

Sensorveiledning eksamen SOS1120 V2016

Oppgave 1.

Riktig svar: De fleste har få straffereaksjoner, men et mindretall har ganske mange.

Det er altså noen ekstremverdier som trekker gjennomsnittet opp og derfor er gjennomsnittet en god del høyere enn medianen.

Oppgave 2.

I oppgave 2 ble ikke spørsmålet forstått som intendert. Denne oppgaven sees derfor bort fra i sensureringen.

- a) En mulighet er å betrakte variabelen som om en kontinuerlig variabel og dermed regne ut gjennomsnitt og sammenligne disse. Man kan da bruke f.eks. t-test. En annen mulighet er å se bort fra at kategoriene er graderte og betrakte variabelen som en kategorisk variabel. Da kan man kode om til en dummy, f.eks. om man har mye tillitt eller ikke, og sammenligne andeler. Man kan da bruke f.eks. kji-kvadrat-test.
- b) Hvis man betrakter variabelen som kontinuerlig er ulempen at variabelen strengt tatt er ordinal og man vet ikke helt hvor langt det er mellom kategoriene. Å regne gjennomsnitt er derfor ikke helt uproblematisk. Hvis man bruker dummykoding, så kaster man bort mye informasjon ved at graderingen blir borte. I begge alternativer er det uklart hva man skal gjøre med kategorien «vet ikke».

Oppgave 3.

Riktig svar: Det er ingen årsakssammenheng mellom X og Y fordi det finnes en tredje variabel, Z, som forklarer både X og Y.

Oppgave 4.

Riktig svar: Usikkerheten ved resultatet blir veldig stor.

(Her vil nok mange svare at man ikke kan generalisere, men siden det er spesifisert at det er tilfeldig trukket er det ikke riktig svar. Usikkerheten kan derimot bli så stor at man ikke får noe ut av det).

Oppgave 5.

- a) Man må forutsette at utvalget er tilfeldig trukket fra populasjonen slik at det ikke er noen systematisk skjevhet i utvalget. Det at det er en svarprosent på 48 gir store muligheter til systematiske skjevheter.
- b) Nei, det er ingen garanti, strengt tatt bare at det ikke er sterk seleksjon på akkurat de variablene.

Oppgave 6

- a) Med «5% nivå» siktes det her til $p < 0.05$. Svaret er dermed 0.05.
- b) Oppslaget er 0.6915, så å få høyere enn dette blir $1 - 0.6915 = 0.31$
- c) Oppslaget er 0.0505, så riktig svar er 0.05.
- d) Oppslaget er 0.9938, så riktig svar er 0.99

Svarene gis med 2 desimaler fordi Insperas tallfelt kun tar to desimaler. Det medfører i seg selv et rimelig slingringsmonn i svaret, men de må i hvert fall treffe riktig linje i tabellen.

Oppgave 7

Riktige utsagn er:

- Hvis konfidensintervallet er for en differanse i gjennomsnitt mellom to grupper, og intervallet rommer verdien null, så vil en t-test konkludere med å **beholde** nullhypotesen
- Hvis man gjør studien mange ganger vil vi forvente at 95% av konfidensintervallene rommer den sanne verdien man ønsker å estimere
- Konfidensintervallet er vel så informativt som punktestimatet

Oppgave 8

- a) Standardfeilen er 0.008.

Utrekning:

$$\text{Varians: } p(1 - p) = 0.142(1 - 0.142) = 0.122$$

$$\text{Standardavvik: } \sqrt{0.122} = 0.349$$

$$\text{Standardfeil: } \frac{0.349}{\sqrt{1808}} = 0.008$$

$$\text{b/c) Konfidensintervallet: } 0.142 \pm 1.96 \times 0.008 = [0.126, 0.158]$$

Oppgave 9

- a) Nullhypotese: Det er ingen forskjell i helseplager mellom menn og kvinner
Alternativ hypotese: Det er en forskjell i helseplager mellom menn og kvinner

- b) Kji-kvadrat test: 19.6

Her må det regnes ut forventede verdier først og deretter kji-verdien:

Forventede verdier:

$$\text{Forventet mann og nei: } \frac{1551 \times 843}{1808} = 723.2$$

$$\text{Forventet kvinne og nei: } \frac{1551 \times 965}{1808} = 827.8$$

$$\text{Forventet mann og ja: } \frac{257 \times 843}{1808} = 119.8$$

$$\text{Forventet kvinne og ja: } \frac{257 \times 965}{1808} = 137.2$$

$$\chi^2 = \sum \frac{(O-F)^2}{F} = \frac{(756-723.2)^2}{723.2} + \frac{(795-827.8)^2}{827.8} + \frac{(87-119.8)^2}{119.8} + \frac{(170-137.2)^2}{137.2} = 19.6$$

c) Nullhypotesen forkastes. Begrunnelse: Med $df=1$ er kritisk grense på 5% nivå er 3.84 og på 1% nivået 6.64. Den utregnede χ^2 -verdien er på 19.6 som er langt høyere enn kritisk grense. De observerte dataene er derfor neppe resultat av slumpvariasjon.

Oppgave 10

a) Nullhypotese: det er ingen forskjell i inntekt mellom de med og uten helseplager.

Alternativ hypotese: det er en forskjell i inntekt mellom de med og uten helseplager.

b) Differanse: $798.4 - 601.0 = 188$

c) Standardfeilen til differansen:

$$\sqrt{[SE(m_1)]^2 + [SE(m_2)]^2} = \sqrt{\left[\frac{343.9}{\sqrt{214}}\right]^2 + \left[\frac{640.9}{\sqrt{1368}}\right]^2} = \sqrt{852.9} = 29.2$$

d) t -verdien er $188/29.2 = 6.5$

e) Vi forkaster nullhypotesen. Begrunnelse: med stort antall frihetsgrader bruker vi normalfordelingen og får kritisk grense 1.96 (5%-nivå). Den utregnede t -verdien er langt høyere enn dette og differansen skyldes neppe slumpvariasjon.

Oppgave 11

a) t -verdi: 19.47 og antall frihetsgrader er 10539

Utregning:

$$\frac{(11.80 - 14.97)}{\sqrt{0.11^2 + 0.12^2}} = \frac{-3.17}{0.1628} = 19.47$$

$$Df = (N-1) + (N-1) = (5579-1) + (4962-1) = 10539$$

Konklusjon: Med $df=10539$ bruker vi normalfordelingen og får kritisk grense 1.96. Utreget t-verdi på 19.47 overstiger denne og vi forkaster nullhypotesen.

b) Ved en enhets økning på ADHD-indeksen synker school attachment (indeks) med 0.06. Altså: økt ADHD-indeks gir mindre tilknytning til skolen.

c) Ved en enhets økning på ADHD-indeksen synker school attachment (indeks) med 0.07. Altså: økt ADHD-indeks gir mindre tilknytning til skolen.

d) Forskjellen i school attachmentindeksen er 2.5.

Her spørres det om forskjellen mellom to predikerte verdier fra regresjonen og man må regne ut en gang for hver verdi. Man behøver ikke legge inn verdier for alle X, men kun den delen som varierer. I dette tilfellet to forskjellige verdier for ADHD. (De øvrige variablene nulles ut, men det er også ok å regne ut for noen gitt verdier av øvrige variable).

Differansen mellom predikerte verdier for to ulike verdier av ADHD kan vi her omtale som $y_1 - y_2$. Øvrige variable er Z.

$$\begin{aligned}y_1 - y_2 &= (b_0 + b \times ADHD_1 + bZ) - (b_0 + b \times ADHD_2 + bZ) \\ &= b \times ADHD_1 - b \times ADHD_2 \\ &= b \times ADHD_1 - b \times ADHD_2\end{aligned}$$

Verdiene som skal settes inn finnes i den vedlagte tabellen for deskriptiv statistikk. Gjennomsnittlig ADHD for kvinner er 11.8 og maksverdi er 54. Vi setter dette inn i ligningen som følger:

$$\begin{aligned}&= (-0.06 \times 54) - (-0.06 \times 11.8) \\ &= -3.24 - (-0.708) \\ &= -2.5\end{aligned}$$

Svaret kan avhenge en del av antall desimaler i mellomregningene.

e) T-test av regresjonskoeffisienten er:

$$t = \frac{-0.07}{0.01} = -7$$

f) 7 er lang høyere enn kritisk grense, så vi forkaster nullhypotesen. Altså: Det er en sammenheng mellom ADHD-skår og tilknytning til skolen.

g) De øvrige variablene er forskjellig fordelt mellom kjønnene, særlig ADHD-indeksen, slik at gruppene i utgangspunktet ikke var like. Når man så kontrollerer for forskjellene slår det ut på konstantleddet.