



Standardointi tekee eri asteikollisista muuttujista vertailukelpoisia

- Standardointi (normeeraus): 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

1



Mitta-asteikon vaikutus analyysimenetelmän ja merkitsevyydestin valintaan

- Peruskysymys: onko asteikko luonteeltaan luokitteleva vai jatkuva?
- Keskiarvon laskeminen joko on mahdollista (ikä, kuukausitulot jne.) tai sitten ei (siviilisäätö, sukupuoli jne.)
- Laatuasteikollisten 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 ristiintaulukkoon tai non-parametriin menetelmiin

2



Tilastollisen menetelmän valinta

		Selitettävä -, riippuva -, y-muuttuja	
		Luokitteleva tai järjestysasteikko	Välimatka tai suhteasteikko
Selittävä -, riippumaton -, X -muuttuja	Luokitteleva tai järjestysasteikko	Ristiintaulukointi, Loglineaariset mallit	T-testi, Manova, Anova
	Välimatka tai suhteasteikko	Logistinen ja multinomiaalinen regressioanalyysi	Regressioanalyysi, polku- ja rakenneyhtälömallit

3



- Jatkuville muuttujille (välimatka- ja suhteasteikollisille muuttujille, joissa mitattu ominaisuus kasvaa tasaisesti muuttujien arvojen kasvaessa) ovat luovallisia kutakuinkin kaikki keskiarvoon perustuvat monimuuttujaiset korrelaatio- ja varianssipohjaiset menetelmät

Ristiintaulukointi eli kontingenssitalukointi

Käyttö:

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

4



Ristiintaulukointi

Päätely:

- Jakaumien silmämääräinen ja numeerinen tarkastelu
- Miten havainnot sirontuvat ja painottuvat soluihin, missä soluissa suurimmat frekvenssit?
- Rivi- ja sarakeprosentit (ks. spss)
- Vertailu odotusfrekvensseihin

Todellinen aineisto:
VSM-taso x 8. lk
2. aste
koulutustaso
KOE = 2 tai useampi virhe
KONTROLLI = 0-1
virhetta

	1=lukio	2=amat tikoulu	3=en tiedä	YHT.
Koe	1	4	4	9
Kontrolli	7	1	1	9
YHT.	8	5	5	18

5



- 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, $df = (r-1)(s-1)$
- Testin havaittu merkitsevyytaso 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)

6

Parametriset ja Ei-parametriset testit
 Tilastolliset testit voidaan jakaa parametrisiin ja ei-parametrisiin testeihin sen mukaan, minkälaisia jakaumia testit käyttävät.

Parametrisilla testeillä on jakaumaoletuksia.
 - vähintään välimatka-asteikko
 - jakaumien normaaliuus

Ei-parametrisillä testeillä ei ole jakaumaoletuksia.
 - testit eivät kuitenkaan yhtä voimakkaita kuin parametriset testit, jonka vuoksi kannattaa käyttää parametrisiä aina kuin siihen on mahdollisuus

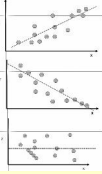
Parametrinen	Ei-parametrinen
Pearsonin tulomomenttikorrelaatio	vs. Spearmanin järjestyskorrelaatio
Riippumattomien otosten t-testi	vs. Mann-Whitney U-testi
Yksisuuntainen varianssi analyysi	vs. Kruskal-Wallis -testi
Toistomittaus	vs. Wilcoxonin testi

Korrelaatio

- 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.

Päätely: 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 riippuvuutta



Korrelaatio

- Selitysaste saadaan korottamalla korrelaatiokerroin (.57) potenssiin 2.
- R² eli korrelaatiokertoimen neliö (jos $r^2 = 0.32$), niin selittävämuuttuja selittää 32% selitettävän muuttujan varianssista

Spearman
 - mittaa kahden muuttujan järjestyslukujen yhteyttä
 - rank correlation, rS

Pearson
 - mittaa kahden satunnaismuuttujan x ja y välisen yhteyden voimakkuutta
 - satunnaismuuttujien oltava vähintään intervalli-asteikolliset ja normaali-jakautuneet

T-testi

- Ryhmien keskiarvotestaus

Kun vertaillaan kahta täysin toisistaan riippumatonta osajoukkoa toisiinsa ja halutaan tietää eroavatko ryhmien keskiarvot toisistaan, voidaan tilastollisena menetelmänä käyttää riippumattomien otosten t -testiä.

Testin oletuksia ovat muuttujan normaalijakautuneisuus ja riippumattomat otokset.

Testisuureen valintaan vaikuttaa ovatko varianssit (keskihajonnat) molemmissa osajoukoissa yhtäsuuret.