

Eksamen ved

Københavns Universitet i

Kvantitative forskningsmetoder

Det Samfundsvidenskabelige Fakultet

14. december 2011

Eksamensnummer: 5

1)

Af boxplottet kan man aflæse, at medianen er 9 og at de 50 % midterste observationer falder mellem 6 og 12, dette er interquartilen, de to streger viser, at den laveste alder er 1 år og den højeste alder er 19 år.

2)

p-p-plot står for probability-probability plot, grafen plotter den kumulative sandsynlighed for en given variabel mod kumulativ sandsynlighed for en bestemt fordeling. Således kan p-p plot vise, hvorvidt data er normalfordelt. Da "locus of Control" ligger sig tæt op af den diagonale linje, kan det konkluderes, at data er normalfordelt.

3)

Først omregnes scoresne til z-scores, med gennemsnit på 0 og standardafvigelse på 1.

$$Z=(35-50)/10 = -1,5$$

$$Z=(65-50)/10 = 1,5$$

I Appendix A1 s.800 i Field (2009) aflæses det at sandsynligheden for at nogen får over 1,5 er 0,00681. Derfor vil  $100 - 0,06681*100*2 = 86,638$  % af populationen have en score mellem 35 og 65.

4)

"Standard error of the mean" for de 16 skolebørn er  $\frac{8}{\sqrt{16}} = 2$

5)

log transformation kan anvendes, når der er en positiv skævhed i fordelingen, således at forudsætningen om normalfordelingen alligevel opfyldes og man således kan anvende de parametriske test.

En transformation af data kan dog skabe problemer senere i processen især, hvis man skal sammenligne med andre data. Et alternativ til at transformere er at anvende bootstrapping, her

antages det, at den skæve fordeling, der observeres i stikprøven, faktisk afspejler en skæv fordeling i populationen.

### 6)

Hvis IQ og uddannelsesniveau korrelerer med 0,5, kan man ved at tage denne værdi i anden udregne  $R^2$ , som kan fortælle, hvor meget varians, der kan forklares af modellen.

$R^2 = 0,25$  betyder, at 25 % af variationen i IQ kan forklares af uddannelsesniveau.

### 7)

Kendall's Tau er en ikke-parametrisk test af korrelations koefficienten, som kan anvendes, hvis forudsætningerne for Pearson's r ikke er opfyldt. Kendall's Tau foretrækkes fremfor Spearman's korrelations koefficient, da Kendall's Tau giver et bedre estimat af korrelationen i populationen.

### 8)

"Partial correlation" er en korrelation mellem to variabler i hvilken effekten af andre variabler holdes konstant. Her kontrolleres der for en eventuel tredje variabels indflydelse på begge de to originale variabler. I "Semi-partial correlation" ser man stadig på de to variable, men ser kun på effekten af den tredje variabel på én af de oprindelige variable.

### 9)

Man kan oftere finde en signifikant effekt i et repeated-measure design, fordi den usystematiske variation, dvs. den variation som ikke kontrolleres eksperimentalt vil være mindre og således er det nemmere at opdage enhver systematisk variation

### 10)

	Sums of squares	df	Mean square	F
Between groups	300	$3-1= 2$	$300/2 = 150$	$150/30 = 5$
Witin groups (error)	900	$32-2= 30$	$900/30 = 30$	
Total	1200	$33-1= 32$		

- A) = se udregninger i tabellen ovenfor
- B) F er signifikant med  $p = 0,05$ , fordi 5 er over grænseværdien på 3,32
- C) Eta-squared =  $\frac{SSM}{SST} = \frac{300}{1200} = 0,25$
- D) Med eta-squared, får vi en mål for effektstørrelse, det er egentlig det samme som  $r^2$ . Dette mål er dog en smule biased, fordi det kun baserer sig på sums of squares og ikke har nogen korrektion for, at der egentlig ønskes et estimat af effektstørrelsen i populationen og ikke blot i stikprøven. Derfor er  $\omega^2$  et bedre mål, fordi den inddrager både variansen forklaret af modellen og fejl variansen. Her ville  $\omega^2$  blive  $\frac{300-2*30}{1200+30} = 0,12$ . Effekten mindskes således, men bliver samtidig et bedre estimat for den egentlige effekt i populationen.
- E) Ved en post hoc test sammenligner man alle grupper mod hinanden, således ville man kunne se om ikke-rygere klarede sig bedre eller dårligere end henholdsvis storrygere, der lige havde røget eller storrygere der ikke havde røget de seneste 3 timer, man ville også kunne se om der var forskel mellem de to grupper af storrygere.

**11)**

En helmert contrast analyse sammenligner alle undtagen den sidste kategori med den gennemsnitlige effekt af alle efterfølgende kategorier. Tennis gruppen ville således blive sammenlignet med den gennemsnitlige effekt hos fodbold, håndbold og cykelløb grupperne sammenlagt. Fodboldgruppen ville blive sammenlignet med den gennemsnitlige effekt hos håndbold og cykelløb grupperne sammenlagt og håndboldgruppen ville blive sammenlignet med den gennemsnitlige effekt hos cykelløbgruppen.

**12).**

- a) De totale frihedsgrader er  $N-1 = 30-1 = 29$
- b) total sum of squares skal opdeles i time-point og error(time-point).  
Time-point vil have  $3-1 = 2$  frihedsgrader og error (time) vil have  $29-2=27$  frihedsgrader.

**13)**

- a) For en person med en månedlig lån på 30.000 kr. og en jobansvars rating på 15 er den forudsagte jobtilfredshed =  $10 + 0,5*30 + 3*15 = 70$

b)  $A = 0,5/0,1 = 5$

$B = 3/1,5 = 2$

c) Indkomsten er det, der bidrager mest til regressionen, idet den standardiserede koefficient er den højeste.

d) Toleransen er  $1/VIF$ , værdier under 0,2 kan være problematiske og værdier under 0,1 indikerer et seriøst problem med "multicollinarity". "Multicollinarity" er, at en "predictor" har en stærk lineær relation til de øvrige "predictorer". I opgaven er toleransen 0,8 og er derfor ikke problematisk ift. forudsætningen.

Toleransen for både salary og job responsibility er ens, fordi der kun er disse to "predictorer" og de har derfor samme indbyrdes relation.

#### 14)

I factor analyse tillader oblique rotationer, at faktorerne korrelerer med hinanden, herunder er der to former man kan vælge mellem i SPSS direct oblim, hvor en konstant delta bestemmer i hvilken grad faktorerne må korrelerer, og promax, som kan anvendes til meget store datasæt.

Oblique rotation foretrækkes ofte af psykologer, da det menes, at faktorer aldrig vil være helt uafhængige og en korrelation i mellem dem således vil afspejle virkeligheden bedre og give en højere økologisk validitet.

#### 15)

	Romantic comedy	Action thriller	Total
Males	Observed: 80 Expected: 50	Observed: 20 Expected: 50	100
Females	Observed: 20 Expected: 50	Observed: 80 Expected: 50	100
	100	100	200

a)  $\text{Pearson's } \chi^2 = (80-50)^2/50 + (20-50)^2/50 + (20-50)^2/50 + (80-50)^2/50 = 72$

b)  $Df = (2-1)*(2-1) = 1$

c)  $\chi^2$  tester forholdet mellem to frekvenser, her er den signifikant, da den er over den kritiske værdi for  $p = 0,01$  på 6,63.

- d) Det kan konkluderes, at der er en signifikant forskel på hvilken filmgenre mænd og kvinder foretrækker. For at udregne effektstørrelse ville man skulle lave en odds ratio.