryhmäparien välillä. Suure J saadaan näiden summana.

Lopuksi J-suure muunnetaan standardoiduksi testisuureeksi, joka noudattaa likimain normaalijakaumaa.

Käyttötilanne:

Sekä testattava muuttuja että ryhmittelevä muuttuja ovat järjestysasteikollisia, ts. vertailtavat ryhmät muodostavat mielekkään järjestyksen; esim. eri lääkeannoksilla hoidetut potilaat.

Esim.

Edellä esitelty Gastryl-tutkimus.

Ohjelmalla SPSS Jonckheeren testisuureeksi saadaan $J = 141,0$, standardoiduksi suureeksi 2,383 ja yksisuuntaiseksi P-arvoksi 0,008. Todetaan, että ryhmämediaanien 66,5, 87,0 ja 99,0 välillä on tilastollisesti merkitsevä nouseva trendi.

SPSS: Valikot ”Analyze” ► ”Nonparametric tests” ► ”K Independent-Samples”. Kohdassa ”Test Type” testiksi valitaan ”Jonckheere-Terpstra” ja kohtaan ”Test variable” viedään muuttuja ”Konsentraatio” ja kohtaan ”Grouping variable” muuttuja ”Tyyppi” ja määritellään sen koodiarvot.

Tulos:

Jonckheere-Terpstra Test ^a	
	Konsentraatio
Number of Levels in Tyyppi	3
N	24
Observed J-T Statistic	141,000
Mean J-T Statistic	96,000
Std. Deviation of J-T Statistic	18,885
Std. J-T Statistic	2,383
Asymp. Sig. (2-tailed)	,017

a. Grouping Variable: Tyyppi

Tulkinta: Tablettityyppien A, B ja C jakaumien lokalisaatioiden (mediaanien) välillä on tilastollisesti merkitsevä ($P=0,017$) monotoninen (kasvava) trendi. Kruskal–Wallisin testi antoi tässä tapauksessa suuremman P-arvon, koska se testaa vain jakaumien lokalisaation heterogeenisyyttä, eikä trendiä, joka tässä tapauksessa on olemassa.

Friedmanin testi ja Pagen testi

Friedmanin testi on jakaumavapaa, eli parametriton yhden ryhmän toistomittausten varianssianalyysille; esim. siten, että n:stä henkilöstä on toistomittauksia, eri olosuhteissa, eri lääkeannoksilla, eri menetelmillä jne. Testiä käytetään silloin, kun parametristen menetelmien vaatimat normaalisuus ja/tai varianssien homogeenisuusoletukset eivät ole voimassa. Friedmanin testi perustuu ainoastaan havaintojen järjestyslukuihin.

Käyttötilanne:

Esim. joukko henkilöitä on tutkittu eri olosuhteissa tai eri arvioijien toimesta. Halutaan tutkia, onko näiden toistojen välillä eroa.

	1	2	...	k
1	y_{11}	y_{12}	...	y_{1k}
2	y_{21}	y_{22}	...	y_{2k}
.
.
.
k	y_{n1}	y_{n2}	...	y_{nk}

Muodostetaan järjestysluvut 1, ..., k riveittäin esim. nousevan suuruusjärjestyksen mukaisesti.

	1	2	...	k
1	r_{11}	r_{12}	...	r_{1k}
2	r_{21}	r_{22}	...	r_{2k}
.
.
.
k	r_{n1}	r_{n2}	...	r_{nk}
Σ	$r_{\cdot 1}$	$r_{\cdot 2}$...	$r_{\cdot k}$

r_{ij} on jokin luvuista 1, 2, ..., k jokaisella i:n arvolla. Friedmanin testi testaa, eroavatko sarakesummat $r_{\cdot j}$ toisistaan.

Testisuure:

$$S = \frac{12}{n \cdot k \cdot (k + 1)} \sum_{j=1}^k r_{\cdot j}^2 - 3 \cdot n \cdot (k + 1)$$

Testisuure S noudattaa likimain χ^2 -jakaumaa vapausastein $df = k - 1$, edellyttäen että k ja/tai n eivät ole kovin pieniä.

Esim. Gastryl-esimerkkiaineisto 2.

Tablettien antotapoja oli kolme: E = tyhjään mahaan, A = antasidin kanssa ja M = aterian jälkeen. Koehenkilöitä oli kahdeksan. Testataan hypoteesia lääkkeen antotapojen välisistä eroista välittämättä tabletin päällysteestä.

Henkilö	E	r_{i1}	A	r_{i2}	M	r_{i3}
1	65	2	63	1	74	3
2	75	1	110	3	99	2
3	65	1	73	3	70	2
4	104	2	103	1	118	3
5	62	1	88	3	87	2
6	68	1	81	3	78	2
7	83	1	94	2	99	3
8	93	1	113	3	98	2
$R_j = \Sigma r_{ij}$		10		19		19
\bar{R}_j		1,25		2,375		2,375

Testisuure: $S = \frac{12}{8 \cdot 3 \cdot (3 + 1)} (10^2 + 19^2 + 19^2) - 3 \cdot 8 \cdot (3 + 1) = 6,75, P=0,0342$

SPSS: Valikot "Analyze" ► "Nonparametric tests" ► "K Related Samples". Kohdassa "Test Type" testiksi valitaan "Friedman" ja kohtaan "Test Variables" vietään muuttujat E, A ja M ja kohdasta "Exact" klikataan "Exact", koska aineisto on pieni.

Ranks		Test Statistics ^a	
	Mean Rank		
E	1,25	N	8
A	2,38	Chi-Square	6,750
M	2,38	df	2
		Asymp. Sig.	,0342
		Exact Sig.	,0375
		Point Probability	,0072

a. Friedman Test

Tulkinta: Eri antotapojen välillä on tilastollisesti merkitsevä ero.

Pagen testi on Friedmanin testin yhteydessä käytetty trenditesti. Se on vastaavanlainen trenditesti kuin Jonckheere–Terpstran testi on Kruskal–Wallisin testin yhteydessä. Jos Friedmanin testin yhteydessä käytetty asetelma on esim. sellainen, että samoilta henkilöiltä määritetyt toistot ovat responssimittauksia eri lääkeannoksilla, niin Pagen testillä voidaan tutkia kasvaako tai laskeeko responssi lääkeannoksen mukaan monotonisesti. Friedmanin testi antaisi vain vastauksen siihen onko lääkeannosten suhteen eroa.

Pagen testi voidaan laskea mm. ohjelmalla StatXact 8. SPSS ei sitä kuitenkaan laske, mutta ilmoittaa Friedmanin testin yhteydessä kullekin toistomuuttujalle järjestyslukujen (R_j) keskiarvot ("Mean Rank"). Sijoittamalla ne osoitteessa <http://www.mv.helsinki.fi/home/sarna/stats/PagenTrenditesti.xls> olevaan Excel-ohjelmaan saadaan laskettua Pagen trenditestille asymptoottinen P-arvo.

Laskenta perustuu testisuureeseen:

$$z_L = \frac{12 \cdot L - 3 \cdot n \cdot k \cdot (k + 1)^2}{k \cdot (k^2 - 1)} \sqrt{\frac{k - 1}{n}}, \text{ missä } L = \sum_{j=1}^k j \cdot R_j$$

z_L noudattaa likimain standardia Normaalijakaumaa.

Esim. Gastryl-esimerkkiaineisto 2.

Pagen trenditesti										
Järjestyslukujen keskarvot toistoissa ("Mean Rank" Friedmanin testin tulosteessa)										
Toistot										
	1	2	3	4	5	6	7	8	9	10
	1,250	2,375	2,375	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Toistojen määrä										
k =	3									
Henkilöiden määrä										
n =	8									
L =				105			z _L =		2,25	
Asymptoottinen P-arvo							1-puolinen		0,0122	
							2-puolinen		0,0244	

Tulkinta:

Eri antotapojen mediaanien välillä on monotoninen trendi, $P=0,0244$. Yksipuolista testausta käytetään silloin, kun ennalta tiedetään trendin suunta, ylöspäin vai alaspäin.

Dunnin testi Friedmanin testin yhteydessä

Mikäli Friedmanin testi on antanut tilastollisesti merkitsevän eron eri ryhmien välillä, niin seuraavaksi tutkijalla yleensä herää kysymys: Mitkä ryhmät eroavat toisistaan? Tällöin syntyy monivertailutilanne ja α -tasoa pitää korjata vertailujen lukumäärällä. Testinä voidaan käyttää Bonferronin korjaukseen perustuvaa Dunnin-monivertailutestiä (Dunn 1964).

Käyttötilanne:

Ryhmiä halutaan verrata joko:

a) pareittain keskenään, jolloin vertailuiden määrä on $k \cdot (k - 1) / 2$ ja **Bonferroni** korjattu merkitsevyys taso on siten:

$$\alpha^* = \frac{\alpha}{k \cdot (k - 1) / 2}$$

b) ryhmiä kontrolli (tai lume) ryhmään, vertailuiden määrä on $k - 1$ ja **Bonferroni** korjattu merkitsevyys taso on siten:

$$\alpha^* = \frac{\alpha}{k - 1}$$

Monet tilastopakettit (esim. SPSS) eivät jostain syystä laske Kruskal–Wallisin testin yhteydessä Dunnin monivertailutestiä. SPSS tulostaa kuitenkin järjestyslukujen keskiarvot \bar{R}_i kullekin ryhmälle, joiden avulla voidaan laskea esim. Excelillä Dunnin testiin liittyvät P-arvot. Laskenta tehdään seuraavasti:

1. Lasketaan ensin

$$z_{\alpha^*} = \frac{|\bar{R}_i - \bar{R}_j|}{\sqrt{\frac{k \cdot (k + 1)}{6 \cdot N}}}$$

Kaavassa $\alpha^* = \frac{\alpha}{k \cdot (k - 1) / 2}$, N on havaintoyksiköiden määrä ja z_{α^*} on normaalijakauman prosenttipiste.

2. z_{α^*} -arvoa vastaava P-arvo saadaan Excelillä kaavalla **$2 \cdot (1 - \text{NORMDIST}(z_{\alpha^*}))$**

3. Näin saatu P-arvo kerrotaan vertailuiden lukumäärällä, $k \cdot (k - 1) / 2$, jos verrataan pareittain kaikkia ryhmiä keskenään tai $k - 1$, jos ryhmiä verrataan kontrolliryhmään.

Esim. Gastryl-esimerkkiaineisto.

Antotapa	\bar{R}_i	Antotapa	\bar{R}_i	Erotus	P-arvo
E	1,25	A	2,38	1,13	0,0733

E	1,25	M	2,38	1,13	0,0733
A	2,38	M	2,38	0	1.00

Esim. Antotapojen E ja A välinen monivertailukorjattu P-arvo saadaan Excel-ohjelmalla <http://www.mv.helsinki.fi/home/sarna/stats/DunnFriedman.xls> seuraavasti:

Dunnin monivertailutesti Friedmanin testille					
k	N	meanR _i	meanR _j	Abs(meanR _i -meanR _j)	
3	8	1,250	2,375	1,125	
$k*(k-1)/2$	$\sqrt{k(k+1)/6N}$			$z(\text{alfa}^*)$	$p(\text{alfa}^*)$
3	0,5000			2,25	0,024448867
					P-arvo
Kaikki parittaiset vertailut:					0,0733
Selitykset: Vertailu kontrolliryhmään:					0,0489
k=toistojen lukumäärä					
N=havaintoyksiköiden lukumäärä toistoissa					
meanR _i =järjestyslukujen keskiarvo toistossa i (Esim. SPSS					
meanR _j =järjestyslukujen keskiarvo toistossa j Laskee nämä)					

Tulkinta:

Koska kaikissa parivertailuissa $P > 0,05$, niin monivertailukorjattuna antotapojen välillä ei ole tilastollisesti merkitsevää eroa.

Kaksisuuntainen parametriton varianssianalyysi

Käyttötilanne:

Esim. kahden tekijän yhdistelykoe, kun mittaukset ovat joko

- järjestysasteikolla tai
- välimatka/suhdeasteikolla ja aineisto on pieni.

Testattavat hypoteesit:

- Onko tekijöiden välillä interaktioita, siis onko toisen tekijän vaikutus erilainen toisen tekijän eri ryhmissä
- Mikäli interaktioita ei esiinny on mielekästä testata kummankin tekijän vaikutuksia erikseen Friedmanin testeillä, muuten ei.

Esim. Gastryl-esimerkkiaineisto 3.

Yhdistelykoe 8 koehenkilöllä. Jokainen koehenkilöistä testaa kaikkia kolmea tablettityyppiä A, B ja C ja kullakin tablettityypillä jokaista lääkkeenantotapaa E, A ja M, missä E = tyhjään mahaan, A = antasidin kanssa ja M = aterian jälkeen. Halutaan testata onko tablettityypin ja antotavan välillä interaktiota, ts. onko tablettityypillä eri vaikutus lopputulokseen riippuen antotavasta.

Potnro	Tabletti-tyyppi	Antotapa	Konsentraatio		Potnro	Tabletti-tyyppi	Antotapa	Konsentraatio
1	A	E	33		1	C	A	34
2	A	E	22		2	C	A	60
3	A	E	18		3	C	A	35

4	A	E	32		4	C	A	37
5	A	E	7		5	C	A	42
6	A	E	35		6	C	A	35
7	A	E	18		7	C	A	40
8	A	E	41		8	C	A	51
1	B	E	18		1	A	M	19
2	B	E	31		2	A	M	21
3	B	E	24		3	A	M	8
4	B	E	50		4	A	M	26
5	B	E	39		5	A	M	5
6	B	E	17		6	A	M	30
7	B	E	50		7	A	M	17
8	B	E	30		8	A	M	23
1	C	E	14		1	B	M	20
2	C	E	22		2	B	M	28
3	C	E	23		3	B	M	21
4	C	E	22		4	B	M	42
5	C	E	16		5	B	M	41
6	C	E	16		6	B	M	15
7	C	E	15		7	B	M	47
8	C	E	22		8	B	M	26
1	A	A	18		1	C	M	35
2	A	A	20		2	C	M	50
3	A	A	16		3	C	M	41
4	A	A	25		4	C	M	50
5	A	A	8		5	C	M	41
6	A	A	32		6	C	M	33
7	A	A	11		7	C	M	35
8	A	A	33		8	C	M	49
1	B	A	11					
2	B	A	30					
3	B	A	22					
4	B	A	41					
5	B	A	38					
6	B	A	14					
7	B	A	43					
8	B	A	29					

Tablettityypin ja antotavan välinen interaktio voidaan testata Friedmanin testillä seuraavasti:

1) Lasketaan tablettityyppien A ja B erotusmuuttujat "AE-BE", "AA-BA" ja "AM-BM" antotavoille E, A ja M. Esim. koehenkilölle 1 antotavalla E erotus on $(33 - 18) = 15$. Tehdään Friedmanin testi näillä erotuksilla.

Ranks		Test Statistics ^a	
	Mean Rank	N	8
AE-BE	2,19	Chi-Square	1,357
AA-BA	2,13	df	2
AM-BM	1,69	Asymp. Sig.	,507

Exact Sig.	,562
Point Probability	,053
a. Friedman Test	

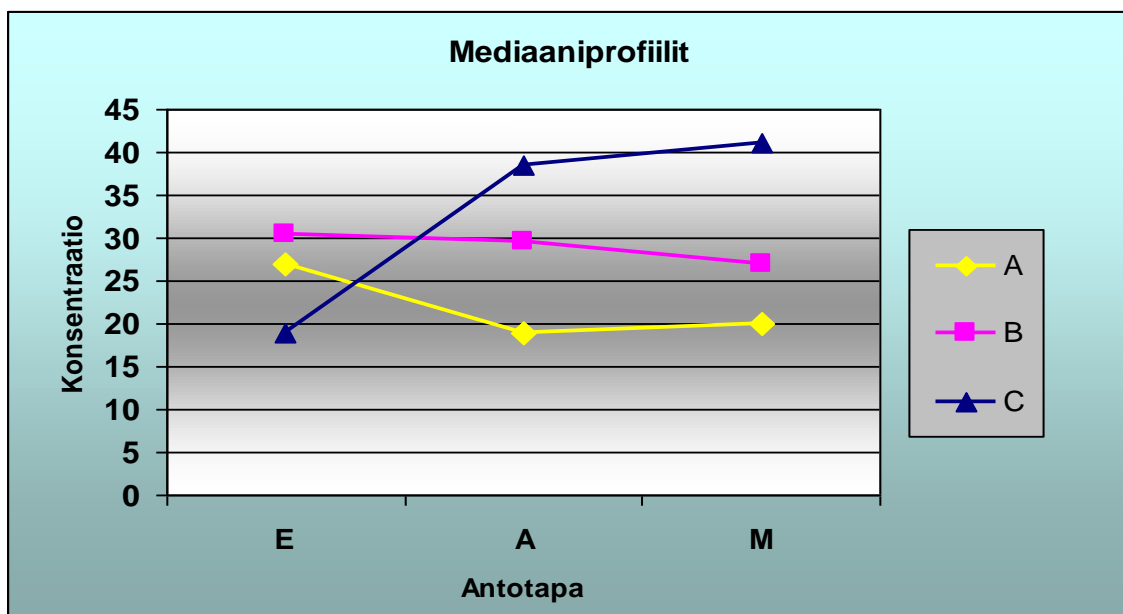
2) Verrataan vastaavalla tavalla tablettityyppien A ja B keskiarvotuloksia tablettityyppiin C, ts. muodostetaan erotusmuuttujat $(AE+BE)/2-CE$, $(AA+BA)/2-CA$ ja $(AM+BM)/2-CM$ ja lasketaan Friedmanin testi näillä muuttujilla:

Ranks		Test Statistics ^a	
	Mean Rank	N	8
$(AE+BE)/2-CE$	3,00	Chi-Square	12,250
$(AA+BA)/2-CA$	1,38	df	2
$(AM+BM)/2-CM$	1,63	Asymp. Sig.	,002
		Exact Sig.	,001
		Point Probability	,001
a. Friedman Test			

3) Lasketaan kohdista 1) ja 2) χ^2 -testisuueet yhteen ja vastaavasti vapausasteet yhteen. Näin saadaan $\chi^2=(1,357+12,250)=13,607$ ja $df=4$. Excelin funktiolla CHIDIST(13,607,4) saadaan $P=0,0087$.

Tulkinta: Lääkkeen tablettityypin ja antotavan välillä on voimakas interaktio.

Interaktion luonteen selvittämiseksi kannattaa piirtää antotapojen mediaaniprofiilit eri tablettityypeille. Ne voidaan piirtää esim. Excelillä seuraavasti.



Tulkinta:

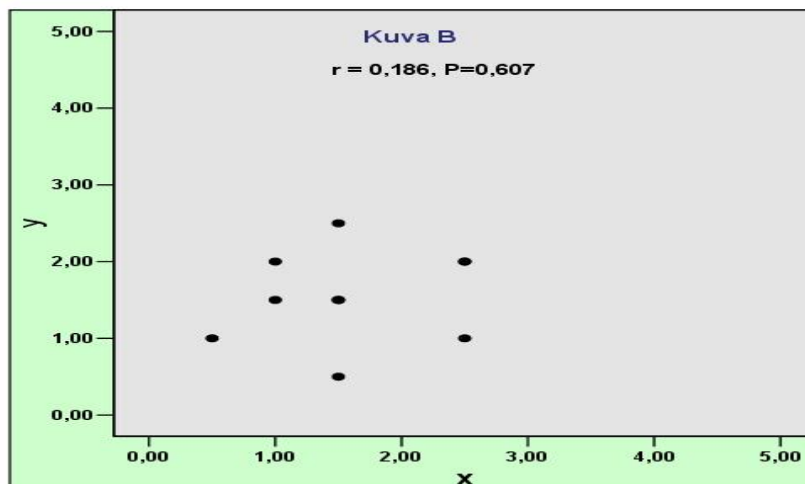
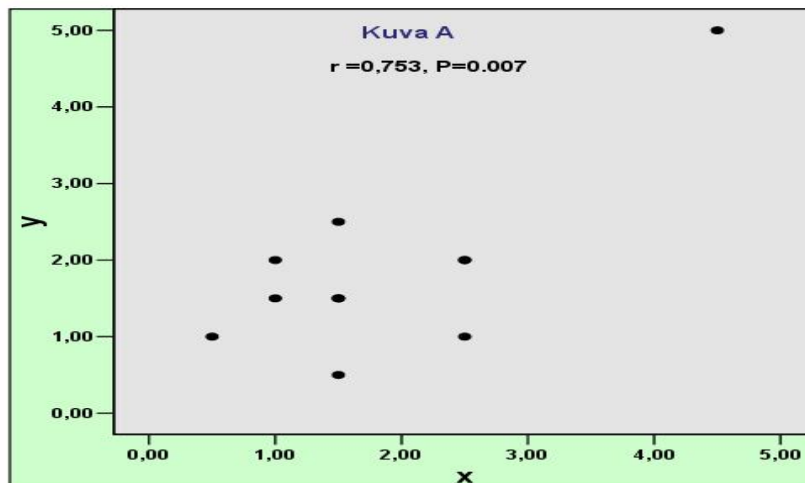
Tablettityypillä C lopputulos riippuu voimakkaasti antotavasta. Muilla tablettityypeillä antotavan vaikutus on pienempi.

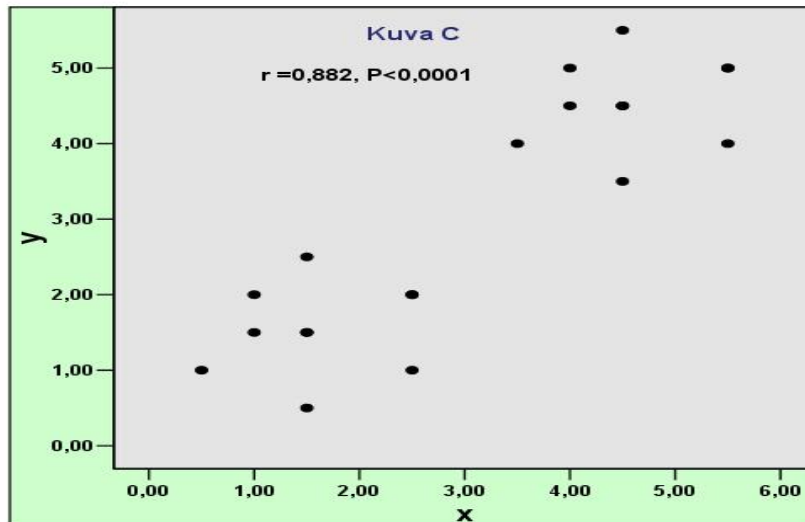
Pearsonin korrelaatiokerroin

Pearsonin korrelaatiokerroin, "Pearson's correlation coefficient" ($r_{x,y}$) Siitä käytetään myös nimitystä "tulo-momentti"-korrelaatiokerron. Se on parametrinen lineaarisen riippuvuuden mitta kvantitatiivisten suureiden x ja y välillä. Parametrinen tarkoittaa, että tarkasteltavien muuttujien x ja y yhteisjakauman tulisi olla likimain normaalin. Muussa tapauksessa esim. poikkeavilla havaintoarvoilla voi olla suuri vaikutus r :n arvoon. r voi saada minkä tahansa arvon väliltä $[-1, +1]$, ja $r = -1$ merkitsee täydellistä lineaarista negatiivista (käänteistä) ja $+1$ positiivista riippuvuutta. $r = 0$ merkitsee, ettei x :n ja y :n välillä ole ollenkaan lineaarista riippuvuutta. $r = 0$ ei merkitse, ettei niiden välillä voisi jotain muunlaista epälineaarista riippuvuutta.

Huom.

Pearsonin korrelaatiokerrointa ylikäytetään, väärinkäytetään ja myös tulkitaan usein väärin lääketieteellisissä artikkeleissa. Esim. korrelaatiokertoimet voivat olla harhaanjohtavia mikäli aineistossa on selvästi muista arvoista poikkeavia havaintoarvoja (kuvat A ja B) tai mikäli aineisto muodostuu osaryhmistä, joiden tiedetään eroavan keskiarvotasolla molempien korreloitavien muuttujien suhteen (kuva C).





Käyttö:

Jotta r olisi validi assosiaation mitta, niin seuraavat vaatimukset tulisi täyttyä:

- muuttujien x ja y yhteisjakauman tulisi olla **normaalinen** (ellei vaatimus täyty edes pitäisi turvautua parametrittomiin korrelaatioihin)
- muuttujien x ja y välillä **ei saa olla teknisiä riippuvuuksia**, jotka säätelevät muuttujien yhteisvaihtelua
- kustakin henkilöstä saa olla vain **yksi havaintopari** aineistossa, josta korrelaatio lasketaan, esim. jos 30 henkilöstä on 3 toistomittausta, niin korrelaatiota ei saa laskea siten, että havaintoarvoja olisi 90! (toistomittausten tapauksessa pitää käyttää toistomittausanalyysijä)

Pearsonin korrelaatiokerroin ja sen keskivirhe lasketaan kaavoilla:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad SE(r) = \sqrt{\frac{1-r^2}{n-2}}$$

Testi:

Korrelaatiokertoimen tilastollista merkitsevyyttä voidaan arvioida testisuurella:

$$t = \frac{r}{SE(r)} = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

Testisuure noudattaa Studentin t-jakaumaa vapausastein (n - 2).

Excel-ohjelma: <http://www.mv.helsinki.fi/home/sarna/stats/PearsonP-arvo.xls>

Luottamusväli:

Korrelaatiokertoimelle kannattaa laskea myös luottamusväli. Luottamusväli antaa paremman käsityksen kuin pelkkä P-arvo siitä, kuinka varma johtopäätös muuttujien x ja y assosiaatiosta voidaan tehdä. Jostain syystä monet tilastopakettit, kuten SPSS, eivät laske r:lle luottamusväliä. Erityisohjelmilla, kuten CIA ja StatXact 8, väli voidaan laskea. Se on helppo ohjelmoida myös Excelille käyttäen seuraavia kaavoja:

1. Lasketaan suureet

$$Z = \frac{1}{2} \cdot \log_e \frac{1+r}{1-r}, \quad A = Z - \frac{Z_{1-\alpha/2}}{\sqrt{n-3}} \quad \text{ja} \quad B = Z + \frac{Z_{1-\alpha/2}}{\sqrt{n-3}}$$

2. ja niiden avulla saadaan luottamusväliksi

$$\text{Alaraja} : \frac{e^{2A} - 1}{e^{2A} + 1} \quad \text{ja} \quad \text{Yläraja} : \frac{e^{2B} - 1}{e^{2B} + 1}$$

Excel-laskin: <http://mv.helsinki.fi/home/sarna/stats/PearsonCL.xls>

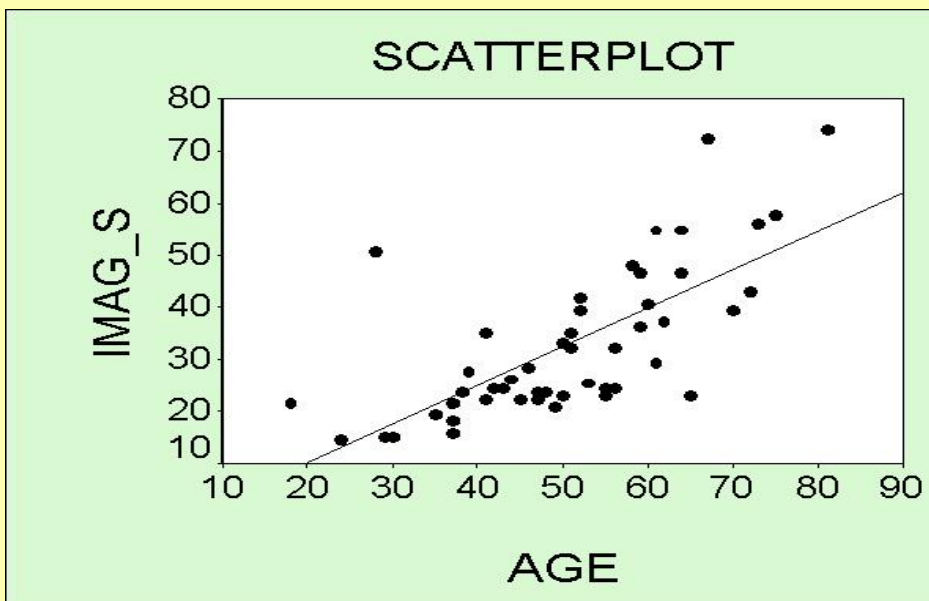
Kahdesta ryhmästä tai kahdesta eri aineistosta laskettujen korrelaatiokerrointen välisen eron tilastollista merkittävyyttä voidaan arvioida z-testillä, joka perustuu yllä olevaan Fisherin transformaatioon Z.

Excel-laskin: <http://mv.helsinki.fi/home/sarna/stats/z-testiKorrelaatiokerroinvertailu.xls>

Esim.

Aineisto:Tietokanta TRIAL. Halutaan tutkia korreloivatko muuttujat "Age" ja "Imag_S" keskenään.

SPSS: Valikot: **"Analyze"** ▶ **"Regression"** ▶ **"Curve Estimation"**:



SPSS: Valikot **"Analyze"** ▶ **"Correlate"** ▶ **"Bivariate"**. Kohdasta **"Correlation Coefficients"** valitaan **"Pearson"**

Correlations

		Age	Imag_s
Imag_s	Pearson Correlation	,708**	1
	Sig. (2-tailed)	,000	
	N	49	49

** . Correlation is significant at the 0.01 level (2-tailed).

Tulkinta:

Saatu korrelaatiokerroin $r = 0,708$ on tilastollisesti merkitsevä, $P < 0,001$, mutta tilastollisen merkitsevyyden lisäksi kannattaa laskea myös luottamusväli. Luottamusvälin pituus riippuu aineistokoosta. Tässä tapauksessa 95 %:n luottamusväliksi saadaan.

Korrelaatiokerroimen 95 % luottamusväli			
n	r		
49	0,7080	0,5329	0,8250

Eli tämän kokoisella aineistolla arvioituna iän ja Imag_S:n välinen korrelaatio perusjoukossa olisi 95 %:n varmuudella tällä välillä.

Korrelatiivisia menetelmiä, esimerkiksi tavallista lineaarista regressiota, käytetään usein, kun osa muuttujista on järjestysasteikollisia. Tällöin tulisi kuitenkin olla erityisen varovainen, sillä järjestysasteikollisten ja jatkuvien muuttujien väliset korrelaatiot saattavat helposti vääristyä, koska harvoin järjestysasteikollisten muuttujien arvojoukko on tasavälinen. Jatkuvien muuttujien ja dikotomisten muuttujien välisissä korrelaatioissa ei sen sijaan ole mitään laskennallista ongelmaa. Sen sijaan testaamisen suhteen on, sillä yllä oleva testisuure toimii tällöin vain likimääräisesti.

Huom.

Korrelaatiokerroimia tulkittaessa kannattaa muistaa seuraavat asiat:

- korrelaatiokerroin on tulkinnallisesti eri asia kuin regressiokerroin
- korrelaatio on symmetrinen assosiaation mitta, $r_{xy} = r_{yx}$
- korrelaatio ja kausaaliiteetti ovat eri asioita; voimakas korrelaatio ei takaa kausaalista riippuvuutta
- tilastollisesti merkitsevä korrelaatio on eri asia kuin kliinisesti merkittävä korrelaatio

Aineistokokoarvio:

Korrelaatiokerroimen perusteella voidaan laskea myös aineistokokoarvioita (**n**). Likimääräinen arvio n:lle saadaan kaavalla:

$$n = \left(\frac{z_{1-\alpha/2} + z_{1-\beta} \cdot \sqrt{1-r^2}}{r} \right)^2 + 2$$

Kaavassa $z_{1-\alpha/2}$ ja $z_{1-\beta}$ ovat normaalijakauman prosenttipisteitä. Vastaavasti, jos tiedetään aineistokoko **n** ja määritellään merkitsevyytensä α , niin yllä olevan kaavan perusteella voidaan laskea tietyn suuruisten korrelaatiokerrointen merkitsevyydestin voima $1-\beta$.

Excel-ohjelma: <http://www.mv.helsinki.fi/home/sarna/stats/PearsonPower.xls>

Korrelaatiokertoimeen perustuva aineistokokoarvio			
r	α	Voima (1- β)	n
0,7000	0,050	0,90	18,87
z		1,9600	1,2816
Voiman arviointi			
r	α	n	Voima
0,5000	0,050	49	95,50 %
$Z_{1-\alpha}$		1,6949	

Tulkinta:

Korrelaatiokertoimen **0,7** toteamiseksi **90 %**:n voimalla ja merkitsevyystasolla **0,05** tarvitaan **19** henkilöä ja kääntäen jos tiedetään, että aineistokoko on **49**, kuten edellä olevassa esimerkissä, niin tällöin suuruusluokkaa **0,5** oleva korrelaatio voidaan todeta tilastollisesti merkitseväksi tasolla **0,05** voimalla **95,5 %**. Tarkempien arvioiden laskemiseksi täytyy käyttää iteratiivista menettelyä (Machin & Campbell 1987 s 89).

Spearmanin järjestyskorrelaatiokerroin

Spearmanin korrelaatiokerroin, (r_s) "Spearman's rho" on parametriton assosiaation mitta kvantitatiivisen tai järjestysasteikollisen suureen x ja y välillä. Mitta saa arvoja väliltä (-1,1). Sitä käytetään erityisesti pienissä aineistoissa. Se lasketaan siten, että aineisto lajitellaan x:n ja y:n suhteen ja annetaan havaintoarvoille järjestysluvut "ranks" ja lasketaan Pearsonin korrelaatiokerroin näiden järjestyslukujen perusteella.

Käyttö:

Spearmanin korrelaatiokerrointa r_s ("Spearman's rho") käytetään järjestysasteikollisten muuttujien välisenä assosiaation mittana. Se on yleisesti käytetty mitta kliinisissä tutkimuksissa, koska se ei edellytä normaalisuutta, eikä ole herkkä poikkeaville havaintoarvoille, joita yleisesti tutkimusaineistoissa esiintyy. Ongelmana on, että monimuuttuja-analyyseissa sillä ei ole käyttöä.

Spearmanin korrelaatiokerroin lasketaan kaavalla:

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n^3 - n}$$

missä d_i on i. henkilön järjestyslukujen erotus korreloitavilla suureilla x ja y. Samojen lukuarvojen tapauksessa järjestyslukuna käytetään niitä vastaavien järjestyslukujen keskiarvoa.

Yllä oleva laskentakaava on alunperin johdettu siten, että Pearsonin kaavaan on sijoitettu alkuperäisten x:n ja y:n lukuarvojen paikalle järjestysluvut. Spearmanin korrelaatioiden tilastollisen merkitsevyyden arvioimiseksi pienillä aineistoilla ($n < 10$) tulisi käyttää eksakteja testejä (esim. StatXact 8 tai erityistaulukoita). Kun aineistokoko on välillä 10 - 30 voidaan käyttää Monte Carlo menetelmiä ja sitä suuremmilla aineistoilla Pearsonin korrelaatiokertoimen t-testiä.

Esim.

Tietokanta TRIAL. Halutaan tutkia korreloituvatko muuttujat "Age" ja "Imag_S" keskenään. Koska Imag_S:n ja iän yhteisjakauma ei ole täysin normaalin, lasketaan Spearmanin korrelaatiokerroin. Ohjelma SPSS antaa arvon $r_s=0,740$ ja sen kaksisuuntaiseksi P-arvoksi. Luottamusväliä SPSS ei laske. Ohjelmalla CIA saadaan 95 %:n luottamusväliksi (0,579, 0,845). Todetaan, että tässä tapauksessa Spearmanin korrelaatiokerroin antaa jonkin verran suuremman arvon kuin Pearsonin korrelaatiokerroin.

StatXact 8 antaa myös Spearmanin korrelaatiokertoimen 95 % - luottamusvälin.

Kendallin Tau ja Somerin D

Kendallin tau "Kendall's tau" on parametrin assosiaation mitta kvantitatiivisen tai järjestysasteikollisen suureen x ja y välillä. Mitta saa arvoja väliltä (-1,1). Sitä käytetään erityisesti pienissä aineistoissa, kuten Spearmanin korrelaatiokerrointakin. Kendallin tau:n avulla, toisin kuin Spearmanin kertoimella voidaan laskea myös osittaiskorrelaatioita $r_{xy.z}$, missä z on muuttuja, jonka vaikutus halutaan x:n ja y:n välisestä assosiaatiosta puhdistaa.

Somerin D, "Somer's D" Parametrin assosiaation mitta kvantitatiivisten tai järjestysasteikollisten suureiden x ja y välillä. Mitta saa arvoja väliltä (-1,1). Se on kilpaileva mitta Kendallin tau:lle.

Käyttö:

Kendallin tau ja Somerin D ovat vaihtoehtoisia mittoja Spearmanin korrelaatiokertoimelle järjestysasteikollisten muuttujien x ja y välisiä riippuvuussuhteita tarkasteltaessa.

Näiden mittojen laskentatapa on erilainen kuin Spearmanin korrelaatiokertoimessa. Siksi ne saattavat antaa hyvinkin erilaisen arvon kuin Spearman, yleensä numeerisesti pienemmän. Molemmat mitat perustuvat konkordanttien (sama tulos) ja diskordanttien (eri tulos) parien määriin eikä samalla tavalla järjestyslukuihin kuten Spearmanin korrelaatiokerroin. Kendallin Tau soveltuu käytettäväksi myös silloin, kun on tarve laskea osittaiskorrelaatioita, eli puhdistaa korrelaatiosta jonkun tai joidenkin muuttujien vaikutus. Somerin D on epäsymmetrinen assosiaation mitta toisin kuin Spearman ja Kendall. Se tarkoittaa, että toista assosioitavista muuttujista tarkastellaan riippuvana ja toista riippumattomana muuttujana.

Cohenin kappa ja painotettu kappa

Cohenin kappa "Cohen's kappa" on luokkamuuttujien välinen yhtäpitävyyden mitta. Mittaa yhtäpitävyyttä esim. kahden tai useamman arvioitsijan tai diagnostisen menettelytavan välillä. Cohenin kappa lasketaan pelkästään yhtäpitävien tulosten pohjalta.

Painotettu kapp "weighted kappa" on luokkamuuttujien välinen yhtäpitävyyden mitta. Mittaa yhtäpitävyyttä esim. kahden tai useamman arvioitsijan tai diagnostisen menettelytavan välillä. Poikkeaa Cohenin kappasta siten, että yhtäpitämättömyydelle annetaan erilaisia painoja.

Käyttö:

Kliinisissä tutkimuksissa usein esille tuleva tarve on arvioida kahden tai useamman mitan yhtäpitävyyttä. Esimerkiksi halutaan tutkia kahden tai useamman arvioitsijan antamien arvioiden tai kahden tai useamman diagnostisen menettelytavan yhtäpitävyyttä. Cohenin kapp ja painotettu kapp soveltuvat yhtäpitävyyden mitoiksi kyseiseen tarkoitukseen silloin, kun yhtäpitävyyttä tutkitaan luokkamuuttujia käyttäen.

Kapp mittaa sitä, kuinka paljon todettu yhtäpitävyys (p_o) poikkeaa pelkän sattuman perusteella odotettavissa olevasta yhtäpitävyydestä p_e .

Tarkastellaan aluksi tapausta, missä verrataan kahta diagnostista menettelyä tai kahden arvioitsijan antamaa arviota samoista näytteistä. Oletetaan, että käytetty luokitus on kaksiarvoinen, eli onko esim. diagnoosi "+ tai -", ja että arviointi tuloksista on muodostettu seuraavanlainen 2X2-taulukko:

		Arvioitsija 2/ Diagnostinen menetelmä 2		
		+	-	
Arvioitsija 1 / Diagnostinen Menetelmä 1	+	a	b	a+b
	-	c	d	c+d
		a+c	b+d	n

Cohenin (Cohen, 1960) **kapp** lasketaan kaavoilla:

$$\kappa = \frac{(p_o - p_e)}{(1 - p_e)}, \text{ missä } p_o = (a + d)/n \text{ ja } p_e = \frac{a+b}{n} \cdot \frac{a+c}{n} + \frac{c+d}{n} \cdot \frac{b+d}{n}$$

Suureen keskivirhe ja Normaalijakauma-aproksimaatioon perustuva likimääräinen $100(1-\alpha)\%$:n luottamusväli lasketaan kaavoilla:

$$\text{Keskivirhe : } SE(\kappa) = \sqrt{\frac{p_o(1-p_o)}{n(1-p_e)^2}}$$

$$\text{Alaraja : } \kappa - z_{1-\alpha/2} \cdot SE(\kappa) \text{ ja Yläraja : } \kappa + z_{1-\alpha/2} \cdot SE(\kappa)$$

Huom.

Ohjelmat CIA ja MedCalc käyttävät näitä laskentakaavoja. Sen sijaan StatXact 8 ja SPSS laskevat kappan keskivirheen hieman monimutkaisemmalla kaavalla (Fleiss, Cohen & Everitt, 1969). SPSS ei jostain syystä laske luottamusväliä ollenkaan. Osoitteessa <http://www.mv.helsinki.fi/home/sarna/stats/KappaCL.xls> oleva ohjelma antaa molemmat vaihtoehdot 2x2-taulukolle.

Esim.

Tietokanta TRIAL. Hepatomegalian diagnosointi kahden klinikon arvioimana. Muuttujat: "Obs_1" ja "Obs_2", jotka ovat (kyllä / ei)-tyyppisiä ilmaisten onko maksa palpoitavissa vai ei. Tulos:

		Arvioitsija 2		
		+	-	
Arvioitsija 1	+	18	6	24
	-	5	18	23
		23	24	47

Havaittu yhtäpitävyys on $p_o = (18 + 18) / 47 = 0,766$. Sattuman perusteella odotettavissa oleva yhtäpitävyys voidaan laskea todennäköisyyslaskennan peruslaskukaavoja käyttäen. "Kyllä-kyllä"-lokeron odotetuksi lukumääräksi tulee $24 \cdot 23 / 47 = 11,74$ ja "ei-ei"-lokerolle tässä tapauksessa sattumalta myös sama $23 \cdot 24 / 47 = 11,74$, joten $p_e = (11,74 + 11,74) / 47 = 0,500$ (kolmeen desimaaliin pyöristettynä).

Ohjelmalla KappaCL.xls (tai StatXact 8:llä) saadaan tulos:

Kapan 95 % luottamusväli									
		Arvioitsija 2			p_o	p_e	κ	SE(κ)	95% CL
		+	-						
Arvioitsija 1	+	18	6	24	0,7660	0,4998	0,5321	0,1234	0,2903
	-	5	18	23					
		23	24	47	Vaihtoehto2		0,1235	0,2901	0,7741
Yhtäpitävyys		76,6 %							

Esim.

Kahden kuvantamismenetelmän, CT:n ja MRI:n vertailu alaselän välilevyn pullistuman ja juurikompression toteamisessa. Aineistona 64 perättäistä avohoitopotilasta, jotka oli lähetetty Amsterdamin yliopistollisen sairaalan neurologian osastolle. Lähetediagnoosi oli LRS ("Lumbosacral Radicular Syndrome"). Potilailla epäiltiin välilevyn pullistumaa nikamavälissä L3-L4, L4-L5 tai L5-S1. Konservatiivinen hoito ei ollut tehonnut näihin potilaisiin. Heiltä tutkittiin kahden havainnoitsijan toimesta yhteensä 456 hermojuurta sekä CT:llä että MRI:llä. Diagnoosivaihtoehdot olivat: juurikompressio todettavissa "+" tai ei todettavissa "-". Haluttiin testata onko yhtäpitävyydessä eroa menetelmien CT ja MRI välillä.

Lähde: Rijn JC, Klemetso N, Reitsma JB et al. 1986.

		Arvioitsija 2			Arvioitsija 2			
		CT	+	-	MRI	+	-	
Arvioitsija 1	+		35	16	51	45	10	55
	-		24	381	405	12	389	401
			59	397	456	57	399	456

SPSS-ohjelmalla testi saadaan valikoista **"Analyze" ► "Descriptive Statistics" ► "Crosstabs"** määrittelemällä taulukon rivit ja sarakkeet ja valitsemalla kohdasta **"Statistics"** testiksi **"Kappa"**.

Koska aineisto on tiivistetyssä muodossa, niin määritellään CT painomuuttujaksi.

Weight Cases

Do not weight cases
 Weight cases by
 Frequency Variable: CT

Crosstabs

Row(s): Rivi
 Column(s): Sarake

Crosstabs: Statistics

Nominal by Interval
 Eta
 Kappa
 Rjisk
 McNemar

Symmetric Measures

CT		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement	Kappa	,587	,059	12,573	,000
N of Valid Cases		456			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Vastaavalla tavalla saadaan painomuuttujan MRI perusteella:

Symmetric Measures

MRI		Value	Asymp. Std. Error ^a	Approx. T ^b	Approx. Sig.
Measure of Agreement	Kappa	,776	,046	16,576	,000
N of Valid Cases		456			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

CT - ja MRI- menetelmillä saatujen kappo-arvojen 95 % - luottamusväliksi tulee **(0,471, 0,702)** ja vastaavasti **(0,686, 0,866)**. MRI-menetelmä antaa tilastollisesti merkitsevästi paremman yhtäpitävyyden (**P=0,0115**, z-testi). Testin voi tehdä osoitteessa: <http://www.kttl.helsinki.fi/sarna/stats/z-testiKahdelleSuureelle.xls>.

Teoreettisesti on mahdollista saada kapalle negatiivinen arvo, mutta käytännössä kappa vaihtelee kuitenkin välillä (0,1). Nolla merkitsee, ettei ole yhtäpitävyyttä ollenkaan ja yksi merkitsee täydellistä yhtäpitävyyttä. Yhtäpitävyyden voimakkuuden tulkitsemiseksi on esitetty monenlaisia nyrkkisääntöjä, esimerkiksi seuraavasti (Landis & Koch, 1977):

k	yhtäpitävyyden voimakkuus
0,00- 0,20	“olematon”
0,21- 0,40	“heikko”
0,41- 0,60	“kohtalainen”
0,61- 0,80	“hyvä”
0,81- 1,00	“erittäin hyvä”

Prevalenssiparadoksi

Yllä olevassa taulukossa esitetyt kappa-suureen rajat ovat korkeintaan vain suuntaa antavia, sillä kappa-suure riippuu voimakkaasti esim. diagnosoitavan taudin tai ilmiön prevalenssista. Esim. jos prevalenssi on suuruusluokkaa **0,50**, eli $(a+b)/n \approx (a+c)/n \approx 0,50$, niin saadaan suurempi kappa kuin mitä saataisiin täsmälleen samanlaisessa testisovelluksessa, jos esim. $(a+b)/n \approx (a+c)/n \approx 0,10$ (eli diagnosoitava asia on harvinainen) huolimatta siitä havaitut yhtäpitävyydet p_o olisivat kummassakin tutkimuksessa täsmälleen samat (ks. esim. Feinstein & Cicchetti 1990).

Hypoteesin H_0 : “ei ole yhtäpitävyyttä ollenkaan” testaaminen voidaan suorittaa z-testillä, mutta yleensä se ei ole kovin mielekästä, koska useimmiten ollaan kiinnostuneita kaplan suuruudesta, eikä siitä onko se yhtäpitävyys tilastollisesti merkitsevä. Luottamusväli kannattaa aina ilmoittaa, koska siitä ilmenee kuinka luotettava kapalle laskettu arvio on.

Yleistykset

Kappa voidaan yleistää useisiin eri tilanteisiin, esim. (rxr)-taulukolle, $r > 2$ ja myös tilanteisiin, missä arvioitsijoita on enemmän kuin kaksi (Fleiss 1971 ja Light 1971). Yleistetyssä kapassa on sama ongelma kuin (2x2)-taukolle lasketussa kapassakin, ts. kaplan suuruus vaihtelee huomattavasti riippuen eri tutkimuksissa todetuista reunafrekvenssijakaumista vaikka havaitut yhtäpitävyydet, testausmenetelmät ja -olosuhteet ja olisivat näissä tutkimuksissa täsmälleen samat.

Cohenin kappa ei riipu ollenkaan taulukon ei-diagonaalelementeistä vaan määräytyy pelkkien päädiagonaalelementtien perusteella. Toisinaan on kuitenkin mielekästä antaa erilaiset painoarvot yhtäpitämättömyydelle, eli painottaa ei-diagonaalelementtejä eri painoilla. Agrestin (1990) on ehdottanut seuraavanlaisia painokertoimia:

$$w_{ij} = 1 - \frac{(i - j)^2}{(r - 1)^2}$$

Näille painoille on ominaista, että yhtäpitävyys on suurempi päädiagonaalia lähempänä olevilla kuin siitä kauempana olevilla lokeroilla. StatXact 8 laskee myös painotetun kaplan. Kahdesta eri taulukosta laskettuja kappa-arvoja ei tulisi verrata keskenään, mikäli reunasummat poikkeavat toisistaan huomattavasti. On helppo konstruoida esimerkki, missä kappa-arvot ovat samat, mutta yhtäpitävyydet (= saman lopputuloksen suhteelliset osuudet) poikkeavat huomattavastikin toisistaan.

Kendallin konkordanssikerroin (W)

Kendallin konkordanssikerroin (W) "Kendall's W", "Coefficient of concordance" on yhtäpitävyyden mitta, joka on välille (0-1) skaalattu versio Friedmanin testisuureesta. Sitä voidaan käyttää järjestysasteikollisen muuttujan yhteydessä mittaamaan esim. N:n arvioitsijan suorittamien arviointien yhtäpitävyyttä. Mikäli arvioinnit poikkeavat täysin toisistaan, niin $W=0$ ja mikäli arvioinnit ovat täysin yhtenevät, niin $W=1$. Kaikkien arvioitsijoiden kesken pareittain laskettujen Spearmanin korrelaatioiden keskiarvo on lineaarisessa yhteydessä W:hen.

Käyttötilanne:

Kaikki edellä tarkastellut assosiaatio- ja yhtäpitävyydmitat soveltuvat tilanteisiin missä samoista havaintoyksiköistä on kaksi mitattua suuretta. Kendallin W soveltuu tilanteisiin, missä esim. samoista henkilöistä on suoritettu enemmän kuin kaksi toistettua mittausta, joiden yhtäpitävyyttä halutaan tutkia. Mittausten tulee olla järjestysasteikollisia.

Esim.

Halutaan tutkia useamman kuin kahden arvioitsijan suorittamia arvioita samoista henkilöistä tai näytteistä tai henkilöiden responsia eri lääkkeitöihin.

Mikäli tällaisia toisistaan riippuvia lopputuloksia on N kappaletta, niin niiden perusteella

voitaisiin laskea $\frac{N!}{2!(N-2)!}$ pareittaista riippuvuuden mittausta, esim. Spearmanin

korrelaatiokerrointa r_s . Mikäli aineistossa ei ole samoja havaintoarvoja, niin voidaan

näyttää, että: $W = \frac{(N-1)}{N} \cdot \bar{r}_s + 1$, missä \bar{r}_s on pareittaisten Spearmanin korrelaatioiden

keskiarvo, joten Kendallin W on suoraan verrannollinen tähän suureeseen \bar{r}_s . W

voidaan laskea myös Friedmanin testisuureesta T_F kaavalla $W = T_F / (N \cdot (K-1))$, missä K on esim. henkilöiden määrä.

Esim. Gastryl-esimerkkiaineisto 2.

Tablettien antotapoja oli kolme: E = tyhjään mahaan, A = antasidin kanssa ja M = aterian jälkeen. Koehenkilöitä oli kahdeksan. Testataan hypoteesia lääkkeen antotapojen välisistä eroista välittämättä tabletin päällysteestä. Tässä aineistossa $N=3$ ja $K=8$.

		Antotapa	Potilas1	Potilas2	Potilas3	Potilas4	Potilas5	Potilas6	Potilas7	Potilas8
1		E	65	75	65	104	62	68	83	93
2		A	63	110	73	103	88	81	94	113
3		M	74	99	70	118	87	78	99	98
Total	N	3	3	3	3	3	3	3	3	3

Taulukossa olevat arvot ovat lääkeainepitoisuuksia.

SPSS: Valikot ”Analyze” ► ”Nonparametric tests” ► ”K Related Samples”. Kohdassa ”Test Type” testiksi valitaan ”Friedman” ja ”Kendall W” ja kohtaan ”Test variables” vietään muuttujat **Potilas1, Potilas2, ..., Potilas8**

Tulos:

Ranks		Test Statistics ^a	
	Mean Rank		
Potilas1	1,83	N	3
Potilas2	6,17	Chi-Square	17,696
Potilas3	1,83	df	7
Potilas4	7,33	Asymp. Sig.	,013
Potilas5	3,00	a. Friedman Test	
Potilas6	3,33	Test Statistics	
Potilas7	5,83	N	3
Potilas8	6,67	Kendall's W ^a	,843
		Chi-Square	17,696
		df	7
		Asymp. Sig.	,013
		a. Kendall's Coefficient of Concordance	

Tulkinta:

Kendallin W on 0,843, joten potilaiden järjestyksen välillä on hyvä yhtäpitävyys eri antotavoissa.

Viitteet

Agresti A. Categorical data analysis. John Wiley & Sons, New York, 1990

Andersen B. Problems with p. Significance testing in medical research. Rockie A/S, Denmark 1984

Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilita. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 1936; 8: 3-62.

Brown LM, Pottorn LM, Hoover RN. Testicular cancer in young men: the search for causes of the epidemic increase in United States. J Epidemiology and Community Health. 1987; 41: 349- 54.

Campbell I. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. Statistics in Medicine 2007;26:3661-3675.

Campbell MJ, Machin D. Medical Statistics, A Commonsense Approach. John Wiley & Sons, Chichester 1999.

Cochran WG. Some methods for strengthening the common χ^2 tests. Biometrics 1954;10:417-451.

Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 1960; 20: 37-46.

Dunn OJ. Multiple comparisons using rank sums. Technometrics 1964; 6: 241-52.

Feinstein AR & Cicchetti DV. High agreement but low kappa: The problems of two paradoxes. Journal of Clinical Epidemiology 1990; 43: 543-9.

Fleiss JL, Cohen J & Everitt BS. Large sample standard errors of kappa and weighted kappa. Psychol. Bull. 1969; 72: 323-7.

Fleiss JL Measuring nominal scale agreement among many raters. Psychol. Bull. 1971; 76: 378-82.

Landis JR & Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-74.

Light RJ. Measures of agreement for qualitative data: Some generalizations and alternatives. Psychol. Bull. 1971; 76: 365 -77.

Makuch RD & Parks WP. Response of serum antigen level to AZT treatment of AIDS. AIDS Research and Human Retroviruses 1988; 4: 315-16.

Pearson ES. The choice of statistical tests illustrated on the interpretation of data classed in a 2x2 table. Biometrika 1947;34:139-167.

Rahe AJ. Tables of critical values for the Pratt matched pair signed rank statistic. Journal of American Statistical Association 1974; 69: 368-73.

Rijn JC, Klemetso N, Reitsma JB et al. Observer variation in evaluation of lumbar herniated discs and root compression: spiral CT compared with MRI. The British Journal of Radiology 2006; 79: 372-377.

Romo M ym. Ratatyökoneenkäyttäjien ja vaununtarkastajien lannerankatutkimus. VR:n tutkimusraportti 1980.

Wulff HR & Schlichting P. Statistical program for the analysis of the results of controlled therapeutic trials and other types of clinical research. Medstat version 2.12 manual 1991.

Biostatistisia ohjeita artikkelien kirjoittamisessa

Yhtenäiset yleisohjeet biolääketieteellisiin lehtiin kirjoittamisesta löytyy netistä sivulta: <http://www.icmje.org/>. Kuten johdannossa todettiin, johtavissa lääketieteen alan lehdissä on jo yli parinkymmenen vuoden ajan kiinnitetty erityisen paljon huomiota julkaistavaksi lähetettyjen käsikirjoitusten ja julkaistujen artikkeleiden tilastolliseen laatuun (esim. Altman 1982, Emerson ja Colditz 1983 sekä Gore ym.1992). Monissa tehdyissä tutkimuksissa on osoitettu, että silti noin puolessa tilastollisia menetelmiä käyttävistä artikkeleista on huomattavia puutteita tai virheitä, jotka ovat riittävän vakavia asettamaan tekijöidensä tekemät johtopäätökset kyseenalaisiksi. Hyvät viitteet näihin arviointitutkimuksiin sekä yhtenäiset ohjeet ja suositukset artikkelien kirjoittajille, editoreille ja arvioijille ovat luettavissa mm. Langin ja Secicin vuonna 1997 sekä Byrnen vuonna 1998 ilmestyneistä kirjoista. Seuraavassa on luettelonomainen esitys joistakin keskeisimmistä artikkeleiden kirjoittamiseen liittyvistä biostatistisista näkökohdista.

Johdanto

Johdanto-osan tarkoituksena on kuvata, miksi kyseinen tutkimus on tehty, ja esitellä tutkittava aihe. Johdanto-osan lopussa tulisi selkeästi kuvata tutkittavat hypoteesit.

Aineisto- ja menetelmäosa

Yleistä

Artikkelien aineisto- ja menetelmäosa tulisi kirjoittaa niin tarkasti, että periaatteessa siinä annettujen tietojen perusteella tutkimus voitaisiin toistaa, jos niin haluttaisiin. Useimmissa tutkimuksissa ainakin seuraavien asioiden tulisi käydä ilmi aineisto- ja menetelmäosasta:

- mitä tutkimusasetelmaa ("study design") on käytetty
 - riittävän tarkka kuvaus!
- kuinka koehenkilöt ja verrokkit on valittu
- sisäänotto- ja poissuljentakriteerit
- mistä perusjoukosta ("sample population") valinta on suoritettu
- koehenkilöiden ja verrokkien määrä
 - vaikuttaa tutkimuksen voimaan
 - millä perusteella tähän määrään on päädytty
- havaintojen keräämis- ja mittaustekniikat
- vastausprosentit havainnoivissa tutkimuksissa ("surveys")
- poispuodonneet ("drop-outs")
 - poisjääneiden kuvaus ja poispuodonamisen syyt

Tilastolliset menetelmät

Kaikki artikkelissa käytetyt tilastolliset menetelmät pitää mainita menetelmäosassa. Standardimenetelmistä, kuten Mann–Whitneyn U-testi tai χ^2 -testi, riittää maininta. Yleisesti ottaen kaikista tavallisimmin käytetyistä tilastopaketeista (kuten SPSS) löytyvistä tilastollisista menetelmistä riittää maininta ja viittaus käytettyyn pakettiin. Tilastopakettien mainitseminen ei ole ohjelmistotalojen mainostamista, vaan on tärkeää siinä mielessä, että eri ohjelmien tuottamat tulokset saattavat jonkin verran erota toisistaan, koska niiden käyttämät laskenta-algoritmit eivät välttämättä ole identtisiä. Yleensä nämä erot kuitenkin ovat pieniä. Ainakin seuraavat asiat tulisi käydä ilmi kohdasta ”Tilastolliset menetelmät”:

- Kaikki käytetyt menetelmät on mainittava
 - missä
 - miten
 - perustelu miksi
- Perusmenetelmistä on mainittava käytetty ohjelmistopaketti ja sen versio, esim. SPSS 17.0 tai StatXact 8.
- Erikoismenetelmistä tulee metodiosassa tai liitteessä olla tarkka kuvaus tai viite johonkin artikkeliin tai ohjelmistoon. Erityisesti monimutkaisempien menetelmien ja mallien kohdalla täytyy perustella, miksi niitä on käytetty, ja arvioitava niiden soveltuvuus.
- On todettava käytettyjen menetelmien oletusten tarkistaminen, normaalisuus, vakiovarianssisuus jne. Pitää mainita kaikki käytetyt muunnokset, esim. log-muunnos.
- Pitää ilmoittaa, kuinka poikkeavien havaintoarvojen kohdalla on menetelty.
- Puuttuvien tietojen käsittely tulee kuvata; onko esim. käytetty jotain puuttuvien tietojen korvaamismenettelyä (imputointia).
- Tulee mainita onko käytetty yksi- vai kaksisuuntaista testausta
- Tutkimuksessa käytetty tilastollinen merkitsevyystaso, esim. $P < 0,05$, tulee mainita.

Useimmissa artikkeleissa on metodiosassa tarve myös kuvata aineistokoon arviointiperiaatteet ja laskelmat. Ainakin seuraavat asiat tulisi silloin mainita:

- Muuttuja/muuttujat, jonka/joiden perusteella aineistokoon arviointi on suoritettu
- Mitä menetelmää on käytetty
- Mitä merkitsevyystasoa on käytetty ($=\alpha$ -virhe, tavallisesti 0,05)
- Mitä voimaa laskelmissa on käytetty (tavallisesti 80 %)
- Kuinka suureen vaikutuksen suuruuteen ($=$ kliinisesti merkittävä ero) voimalaskelma perustuu.

Mikäli aineistokokoa ei ole etukäteen kuvattu tai se asetelman luonteesta johtuen on ollut vaikea laskea, voi pohdintaosassa tehdä laskelmia myös jälkikäteen.

Tulososa

Aineiston kuvailu

Aineisto tulee esitellä riittävästi aina ennen tulosten esittelyä. Useimmissa artikkeleissa aineiston esittelyyn käytetään taulukkomuotoista esitystä (taulukko 1), jossa aineiston keskeisimpien tulosten kannalta tärkeimpien taustamuuttujien prosenttiosuudet, keskiluvut (kuten keskiarvot tai mediaanit) ja hajontaa kuvaavat suureet (kuten SD, vaihteluväli) esitetään mahdollisesti ryhmiteltynä jonkin tutkimuksen kannalta keskeisen ryhmittymän mukaan. Hoitokokeissa taulukossa 1 pyritään antamaan kuva, ovatko vertailtavat ryhmät taustoiltaan, eli lähtötason muuttujien suhteen vertailukelpoiset. Mikäli on kyseessä satunnaistettu koe, niin taustamuuttujien kaikki mahdollisesti erot johtuvat sattumasta. Eroilla voi kuitenkin tulosten tulkinnan kannalta olla merkitystä. Havainnoivissa tutkimuksissa vertailtavien ryhmien taustamuuttujissa todetut erot huomioidaan sekoittavina tekijöinä tuloksia analysoitaessa.

Tulosten esittäminen

Tulosten esittämisessä tulisi kiinnittää huomiota ainakin seuraaviin asioihin:

- Prosenttiluvuissa on käytävä ilmi nimittäjä; esim. väheneminen 25 % → 20 % saatetaan ilmaista joko 5 %:na tai 20 %:na, riippuen siitä, tarkoitetaanko prosenttiyksiköitä (nimittäjä 100) vai suhteellista muutosta (nimittäjä 25).
- Pienten lukumäärien yhteydessä prosentit voivat olla harhaanjohtavia, esim. jos nimittäjä on 20, niin prosenttiluvut ovat ns. modulo 5, eli hyppivät viiden välein.
- Prosenttilukujen tarkkuuteen pätee sääntö: annetaan niin monta desimaalia, että osoittajassa oleva luku on rekonstruoitavissa.
- Merkintää $\bar{x} \pm \text{SD}$ tulisi välttää. Monet lääketieteellisistä lehdistä (BMJ, Lancet, NEJM) eivät suosittele tätä merkintätapaa. Merkintä tarkoittaisi, että mikäli jakauma olisi normaalin, niin suunnilleen 68 % muuttujan x havaintoarvoista sisältyisi tähän väliin. Julkaisuissa näkee runsaasti tapauksia, missä jakauma on niin vino, että miinusmerkki ei edes voi tulla kyseeseen. Parempi merkintätapa on: \bar{x} (SD =...). SD mittaa hajontaa mielekkäästi vain symmetrisissä jakaumissa.
- Merkintä $\bar{x} \pm \text{SE}$, jossa SE on keskiarvon keskivirhe, tarkoittaa keskiarvon 68 %:n luottamusväliä. Mikäli piirroksissa halutaan ilmaista esim. arvioidun keskiarvokäyrän tarkkuutta, se usein ilmaistaan usein keskiarvopisteen ympärille molempiin suuntiin piirretyllä SE:n mittaisella janalla. Mikäli tavoitteena on kuvata havaintopisteistöön sisältyvää hajontaa, esitetään piirroksissa silloin SD.
- Merkitsevyydestien antamien tulosten esittäminen: (P=..., testi).
- Kannattaa aina harkita, onko testaaminen mielekkästä, sillä usein luottamusvälit ovat mielekkäämpiä. Erityisesti, kun ilmoitetaan esim. lääkehoidon tehoa kahden vertailtavan ryhmän välillä, niin pelkän P-arvon ilmoittaminen antaa lukijalle hyvin vähän tietoa. Ehdottomasti pitäisi ilmoittaa lasketun tehoeron ohella myös sen luottamusväli. Yleensä käytetään 95 % luottamusvälejä, vaikka se ei mihinkään matemaattiseen tosiseikkaan perustukaan.

- Merkitsevyysrajan $\alpha = 0,05$ käyttö johtaa helposti merkitsevien tulosten yliedustukseen; suoritettuja **testejä** on usein raportoituja **enemmän** ja silloin α -virhe kasvaa
- Yleensä kannattaa ilmoittaa tarkka P-arvo, esim. 0,0612 tai 0,0425. Tilastopaketit antavat P-arvon usein neljän desimaalin tarkkuudella ja esim. myöhempiä mahdollisia meta-analyysyjä silmällä pitäen näinkin suurta tarkkuutta voidaan käyttää, vaikka yleensä kaksi merkitsevää numeroa riittää.
- Merkitsevyydet, jotka ovat alle 0,0001, eivät ole mielekkäitä, esim. mikäli ilmoitetaan testin tuloksena $P = 2,6 \cdot 10^{-8}$, niin on testauksen tarkoitus ymmärretty väärin. Tämä tulos kannattaa ilmaista muodossa $P < 0,0001$.
- Jos todetaan, että esim. jokin hoitoero ei ole tilastollisesti merkitsevä, esim. $P > 0,05$, niin se ei tarkoita, että vaikutusta ei ole. Vaikutus voi olla esim. niin pieni, että se ei tule käytetyllä aineistokoolla näkyviin. Nollahypoteesia ei voi koskaan todistaa oikeaksi.
- Hyvä grafiikka on usein parempi kuin raskaat taulukot. Monet lehdet toivovat mm. hajontadiagrammeja tai laatikko-jana-tyyppisiä esitysmuotoja. Jos aineisto on pieni, kannattaa kaikki havaintopisteet esittää kuvassa.
- Kuvien ja taulukoiden tulee olla itsensä selittäviä; ts. ne pitää varustaa niin hyvin selityksin, että lukija voi ymmärtää niiden sisällön lukematta koko artikkelia. Kolmiulotteisen grafiikan käyttämisessä on oltava tarkkana, sillä se saattaa helposti olla harhaanjohtavaa. Mikäli tutkittava asia ei ole kolmiulotteinen, ei kolmiulotteista grafiikkaa pidä tieteellisissä artikkeleissa käyttää. Liian monimutkaisia taulukoita tulisi välttää, samoin kuvien ja taulukoiden yhdistelmiä.

Virheitä tulosten analysoinnissa

Tavallisimpia virheitä tulosten analysoinnissa:

- parittaisen asetelman käyttämättä jättäminen
- testien tai menetelmien soveltuvuuden tarkistaminen; ts. ovatko testien ja menetelmien käyttöedellytykset (olettamukset) voimassa
- toistomittausten tulkitseminen riippumattomiksi havainnoiksi
 - toistomittausten yhteydessä täytyy käyttää siihen tarkoitettuja testejä ja menetelmiä
- poikkeava havaintoarvojen merkityksen huomioimatta jättäminen
 - paljonko tulos muuttuu, jos havaintoarvot jätetään pois
- puuttuvien tietojen käsittelyssä on paljon virheellisiä menettelytapoja
 - eri korvausmenettelyt (esim. ohjelma SOLAS)

Kliiniset kokeet

Kliinisiin kokeiden laadun parantamiseen on viime vuosikymmeninä kiinnitetty paljon huomiota ja aiheesta on julkaistu monia ohjeistoja, esim. "CONSORT statement" (Begg ym.1996), joiden tavoitteena on paitsi laadun parantaminen, niin kliinisten kokeiden toteutuksen ja raportoinnin yhtenäistäminen. Nämä ohjeet löytyvät myös Internetistä: www.consort-statement.org/

Kliinisten kokeiden raportoinnissa tulisi kiinnittää erityistä huomiota mm. seuraaviin asioihin

- riittävän tarkka koejärjestelyjen kuvaus
- onko tutkimus avoin vai suljettu
- sokkouttamismenettelyt (yksöissokko, kaksoissokko)
- hoitojen tarkka kuvaus
- hoitomyöntyvyyden tutkiminen ja huomioonottaminen analyyseissä
- käytetty satunnaistamismenetelmä
 - lohkokoko, osittaminen
- lopputulosmuuttujan/-muuttujien kuvaus
 - käytetyt muunnokset ilmoitettava
- hoitokokeissa tulisi suositusten mukaan olla mieluiten vain yksi ensisijainen lopputulosmuuttuja – toissijaiset lopputulosmuuttajat eksploraatiivisia, tutkimushypoteeseja luovia
- mikäli on useampia ensisijaisia lopputulosmuuttujia, joudutaan helposti monivertailuongelmaan
- tutkimuksen voima lasketaan päälopputulosmuuttujan perusteella
- toissijaiset lopputulosmuuttujat ovat lähinnä eksploraatiivisia, uusia tutkimushypoteeseja luovia
- usean hoidon samanaikaisissa kokeissa kaikki ennalta suunnitellut kontrastit tulee kirjata tutkimussuunnitelmaan; esim., että hoitojen A, B ja C vertailun lisäksi hoitoja A ja B tullaan myös yhdessä vertaamaan hoitoon C.

Satunnaistaminen takaa kliinisissä kokeissa hoitojen allokoitumisen harhattomasti eri ryhmien välillä, mutta se ei takaa, että lähtötilanteessa kaikkien lopputulokseen vaikuttavien muuttujien jakaumien samankaltaisuutta etenkin, jos aineistot ovat pienet.

Altman ja Dore (1990) tutkivat 80 artikkelia satunnaistetuista kliinisistä kokeista, jotka oli julkaistu jossakin seuraavista neljästä lehdestä: Annals of International Medicine, British Medical Journal, The Lancet tai New England Journal of Medicine. Arviointiraati kiinnitti erityistä huomioita lähtötason muuttujavertailuja koskeviin virheisiin ja puutteisiin. Seuraavassa luettelossa on esitetty valikoituja **tuloksia todetuista lähtötason muuttujia koskevan käsittelyn puutteista:**

- 19 raportissa käytettiin jatkuvien muuttujien hajonnan kuvaamisessa **keskivirhettä (SE)**, yhdessä **luottamusvälejä** ja 13 artikkelissa ei ilmoitettu hajontaa ollenkaan, joten puutteita oli 49 %:ssa tapauksista.
- Lähtötason muuttujien kuvaamisessa hajonnan ilmoittaminen on oleellista. Suunnilleen symmetrisillä jakaumilla ilmoitetaan keskihajonta (SD) ja muuten esimerkiksi **vaihteluväli ("range")**, valikoituja **prosenttipisteitä ("percentiles")** tai joskus, hyvin pienissä tutkimuksissa, jopa koko aineisto. Keskivirhe (SE) ei ole hajonnan mitta lähtötason muuttujille vaan erilaisten estimaattien luotettavuuden

mitta. Myöskään luottamusvälien käyttö ei ole mielekästä kun kuvataan lähtötason eroja. Luottamusvälejä käytetään hoitoeftien tai niiden erojen arviointiin.

- Hypoteesien testaamista lähtötason muuttujien suhteen käytettiin 58 %:ssa tutkimuksista, mutta vain joka kolmannessa artikkelissa ilmoitettiin käytetty menetelmä. Testausmenetelmän ilmoittaminen on aivan oleellista ilmoitetun **P-arvon** kannalta.
- 46 artikkelissa ilmoitettiin yhteensä noin kuusisataa hypoteesien testausta lähtötason muuttujien suhteen, ja niistä 4 % antoi tilastollisesti merkitsevän tuloksen ($P < 0,05$).

Yleensä tutkimuksissa testataan lukuisia lähtötason muuttujia ja ennusteen kannalta relevanteimmat ilmoitetaan. Suoritettujen testien määrä paisuu kuitenkin tavallisesti suureksi ja silloin törmätään **monivertailuongelmaan** (“**multiple testing**”). Esimerkiksi, jos testataan 20 lähtötason muuttujaa, voidaan olettaa, että ainakin yhdessä niistä saadaan tilastollisesti merkitsevä ero ($P < 0,05$), vaikka eroa ei todellisuudessa olisi olemassakaan. Tässä arvioissa oletetaan lähtötason vertailut toisistaan riippumattomiksi, mitä ne eivät käytännössä tietenkään ole.

- **Hypoteesien testaaminen** ei ole paras mahdollinen ryhmien vertailukelpoisuuden arviointikeino. Paljon tärkeämpää olisi tietää prognostisten tekijöiden voimakkuus ja ryhmien välinen mahdollinen epätasapaino.
- Noin puolessa tutkimuksista lähtötason eroja ei ollut korjattu tai huomioitu mitenkään tuloksia analysoitaessa. Joka neljännessä **korjaamiseen** (“adjustment”) käytettiin tilastollista mallintamista.
- Lähtötason muuttujien vaikutuksen korjaaminen mallittamalla oli vielä 1980-luvun alussa jonkin verran kiistelty kysymys. Nykyisin adjustointiin suhtaudutaan kuitenkin melko yksimielisen hyväksyvästi. **Logistinen malli** ja **Coxin malli** ovat tavallisimpia keinoja adjustoinnin suorittamiseksi.
- Lähes kaikissa (91 %) artikkeleista kommentoitiin jollakin tavoin lähtötasojen samanlaisuutta. Vertailu oli kuitenkin asiallista vain 59 %:ssa tapauksista. Suurimpina puutteina olivat informoinnin riittämättömyys ja virheet adjustoinnissa

Laboratoriotutkimukset

Useimmat CONSORT-ohjeista pätevät myös laboratoriotutkimuksiin. Kliinisissä kokeissa tarkasteltu ihminen korvataan vain esim. sanalla koe-eläin tai näyte. Satunnaistaminen ei luonnollisestikaan useimmiten ole laboratoriotutkimuksissa mahdollista. Siitä huolimatta myös niissä systemaattisen virheen mahdollisuuteen tulee kiinnittää erityistä huomiota, samoin mittalaitteiden virhearvioihin sekä mittauksen ja kokeiden toistettavuuteen. Diagnostisen tarkkuuden ja diagnostisten suureiden yhtäpitävyyden arvioinnista löytyy ohjeita mm. verkkosivulta: www.consort-statement.org/stardstatement.htm.

Meta-analyysi, systemaattinen katsaus

Kliinisten tutkimusten perusteella tehdään nykyisin hyvin runsaasti systemaattisia katsauksia ja meta-analyysyjä. Kyse on osittain muoti-ilmiöstä, osittain siitä, että niiden tekeminen on viime aikoina huomattavasti helpottunut kattavien tietokantojen, esim. Cochrane-tietokanta (www.cochrane.org/index0.htm), ansiosta. Muun muassa lääkeyritykset ovat velvollisia antamaan tietoja myös sellaisista tutkimuksista, jotka ovat epäonnistuneet, jotka on keskeytetty tai joista ei ole tehty tieteellistä julkaisua. Aiemmin näitä tutkimuksia jäi runsaasti ns. pöytälaatikkoon. Meta-analyyseista on olemassa useita standardeja, esim. **MOOSE** (Meta-analysis Of Observational Studies in Epidemiology) ja **QUOROM** (Quality Of Reporting Of Meta-analysis). Nämä ohjeet ovat elektronisessa muodossa verkkosivulla: www.consort-statement.org/Initiatives/complements.htm Chalmersin ja Altmanin kirjassa on myös paljon hyviä ohjeita systemaattisten katsausten tekoon.

Havainnoivat tutkimukset

Havainnoivien tutkimusten kohdalla ei tavallisesti noudateta läheskään yhtä tarkkoja standardeja kuin kliinisissä kokeissa, vaikka oikeastaan pitäisi. Tästä syystä ne ovat erityisen herkkiä mm. erilaisten sekoittavien tekijöiden vaikutukselle. Lukijan on usein vaikea saada havainnoivien tutkimusten artikkeleista tarkkaa käsitystä siitä, kuinka tekijät ovat aineistonsa keränneet, onko tehty minkäänlaista aineistokoon arviointia ja kuinka sekoittavia tekijöitä on kontrolloitu. Ainakin seuraavien asioiden tulisi havainnoivien tutkimusten artikkeleista käydä selville:

- Käytetyn tutkimusasetelman tulisi ilmetä jo abstraktissa
- Tutkimushypoteesit eli miksi työ on tehty. Ne mainitaan yleensä johdanto-osan lopussa
- Aineisto-osassa kuvataan tutkittavat ryhmät, mahdollisesti käytetyt voimalaskelmat, osallistumisprosentit ja kadon arviointi. Katoanalyysien teko on havainnoivissa tutkimuksissa erittäin tärkeää erilaisten selektiomekanismien selvittämiseksi.
- Käytetyt tilastolliset menetelmät ilmoitetaan edellä esitettyjen yleisohjeiden mukaisesti.
- Taulukossa 1 yleensä kuvataan tutkittava aineisto keskeisimpien ryhmitystekijöiden mukaisesti.
- Tärkeimpiin lopputuloksiin kannattaa aina laskea luottamusvälit. Niiden perusteella lukija voi muodostaa käsityksen tulosten luotettavuudesta.
- Pohdintaosassa tulisi kiinnittää huomiota paitsi keskeisimpien tulosten luotettavuuteen ja myös siihen, kuinka käytetyt tilastolliset menetelmät soveltuvat kyseiseen tutkimukseen.

Viitteet

www.icmje.org/

www.consort-statement.org/

www.consort-statement.org/stardstatement.htm

www.cochrane.org/index0.htm

www.consort-statement.org/Initiatives/complements.htm

www.kttl.helsinki.fi/sarna

Altman DG. Statistics in medical journals. *Statistics in Medicine* 1982;1:59-71.

Altman DG, Dore CJ. Randomization and baseline comparisons in clinical trials. *Lancet* 1990;335:149-53.

Begg C, Cho M, Eastwood S, ym. Improving the quality of reporting of randomized controlled trials: the **CONSORT** statement. *JAMA* 1996;276:637- 9.

Byrne DW. *Publishing Your Medical Research Paper: What They Don't Teach In Medical School*. Lippincott Williams & Wilkins, Philadelphia PA 1998, ISBN: 0-683-30074-1.

Chalmers I, Altman DG. *Systematic reviews*, BMJ Publishing Group 1995.

Emerson JD, Colditz GA. Use of statistical analysis in the *New England Journal of Medicine*. *NEJM*.1983; 309,709-713.

Gore S, Jones G, Thompson SG. *Lancet's statistical review process: areas for improvements by authors*. *Lancet* 1992;340:100-102.

Lang TA, Secic M. *How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors and Reviewers*. American College of Physicians, Philadelphia PA 1997. ISBN 0-943126-44-4.

Huom.

Tämä moniste on elektronisessa muodossa osoitteessa:

www.mv.helsinki.fi/home/sarna/opetus nimellä **moniste osa1**. Siihen liittyvä sanasto on samalla sivulla.