

Riston osuus (kysy torstai 15min)

Käsitteitä

- Luottamusväli
- Standardointi (tutustutaan Karma & Komulainen aineistoon)

Parametrittomat testit

- Parametrian edellytysten tarkistaminen
- Khiin neliön testi
- Mann-Whitney U-testi
- Wilcoxon
- Kruskal-Wallis

Faktorianalyysi (...regressioanalyysi)

- Käyttö (eksploraatiivinen / konfirmatorinen)

1

Tilastollisesta päätöksenteon perusteita

Karma-Komulainen s. 65 ->

Yleistäminen (kuinka hyvin otos "sample" vastaa perusjoukkoa "population")

Luottamusväli

- Perusjoukon keskiarvoa (tai vastaavaa tunnuslukua) ei voida otoksen perusteella määrittää tarkasti. Otoksen perusteella voidaan vain sanoa, millä todennäköisyydellä arvo sijoittuu tietyille väleille. Yleensä lasketaan vähintään 95% todennäköisyys (vrt. 99 ja 99,9%).

- 10 oppilaan perusteella saatu keskiarvo on toki sattumalle alttiimpi (eli suurempi luottamusväli) kuin 100:n tai 1000:n oppilaan vastaava....

2

Luottamusväli

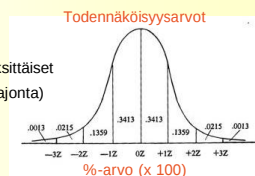
Normaalijakaumaa hyödyntävä päättely

Esim. 1. summamuuttujat muistuttavat normaalijakaumaa (muista muuttujan testaus), ja (äärimmäinen esimerkki: esim. 2. useista kyllä – ei – vaihtoehtoa hyödyntävistä kyselyistä voidaan laskea keskiarvo, joka alkaa muistuttaa normaalijakaumaa...)

Teoreettisiin todennäköisyyksien arviointiin käytetään Gaussin käyrää, jonka yhteensopivuutta/vastaavuutta verrataan käytännön tutkimustilanteessa saatuun otantajakaumaan (joka suhteutetaan aineiston kokoon / hajontaan)

Kaksi sovellustilannetta:

- Otoksen jakauma: Miten otoksen yksittäiset pistemäärät jakautuvat (keskiarvo ja hajonta)
- Otantajakauma: Otoksista tuotettu keskilukujen jakauma (%)



Todennäköisyyksiarvot

%-arvo (x 100)

Luottamusvälin laskeminen

Keskiarvon luottamusväli. Tarvitaan tieto:

$$S = \frac{s}{\sqrt{n}}$$

a) Keskiarvo (standard error) = muuttujan keskihajonta / sqrt(n)

b) 95%:n luottamusväli = keskiarvo +/- 1.96*keskiarvo

Esim. kun testin N=250, ka=50, hajonta 8,

a) → 8 / neliöjuuri 250:sta = 0.51

b) → 50 +/- (1.96 * 0.51) → 1) = 50 +/- 1 (eli 49 – 51)

(Voitaisiin myös laskea 99%:n tai 99,9%:n luottamusväli!) vrt. s. 80-81

Luottamusvälin laskeminen prosenteista:

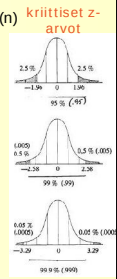
esim. natojäsenyyden kannatukselle kun 45% äänestänyt puolesta ja 55% vastaan (N=1000)

→ Keskiarvo = sqrt(.45*.55/1000) = 1.57

→ 95%:n luottamusväli kannatukselle: p+/- 1.96* keskiarvo

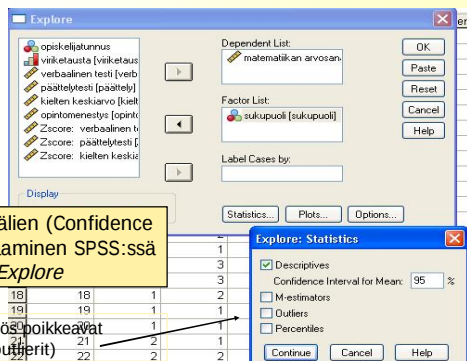
eli 45 +/- (1.96 * 1.57) → 3,01, väli on siis 41.9%-48.1%

→ tarkoittaa samaa kuin virhemarginaali (vrt. vaaligallupit)



4

Luottamusväli SPSS:ssä



Luottamusvälin (Confidence intervals) saaminen SPSS:ssä
Analyze -> Explore

(tästä saisi myös poikkeavat havainnot eli outlierit)

5

Standardointi tekee eri asteikollisista muuttujista vertailukelpoisia

- Standardointi (muuttujien normeerus): eli muuttujan arvosta vähennetään sen keskiarvo ja jaetaan hajonnalla

→ arvot verrattavissa minkä muun normeeratun muuttujan kanssa

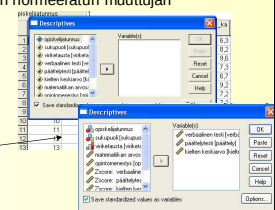
→ puhutaan Z-pisteistä

→ voidaan muodostaa esim. summamuuttujia

→ miten z-pisteet tehdään

Analyze -> Descriptive statistics

→ Descriptives



6



Mitta-asteikon vaikutus analyysimenetelmän ja merkitsevyystestin valintaan

- Peruskysymys: onko asteikko luonteeltaan luokitteleva vai jatkuva?
- Keskiarvon laskeminen joko on mahdollista (ikä, kuukausitulot, pituus, iloisuus jne.) tai sitten ei (siviilisääty, sukupuoli, erityisopetus/yleisopetus... jne.)
- Laatueroasteikollisten muuttujien lisäksi luokitteleviksi luetaan järjestysasteikolliset muuttujat, joilla kylläkin on suunta, mutta luokkaväleistä ei vallitse yksimielisyyttä (esim. ammatit)
- Luokitteleville muuttujille soveltuvat analyysit ja tilastolliset testit pohjautuvat joko ristiintaulukoon tai non-parametrisiin menetelmiin
- Jatkuville muuttujille (välimatka- ja suhdeluku-asteikollisille muuttujille, joissa mitattu ominaisuus kasvaa tasaisesti muuttuja-arvojen kasvaessa) ovat luovallisia kutakuinkin kaikki keskiarvoon perustuvat monimuuttujaiset korrelaatio- ja varianssipohjaiset menetelmät

7

PARAMETRISET JA EI-PARAMETRISET TESTIT

Tilastolliset testien jako sen mukaan minkälaisia jakaumia testit käyttävät.

Parametrisilla testeillä on jakaumaoletuksia.

- vähintään välimatka-asteikko
- jakaumien normaliteetti

Ei-parametrisillä testeillä ei ole jakaumaoletuksia.

- testit eivät ole yhtä voimakkaita kuin parametriset testit, jonka vuoksi kannattaa käyttää parametrisiä aina kuin siihen on mahdollisuus

Parametrinen

Pearsonin tulomomenttikorrelaatio
järjestyskorrelaatio

Ei-parametrinen

vs. Spearmanin

Riippumattomien otosten t-testi

vs. Mann-Whitney U-testi

Yksisuuntainen varianssi analyysi

vs. Kruskal-Wallis -testi

Toistomittaus

vs. Wilcoxonin testi

8



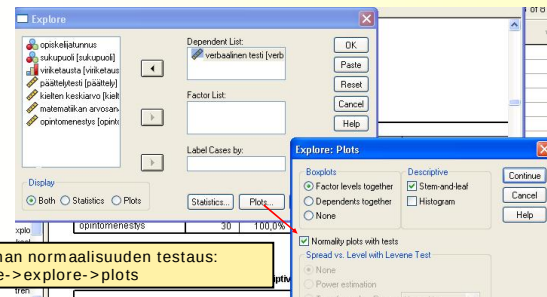
Tilastollisen menetelmän valinta

	Selitettävä -, riippuva -, y-muuttuja	
	Luokittelu- tai järjestysasteikko	Välimatka tai suhdeasteikko
Selittävä -, riippumaton -, X -muuttuja	Luokittelu- tai järjestysasteikko	Ristiintaulukointi, T-testi, Logineaariset mallit
	Välimatka tai suhdeasteikko	Logistinen ja multinomiaalinen regressioanalyysi, polku- ja rakenneyhtälömallit

9



Aineiston parametrisuuden testaaminen



Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk
	Statistic	df	Sig.	
verbaalinen testi	,149	30	,088	,962

a. Lilliefors Significance Correction

Jos sig. arvo on < .05, jakauma poikkeaa tilastollisesti merkitsevästi normaalista

Jakauman normalisuuden testaus = onko muuttuja normaalisti jakautunut.

HUOM! Graafinen tarkastelu myös!!

Vielä vinoudesta: jos vinousarvot pienempiä kuin .50 tai suurempia kuin .50 -> kannattaa testata muuttujan normalisuutta. Jos *kovin vino* -> valitaan ei-param. menetelmät jotka eivät herkkiä jakauman muodolle (tarkin kriteeri: poikkeama symmetriasta, jos vinousarvo > kuin 2 x keskivirheensä)

11



Ristiintaulukointi (eli kontingenssitalukointi)

Käyttö:

- Aineiston kuvaaminen
- Kahden luokittelevan muuttujan välisen yhteyden selvittäminen
- Muuttujien jakaumien vertaaminen eri ryhmissä
- Soveltuu kategorisille (laatuero- tai järjestysasteikko) muuttujille ja luokitelluille numeerisille (välimatka- tai suhdeasteikko) muuttujille

Analyysin tarkoitus:

Riippuvuuden tilastollisen merkitsevyyden testaaminen

- X²-testi (Khiin neliön testi, riippumattomuustesti)

Hypoteesit ja niiden testaus:

- H₀= muuttujat ovat riippumattomia
- H₁= muuttujat eivät ole riippumattomia
- Vapausaste (df) lasketaan taulukon rivien ja sarakkeiden lukumäärästä, df=(r-1)(s-1)

12

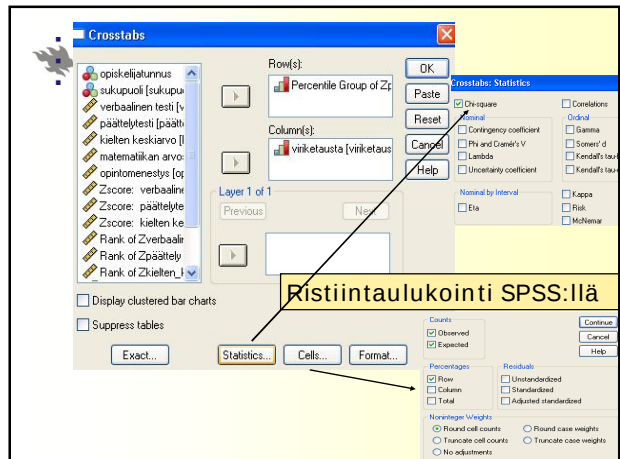


- Testin havaittu merkitsevyystaso eli p-arvo riippuu testisuureen arvosta ja vapausasteesta
- P-arvo kertoo erehtymisriskin suuruuden, kun testattava nollahypoteesi hylätään (eli mikä on riski sille, että riippuvuus johtuu sattumasta)
- Yleensä hylätään nollahypoteesi (eli todetaan, ettei aineisto tue nollahypoteesia), jos erehtymisriski on korkeintaan 5% eli p-arvo $< 0,05$)

Käytön edellytykset:

- enintään (max.) 20% odotetuista frekvensseistä saa olla pienempiä kuin 5, ja
- pienin odotettu frekvenssi = 1, eli ei saa olla odotusarvoltaan tyhjiä soluja tai odotetulta frekvenssiltään tyhjiä soluja saa olla siellä täällä, ei kuitenkaan kokonaisia nollarivejä tai -sarakkeita

13

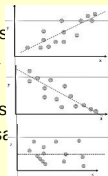


Ristiintaulukointi SPSS:llä



Korrelaatio (kertausta)

- Kertoo kahden muuttujan välisestä lineaarisesta yhteydestä eli kuinka hyvin toisen muuttujan avulla voidaan ennustaa toisen muuttujan vaihtelua
 - Ei paljasta käyräviivaista yhteyttä (vrt. ristiintaulukointi)
 - Ei selitä syy-seuraus-suhdetta.
- Päätelmä:** Korrelaatiokerroin on lineaarisen riippuvuuden voimakkuutta kuvaava tunnusluku (voi saada arvoja, jotka ovat -1:n ja 1:n välillä)
- Jos kerroin on lähellä arvoa +1, muuttujien välillä on voimakas positiivinen korrelaatio → toisen muuttujan arvojen kasvaessa myös toisen muuttujan arvot kasvavat
 - Jos kerroin on lähellä arvoa -1, muuttujien välillä on voimakas negatiivinen korrelaatio → toisen muuttujan arvojen kasvaessa toisen muuttujan arvot pienenevät
 - Jos kerroin on lähellä arvoa 0, ei muuttujien välillä ole lineaarista



15



Korrelaatio ei-parametrisillä menetelmillä

Spearman

- mittaa kahden muuttujan järjestyslukujen yhteyttä
- rank correlation, r_s

Pearson (kertausta)

- mittaa kahden satunnaismuuttujan x ja y välisen yhteyden voimakkuutta
- satunnaismuuttujien oltava vähintään intervaleilleasteikolliset ja normaalijakautuneet

HUOM:

- Selitysaste saadaan korottamalla korrelaatiokerroin (.57) potenssiin 2.
- R^2 eli korrelaatiokertoimen neliö (jos $r^2 = 0.32$), niin selittäväämuuttuja selittää 32% selitettävän muuttujan varianssista

16