

Psykologisk institutt

Eksamensoppgave i PSY2017/PSYPRO4317 Statistikk og kvantitative forskningsmetoder

Faglig kontakt under eksamen: Eva Langvik

Tlf.: Psykologisk institutt 73591960

Eksamensdato: 21.5.2013

Eksamenstid (fra-til): 09.00-13.00

**Hjelpemiddelkode/Tillatte hjelpemidler: Alle bøker og skrevne hjelpemidler er tillat,
det er også tillatt med kalkulator uten sender og mottaker**

Annen informasjon:

Målform/språk: Bokmål

Antall sider: 4

Antall sider vedlegg: 0

Kontrollert av:

ALLE OPPGAVER SKAL BESVARES

Oppgave 1

En ny doktorgrad ved NTNU viser at hvordan man opplever helsa si, er sterkt assosiert mellom ektefeller og samboere. Likhet i utdanning kan være en av mekanismene bak denne sammenhengen. Høyt utdannede har ofte bedre helse enn de med lavere utdanning. Avhandlingen antyder også at helsa vår påvirkes av partnerens utdanning. For eksempel har personer med lavere utdanning ofte bedre helse hvis de er sammen med en partner med høyere utdanning. Samtidig har partnere med ulik utdanning ganske lik oppfatning av helsa si.

Du ønsker å undersøke sammenhengen mellom partnere, helse, og utdanning i et mindre utvalg du har samlet inn ved å generere en liste over 100 tilfeldige personer i skatteliste. Av disse var det 60 som var gift eller hadde samboer, og av disse igjen var det 40 par (80 personer) som returnerte spørreskjema med svar fra begge.

De 80 respondentene oppga egen utdanning og hvor god helsa deres var. Variabelen «Helse» er selvrapportert på en kontinuerlig skala hvor maksimal skåre er 30 og lav skåre viser til dårlig helse og høy skåre indikerer god helse. Utdanningsnivå er operasjonalisert som følger: 1= Lav/ingen utdanning (Inkludert videregående skole 1-2 år), 2 = Middels utdanning (videregående skole/annen utdanning 3 eller flere år), 3=Høy utdanning= 3(Høyskole eller universitet).

- a) Er det en sammenheng mellom eget utdanningsnivå og selvrapportert helse? Formuler hypotesene. Fyll inn tallene og tekst som mangler i tabell 1 (X:1-X:5). På X:5 kan symbolene > eller < benyttes. Hva kan tabellen, i ferdig utfylt form, fortelle oss om sammenhengen mellom utdanningsnivå og selvrapportert helse? Hvilken test er benyttet i tabell 1 og hvorfor er denne testen benyttet?

Tabell 1

	Sum of Squares	df	Mean Square	X:3	Sig.
Between Groups	163.474	2	X:2	X:4	X:5
Within Groups	596.526	77	7.747		
Total	760.000	X:1			

- b) Gjør rede for innholdet i tabell 2. Beskriv kort hva de ulike kolonnene gir oss informasjon om. Hva kan tabell 2 fortelle oss om sammenhengen mellom utdanningsnivå og helse utover det vi får vite i tabell 1?

Tabell 2

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1.00	15	21.8667	2.44560	.63145	20.5123	23.2210
2.00	36	23.5000	2.88345	.48058	22.5244	24.4756
3.00	29	25.7241	2.81446	.52263	24.6536	26.7947
Total	80	24.0000	3.10165	.34678	23.3098	24.6902

Oppgave 2

I det samme forskningsprosjektet ønsker du også å studere hvor like partene i et ekteskap/samboerforhold er når det gjelder selvrapportert helse.

- a) Er det en sammenheng mellom kvinnen og mannens helse i dette utvalget? Formuler hypotesene og regn ut. Bruk signifikansnivå.01.

Benytt deg av verdiene i tabellen under for å regne ut.

	Kvinner «Helse» (n 40)	Mann «Helse» (n 40)	Mann*Kvinne
Gjennomsnitt	23.95	24.05	
Standardavvik	2.81	3.40	
Σ «Helse»	958	962	
Σ «Helse» ²	23252	23588	
Σ Kvinne «Helse»*Mann «Helse»			23324

- b) Av de to variablene "Utdanning" og "Partners/ektefelles helse", hva forklarer mest av variasjonen i selvrapportert helse? Undersøk om det er sterkest sammenheng mellom utdanning og helse (1 a) eller mellom helsen til partene i et ekteskap/samboerforhold (2a)? Vis utregninger og forklar hva resultatet betyr.
- c) Har vi mest statistisk styrke (Power) i 1 a eller 2 a? Bruk resultatene i oppgave 2 b og informasjon gitt i oppgavesettet, og begrunn svaret. Ingen utregninger nødvendig.
- d) For hele populasjonen er gjennomsnittet (μ) på skalaen «Helse» 23.02 og standardavviket i populasjonen (σ) er 3. Er mennene i dette utvalget signifikant forskjellig fra populasjonsgjennomsnittet? Formuler hypotesene og regn ut. Benytt signifikansnivå .05. Forklar hva resultatet betyr.

Oppgave 3

I utvalget så har 12 menn og 17 kvinner universitets eller høyskoleutdannelse. Er det en signifikant overrepresentasjon av kvinner i denne gruppen?

Formuler hypoteser og regn ut, benytt signifikansnivå .05

Oppgave 4

- a) Gjør rede for de ulike typene t-tester. Hva er forutsetningene for å benytte de ulike t-testene og hva gir de svar på?
- b) Beskriv en forskningsdesign hvor det passer å benytte en av t-testene. Gjør rede for hovedelementene i designet. Formuler hypotesene og gjør rede for variablene som inngår. (ingen utregning).
- c) Hvilke egenskaper ved forskningsdesignet du har beskrevet i 4 b er avgjørende for gyldigheten til resultatet av t-testen?

Oppgave 5

Gjør rede for hovedforskjellen mellom regresjon og korrelasjon.

Hva er forutsetningene for å kunne gjøre en regresjonsanalyse?

Hvilken funksjon har parameterestimatene a og b i en regresjonsligning?

Oppgave 1

En ny doktorgrad ved NTNU viser at hvordan man opplever helsa si, er sterkt assosiert mellom ektefeller og samboere. Likhet i utdanning kan være en av mekanismene bak denne sammenhengen. Høyt utdannede har ofte bedre helse enn de med lavere utdanning. Avhandlingen antyder også at helsa vår påvirkes av partnerens utdanning. For eksempel har personer med lavere utdanning ofte bedre helse hvis de er sammen med en partner med høyere utdanning. Samtidig har partnere med ulik utdanning har ganske lik oppfatning av helsa si.

Du ønsker å undersøke sammenhengen mellom partnere, helse, og utdanning i et mindre utvalg du har samlet inn ved å generere en liste over 100 tilfeldige personer i skattelistene. Av disse var det 60 som var gift eller hadde samboer, og av disse igjen var det 40 par som returnerte spørreskjema med svar fra begge.

Variabelen «Helse» er selvrapportert på en kontinuerlig skala hvor maksimal skåre er 30 og indikerer veldig bra helse. Utdanningsnivå er operasjonalisert som følger: 1= lav/ingen utdanning (Inkludert videregående skole 1-2 år), 2 = Middels utdanning (videregående skole/annen utdanning 3 eller flere år), 3=Høy utdanning= 3(Høyskole eller universitet).

- a) Er det en sammenheng mellom utdanningsnivå og selvrapportert helse? Formuler hypotesene. Fyll inn tallene og tekst som mangler i tabell 1 (X:1-X:5). På X:5 kan symbolene > og < benyttes. Hva kan tabellen, i ferdig utfylt form, fortelle oss om sammenhengen mellom utdanningsnivå og selvrapportert helse? Hvilken test er benyttet i tabell 1 og hvorfor er denne testen benyttet? (Svar: her er det blitt gjort en enveis ANOVA- analyse av variasjon, -grupperingsvariabelen har mer enn to nivå, passer best, avh.var er kontinuerlig, flere enn to t-tester øker sjansen for type 1 feil, mer enn 5 prosent sjanse for tilfeldighet. Tabellen viser at det er en signifikant forskjell mellom minst to av gjennomsnittene- ikke noe om HVEM som er forskjellig. Her må studentene skjønne at de selv kan velge signifikansnivå, -bør helst få frem at det er sig på .01 nivå, da SPSS output gir eksakt p verdi. Tabellen forteller oss at variasjonen mellom gruppene er større enn den innad i gruppene, men ikke noe mer om hvor forskjellene ligger.
- b) Gjør rede for innholdet i tabell 2. Beskriv kort hva de ulike kolonnene gir oss informasjon om. Hva kan tabell 2 fortelle oss om sammenhengen mellom utdanningsnivå og helse utover det vi får vite i tabell 1? Her er det KI til gjennomsnitt. En litt vanskelig oppgave, da de ikke er vant med å se slike output. Men, de vet at KI for et gjennomsnitt som ikke overlapper med KI for et annet vil være signifikant forskjellig fra hverandre. Her ser man hvor forskjellene ligger, -ser noe om usikkerheten til estimatet; stor KI= mer usikkerhet. Ser deskriptiv statistikk, gjennomsnitt og variasjon i de tre ni-

våene, og ser størrelsen på gruppene. Her er gruppene ulike, og litt ulik variasjon, noe som ikke er helt bra.

Tabell 1

	Sum of Squares	df	Mean Square	X:3(f)	Sig.
Between Groups	163.474	2	X:2 (81.74)	X:4(10.55)	X:5(p<.01)
Within Groups	596.526	77	7.747		
Total	760.000	X:1 (79)			

Tabell 2

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1.00	15	21.8667	2.44560	.63145	20.5123	23.2210
2.00	36	23.5000	2.88345	.48058	22.5244	24.4756
3.00	29	25.7241	2.81446	.52263	24.6536	26.7947
Total	80	24.0000	3.10165	.34678	23.3098	24.6902

Oppgave 2

I det samme forskningsprosjektet ønsker du også å studere hvor like partene i et ekteskap/samboerforhold er når det gjelder selvrapportert helse.

- a) Er det en sammenheng mellom kvinnen og mannens helse i dette utvalget? Formuler hypotesene og bruk signifikansnivå.01. Benytt deg av verdiene i tabellen under for å regne ut.

	Kvinner «Helse» (n 40)	Mann «Helse» (n 40)	Mann*Kvinne
Gjennomsnitt	23.95	24.05	
Standardavvik	2.81	3.40	
Σ «Helse»	958	962	
Σ «Helse» ²	23252	23588	
Σ Kvinne «Helse»*Mann «Helse»			23324

H0: $\rho = 0$

H1: $\rho \neq 0$

Bruk tabell for kritiske verdier for r, sig. 0.01 , df 38, .45

Correlations

		helse Kvinne	Helse-Mann
helse Kvinne	Pearson Correlation	1	.762**
	Sig. (2-tailed)		.000
	N	40	40
HelseMann	Pearson Correlation	.762**	1
	Sig. (2-tailed)	.000	
	N	40	40

**. Correlation is significant at the 0.01 level (2-tailed).

- c) Av utdanning eller partner/ektefelles helse, hva forklarer mest av variasjonen i selvrapportert helse? Undersøk om det er ster-

kest sammenheng mellom utdanning og helse (1 a) eller mellom helsen til partene i et ekteskap/samboerforhold (2a)? Vis utregninger og forklar hva resultatet betyr.

Svar: (Kvadrert eta SS gruppe / SStotal = .45, tilsvarer r i andre. Sterk effekt, (regner man ut kvadrert omega blir verdien noe lavere, - har ingen betydning hva studentene velger), r mellom partnernes helse er .76, noe som tilsvarer en r i andre på .577, .58, altså 58 prosent av variasjonen i selvrapportert helse kan forklares ut av partnerens selvrapporterte helse, mens kun 45 % av variasjonen kan forklares ut fra utdanning.

Sterkest effekt i 2 a, altså partners helse har mer å si enn utdanning.

- c) Basert på resultatene i oppgave 2 b og informasjon gitt i oppgavesettet, har vi mest statistisk styrke (Power) i 1 a eller 2 a? Begrunn svaret. Ingen utregninger nødvendig.
Svar: Begge parametriske tester- bra- mer styrke enn ikke parametrisk. Større effektstyrke i 2a- indikerer mer statistisk styrke, Samme utvalg,
- d) For hele populasjonen er gjennomsnittet (μ) på skalaen «Helse» 23.02 og standardavviket i populasjonen (σ) er 3. Er mennene i dette utvalget signifikant forskjellig fra populasjonsgjennomsnittet? Formuler hypotesene og regn ut. Benytt signifikansnivå .05. Forklar kort hva resultatet betyr.

Svar: Z for å teste hypoteser om gjennomsnitt siden MJU er kjent. $\frac{24.05 - 23.02}{1.03 / \sqrt{40}} = \frac{1.03}{.161} = 6.32$ dvs, signifikant forskjellig. Resultatet betyr at mennene som gruppe skårer høyere på helse, altså- at disse mennene er litt friskere enn landsgjennomsnittet, -kan i utgangspunktet by på generaliseringsproblematikk men ingen grunn til å tro at sammenhengen mellom variablene er annerledes her enn ellers. Spør uansett ikke om implikasjoner, kun hva resultatet betyr. Forkaster H_0 . Her kan det komme variasjoner, one sample, z-skårer for enkeltobservasjoner osv, men Z-for gjennomsnitt (med standardfeil i stedet for standardavvik) er helt klart MEST rett siden det står at Sigma er kjent.

Oppgave 3

I utvalget så har 12 menn og 17 kvinner universitets eller høyskoleutdannelse. Er det en signifikant overrepresentasjon av kvinner i denne gruppen? Formuler hypoteser og regn ut, benytt signifikansnivå .05

Svar: Hypoteser: Det er lik kjønnsfordeling i gruppen med høy utdanning. Det er ikke lik osv

Gof, $\frac{29 - 2 \cdot 14.5}{14.5} = \frac{14.5 - 12}{14.5} = .172$, kritisk verdi = XX, Ingen signifikant kjønnsforskjell når det gjelder det å ha høyere utdanning. Beholder H_0

Oppgave 4

a) Gjør rede for de ulike typene t-tester. Hva er forutsetningene for å benytte de ulike t-testene og hva gir de svar på?

Svar: one-sampel,- tester et gruppegjennomsnitt opp mot et referansepunkt. Når man ikke vet sigma. Må ha kontinuerlig variabel for å lage gjennomsnitt (gjelder alle t-tester). Må ofte vite gjennomsnitt i populasjonen (mju). Uavhengig/urelatert: to gjennomsnitt som er uavhengig av hverandre, grupperingsvariabel med to verdier, en avhengig variabel som er kontinuerlig, helst lik varians (testes med levene's test). Avhengig/relatert; når det er avhengighet mellom parvise observasjoner. Må ha data for begge par, ellers må de elimineres, normalfordelt avhengig variabel. To observasjoner (eks. kvinner menn i et par, eller mest vanlig, pre-test-post-test. Kun to målepunkt, ellers family-wise error)

b) Beskriv en forskningsdesign hvor det passer å benytte en av t-testene. Gjør rede for hovedelementene i designet. Formuler hypotesene og gjør rede for variablene som inngår.

Svar: se a), -viktig at de presiserer variablene, avhengig/uavhengig, definerer populasjonen, utvalgstrategi, operasjonaliseringen av variabelen. Design kan være Pre-test-post-test, en gruppe. Enkeltpersondesign (statistisk gruppesammenligning) Kvasi-eksperimentdesign, to grupper, ikke pre-test.

c) Hvilke egenskaper ved forskningsdesignet du har beskrevet i 4 b er avgjørende for gyldigheten til resultatet av t-testen? Reliable mål, kontroll av andre faktorer om det er et eksperiment, riktig utvalg, kontroll over trusler for indre validitet, bruk av reliable mål osv osv. Viktigste elementene er: utvalg, inndeling i grupper, grad av kontroll, operasjonalisering av variabler, reliable mål, indre vs ytre validitet osv.

Oppgave 5

Gjør rede for hovedforskjellen mellom regresjon og korrelasjon. Hva er forutsetningene for å kunne gjøre en regresjonsanalyse? Hvilken funksjon har a og b i en regresjonsanalyse?

Svar: korrelasjon som deskriptivt mål på en sammenheng mellom to «likeverdige variabler». Kan signifikant teste r (Rho) for å se om den er signifikant forskjellig fra 0, -da er det slutningsstatistikk. Regresjon går videre, ønsker å predikere verdier på en variabel (y) basert på verdier av (X), behandler den ene som prediktor(uavhengig) og den andre uavhengig. Begge analysene krever et lineært forhold mellom variablene, normalfordelt y variabel osv. I regresjon bruker man normalligninger for å regne ut hvor regresjonslinja (den man benytter til prediksjonen skal starte(a) og hvor bratt den skal være(b). a er y 's verdi når x er 0, og b er hvor mye y øker når x øker med en enhet. a og b definerer hvor linja skal gå for å minimere feilprediksjon, og den tar hensyn til alle observasjoner, slik at uteliggere kan forskyve linja mye i små utvalg. Både korr og reg bruker Co-variants i utregning. B og r er lik når standardavvikene til r er like, derfor vil standardisert beta tilsvare r i en enkel regresjon. I regresjon kan man også putte inn flere uavhengig variabler og da kalles det multippel regresjon og andel forklart varians symboliseres med stor R i andre i stedet for liten r i andre som i korrelasjon sier noe om hvor mye av variasjonen i den ene variablene som forklarer i d