**PSYC3101: Løsningsforslag til Pål og Dag Erik’s øving i estimering av reliabilitet og skalakonstruksjon**

Dette er et kort løsningsforslag til øvingen. Ideelt sett burde nok svarene vært mer utfyllende på enkelte spørsmål, men det kan dere jo selv sørge før. Merk at jeg har brukt begrepene spørsmål, indikatorer og items litt om hverandre, men de henspeiler på det samme.

**OPPGAVE 1 Bli kjent med spørsmålene:**

1. Hvilket responsformat har spørsmålene? (sett kryss der det passer)

Semantisk differensial

Likert skala

Visuell-analog skala

1. Hva menes med ”parallelle målinger”?

Litt forenklet sagt at alle spørsmålene er like gode indikatorer/ like gode til å fange opp den latente variabelen vi ønsker å måle, f eks at spørsmål 1 er like god som spørsmål 5 til å fange opp den latente variabelen vi ønsker å måle. I utgangspunktet kan vi betrakte de åtte spørsmålene som åtte repeterte mål på den latente variabelen vi ønsker å måle. Men det er ikke sikkert at alle spørsmålene er like gode til å fange opp den latente variabelen vi ønsker å måle, dette skal vi se nærmere på i øvingen.

1. Er det noen av spørsmålene ”snudd” i forhold til de andre (dvs. betyr en høy skåre på noen av spørsmålene det motsatte av hva det betyr på andre)?

Hvis ja, skriv opp hvilke (nr. på spørsmål): spm 3 og spm 7 er snudd og disse må rekodes før vi kan estimere reliabiliteten til sumskåren av alle spørsmålene – ellers blir det helt feil.

1. Hvordan kan man lage et ”samlet mål”/sumskåre for disse spørsmålene? Hva ville eventuelt en ”sumskåre” være et uttrykk for? Og – hva er maksimalt oppnåelig sumskåre en person kan ha?

Stikkord: Man kan lage et slikt samlet mål gjennom å addere sammen skåren på alle de åtte spørsmålene, etter at vi har snudd/rekodet skårene på spm3 og spm7. Sumskåren man da får kan sees på som et utrykk for den skåren man har på den latente variabelen (som vi f eks kan kalle ”preferanse for risikotaking i trafikken”). En person kan maksimalt oppnå 56 poeng på sumskåren (7\*8).

**OPPGAVE 2 BEREGNING AV SPLIT-HALF RELIABILITET**

a) Split-half reliabilitet er:

Korrelasjonen mellom to målinger på forskjellig tidspunkt av sumskåren av de åtte spørsmålene, korrigert for antall spørsmål i skalaen

Den gjennomsnittlige korrelasjonen mellom alle de åtte spørsmålene korrigert for antall spørsmål i skalaen

Korrelasjonen mellom sumskåren av to halvdeler av de åtte spørsmålene, korrigert for antall spørsmål i skalaen

d) Hvilke spørsmål inngår i de to halvdelene?

NB! Her er dette avhengig av hvilken rekkefølge dere har klikker variablene over i boksen ”Items” i SPSS. I dette eksempelet er det:

Halvdel 1: spm1, spm2, spm4, spm5

Halvdel 2:spm 6, spm 8, spm 3 (rekodet) og spm 7 (rekodet)

e) Hva blir korrelasjonen mellom de to halvdelene? Er denne det endelige målet på reliabilitet vi bruker?

Korrelasjonen er lik .672, *hvis vi du har samme spørsmål i hver halvdel som nevnt i oppgave d).* Det er ikke det endelige målet vi bruker, se oppgave f).

f) Hva er Spearman-Brown korrigering og hva blir split-half reliabiliteten til skalaen hvis vi bruker denne korreksjonen? Du kan bruke denne formelen hvis du vil regne ut dette for hånd:

, der *r* er korrelasjonen mellom de to halvdelene

Reliabiliteten etter Spearman-Brown korreksjon blir:\_\_0.804\_\_\_

Logikken bak denne korreksjonen er at når vi deler skalaen i to halvdeler og korrelerer disse med hverandre, så estimerer vi reliabiliteten til halve skalaen, dvs. en skala bestående av 4 spørsmål/items. Vi vil gi et estimat på reliabiliteten til hele skalaen på åtte spørsmål (reliabiliteten til en sumskåre bestående av åtte spørsmål). Hvis vi har åtte målinger (av den latente variabelen) isteden for fire, tror vi andelen av målefeil i sumskåren av åtte spørsmål er mindre enn andelen målefeil for en sumskåre basert på fire spørsmål. Det er derfor gjør vi denne korreksjonen. Dette forutsetter imidlertid at alle spørsmålene kan betraktes som parallelle målinger.

Altså: jo flere målinger vi har av en egenskap (latent variabel), jo mindre effekt vil tilfeldige målefeil i hver indikator ha når vi summerer alle spørsmålene til en sumskåre (gitt at spørsmålene/items er korrelerte med hverandre da).

h) Ser du noen svakheter med å beregne reliabilitet på denne måten?

Det er først og fremst fordi den beregnede reliabiliteten er sårbar for hvilke items som inngår i de to halvdelene – her har vi jo mange ulike kombinasjoner av de åtte spørsmålene vi kan lage to halvdeler på og hvis dere prøver litt på egen hånd, så vil dere finne ut at relaibilitetsestimatet vil blir litt forskjellig for hver kombinasjon.

**OPPGAVE 3 – ESTIMERING AV RELIABILITET MED CRONBACH’S ALPHA**

**a**) Hva er Cronbach’s alpha et mål på? (gi et kort svar)

Dette er et mål på reliabilitet for sumskåren til en skala bestående av spørsmål/indikatorer (items) målt på samme tidspunkt. Alfa er et mål på hvor stor andel av den totale variansen i sumskåren til denne skalaen som antas å være sann varians, dvs. det som ikke er målefeil.

Alpha er også gjennomsnittet av alle mulige estimat for split-half for skalaen, dvs. om vi hadde estimert ut split-half reliabilitet til alle mulig kombinasjoner av to halvdeler de 8 spørsmålene og beregnet gjennomsnittet av disse.

b) Start først med å beregne Cronbach’s alpha for de åtte spørsmålene ved førmålingen. Husk at du må bruke de rekodede spørsmålene!

**Tolkning av resultatene**

1. Les resultatutskriften (output) og svar på disse spørsmålene:

2. Hva er verdien på Cronbach’s alpha? \_0,79, evt. 0,781 for standardiserte items\_\_\_\_

3. Har noen av spørsmålene en gjennomsnittskåre (mean) som er langt fra det naturlige midtpunktet?

Hvilket spørsmål ligger lengst fra det naturlige midtpunktet?

Spørsmål 7 er det spørsmålet som avviker mest fra det naturlige midtpunktet (som er 4). Resten av spørsmålene har ikke så stor avvik.

4. Hva er den gjennomsnittlige (average) inter-item korrelasjonen: 0,309\_\_\_\_

5. Hva er minimum og maksimum korrelasjonen mellom de 8 spørsmålene? Min = 0,064, maks = 0,734

6. Har noen av spørsmålene en (corrected) item-total correlation under .30? \_\_spørsmål 7\_\_

7. Vil alpha-koeffisienten øke hvis du utelater noen av spørsmålene fra skalaen?

Ja, spørsmål 7 og spm 5

8. Hvis du svarte ja på forrige spørsmål – utelat det ”dårligste” itemet (spørsmål) og gjør analysen på nytt. Er det noen andre items som kan utelates nå? Hvis ja, utelat dette også og gjør analysen på nytt.

Utelater først spørsmål 7. Så viser analysen at spørsmål 5 også bør kuttes ut.

MEN: husk at det å utelate spørsmål 7 og spørsmål 5 også må begrunnes ut i fra andre hensyn. Spørsmål 7 er gjennomgående lavt korrelert med de andre spørsmålene, og det tyder på at det ikke måler det samme som de andre like godt. Videre ser vi f eks at spørsmål 7 var svært skjevfordelt og de fleste svarte likt på dette spørsmålet (dvs. lite varians). Da det ikke er så stort poeng å ta med spørsmål som de fleste svarer likt på, foreslår jeg å kutte det ut av ovennevnte hensyn. Men alfaverdien blir ikke så mye høyere av å utelate spørsmålet.

Spørsmål 5 er egentlig et dårlig formulert spørsmål, siden vi spør om to ting i samme spørsmål (dårlige veier OG gamle biler). Det kan være vanskelig å svare på slike spørsmål, og jeg ville kuttet det ut pga dette.

Men i andre tilfeller kan det forsvares å ta med spørsmål som kan gi en litt lavere, men likevel akseptabel alfa, hvis spørsmålet vil gi en ”bredere” måling av det vi ønsker å måle. Husk at det ikke er noe problem å få en høy alfaverdi hvis vi stiller omtrent det samme spørsmålet flere ganger, da vil vi jo få en veldig smal måling av det vi ønsker å måle. Tenk gjerne litt i forhold til validitet her og representativitet av items/indikatorer (se forelesningsnotater og lærebok under validitet og logisk analyse).

Kan du oppsummere hva du fant – kan du lage en skala med tilfredsstillende reliabilitet for førmålingen, og hvilke spørsmål består denne i så fall av?

Alpha: 0.817, evt. 0,819 (standardisert)

Antall spørsmål: 6 (alle unntatt spørsmål 5 og 7)

Gjennomsnittlig inter-item korrelasjon: 0.429

9. Beregn alpha med kun tre spørsmål. Velg ut spørsmål 1, 2 og 8 og beregn alpha for disse. Hva blir

Alpha: 0.741

Antall spørsmål: 3

Gjennomsnittlig inter-item korrelasjonen: 0,489

Sammenlign disse tallene med det du fant i spm 8. Hvorfor tror du alpha-verdien blir forskjellig?

Det er først og fremst fordi vi har færre spørsmål i skalaen nå, 3 sammenlignet med 8. Siden antall spørsmål/indikatorer (items) tas hensyn til i utregningen av Alpha-koeffisienten (i tillegg til den gjennomsnittlige interkorrelasjonen), vil alphaverdien lett bli lavere, selv om den gjennomsnittlige inter-korrelasjonen faktisk er sterkere i dette tilfellet hvor vi kun har 3 spørsmål

Grunnen til at antall spørsmål/items tas hensyn til, er at jo flere målinger vi har av en egenskap (latent variabel), jo mindre effekt vil tilfeldige målefeil i hver indikator ha når vi summerer alle spørsmålene til en sumskåre (gitt at spørsmålene/items er korrelerte med hverandre da).

**OPPGAVE 4 – BEREGNING AV TEST-RETEST RELIABILITET**

1. Hva er test-restest reliabilitet og hvordan kan du estimere denne i dette tilfellet?

Test-retest kalles også en stabilitetskoeffisient fordi vi har mål på de samme spørsmålene på to forskjellige tidspunkt. Test-retst reliabilitet beregnes gjennom å korrelere sumskårene av spørsmålene på de to tidspunktene med hverandre. Her trenger vi ingen Spearman-Brown korreksjon som under split-half, fordi sumskåren på hvert tidspunkt består av alle (seks) spørsmålene i skalaen

1. Korreler variabelen ”førmåling” med variabelen ”ettermåling” (velg Analyse, Correlate og Bivariate). Hva blir korrelasjonskoeffisienten, *r*, og hva kan den brukes som et mål på i dette tilfellet?

Korrelasjonskoeffisienten blir:.682

Og kan i dette tilfellet brukes som et mål på test-retest reliabiliteten til testen – eller andel sann varians i forhold til total varians i sumskåren.

1. Sammenlign korrelasjonen med Cronbach’s alpha. Er det forskjell på de to målene?

Ja, test-retest reliabiliteten er litt lavere enn alpha

1. Hva synes du er den beste måten å estimere reliabiliteten til de åtte spørsmålene på – test-retest eller Cronbach’s alpha? Gi en kort begrunnelse.

Test-retest reliabilitet er best egnet for å egenskaper som regnes å være stabile over tid (f eks personlighet, intelligens). I dette tilfellet måler en egenskap (holdninger), og det kan være litt problematisk å hevde at dette er en egenskap som er stabil over tid – særlig i dette tilfellet da over halvparten av utvalget ha vært utsatt for en holdningskampanje mellom de to målingene. Da er det nok litt tvilsomt å benytte test-retest reliabilitet. Det at det ikke er så høy korrelasjon mellom målingene gjort på to tidspunkt kan skyldes både reelle endringer i holdninger (dvs. endring i sann skåre) og tilfeldige målefeil.

Cronbach’s alpha er nok bedre i dette tilfellet, siden vi antar at holdninger kan endre seg over tid (særlig når det har vært en intervensjon i mellomtiden som i dette tilfellet). Derfor er det bedre å måle reliabilitet for holdningsmålene på hvert tidspunkt separat med alpha i dette tilfellet. Alpha er igjen å foretrekke fremfor split-half reliabilitet (se oppgave 3a).

Er vi derimot mest interessert i å undersøke stabilitet over tid for et måleinstrument, er nok test-retest bedre egnet.