

Løsningsforslag til Eksamen høst 2011

Oppgave 1

a) Parvis, det er samme kvinne. Migreanfallene uten og med medisin for en kvinne er avhengige variable. Det blir derfor direkte feil å analysere dette som to uavhengige utvalg. La X_i være antall årlige anfall uten Nembuxil, og Y_i være antall anfall for samme kvinne med bruk av Nembuxil (kvinne nr. i). Vi lar D_i være $X_i - Y_i$.

Modell: $D_i \sim N(\mu_D, \sigma_D)$, der alle D_i -ene er uavhengige. $i = 1, 2, \dots, 9$.

Estimert effekt av Nembuxil ($\hat{\mu}_D = \bar{D}$) er 2,9. ($\hat{\sigma}_D = s_D = 3,3$)

b) $H_0: \mu_D = 0$ (det vil si ingen effekt) mot $H_1: \mu_D > 0$ (det vil si positiv effekt)

$$\text{Testobservator: } T = \frac{\bar{D}}{s_D/\sqrt{9}} = \frac{2,9}{3,3/3} = 2,64.$$

Må sjekkes med t-tabell med 8 frihetsgrader som er 1,86 ved 5 % nivå. Dermed kan vi forkaste H_0 og påstå av Nembuxil har positiv effekt.

(For øvrig er p-verdien 0,015, men den er det ikke spurt om)

c) Resultatet kan sammenfattes slik:

	A	B	C	Totalt
Redusert	24	26	27	77
Ikke redusert	8	5	10	23
Totalt	32	31	37	100

$H_0: P(\text{Reduksjon} \cap \text{Type}_i) = P(\text{Redusjon}) * P(\text{Type}_i)$ for alle typer medisin. (Uvhengighet)

$H_1: P(\text{Reduksjon} \cap \text{Type}_i) \neq P(\text{Redusjon}) * P(\text{Type}_i)$ for minst en type medisin. (Avhengighet)

Vi får følgende:

Expected Counts

A	B	C
24.64	23.87	28.49
7.36	7.13	8.51

Chi-square Components

A	B	C
0.02	0.19	0.08
0.06	0.64	0.26

Dette gir $Q = 1,25$. Dette er mindre enn 4,61 som er forkastningsgrense på 10 % nivå.

Vi kan ikke forkaste H_0 , og har dermed ikke klart å påvise sammenheng mellom type og effekt.

Oppgave 2

a) For begge modeller har α ingen fysisk tolkning da vi ikke kan tenke oss bjørner uten brystomkrets eller nakkeomkrets.

For Modell 1 tolker vi β som forventet (gjennomsnittlig for alle bjørner i populasjonen) økning i vekt dersom brystomkrets øker med 1 cm, og σ som standardavvik i vekt for alle bjørner med samme brystomkrets.

For Modell 2 tolker vi β som forventet (gjennomsnittlig for alle bjørner i populasjonen) økning i vekt dersom nakkeomkrets øker med 1 cm, og σ som standardavvik i vekt for alle bjørner med samme nakkeomkrets.

Modell 1: $\hat{\alpha} = -133,52$, $\hat{\beta} = 2,73$,

$\hat{\sigma}$ finner du ved å ta kvadratroten av (SSresidual/22)

Modell 2: $\hat{\alpha} = -118,23$, $\hat{\beta} = 4,25$.

Siden det er samme respons er SSTotal den samme for begge modeller: SSTotal = 113940, dermed blir SSresidual : $113940 - 96678 = 17262$.

$\hat{\sigma}$ finner du ved å ta kvadratroten av (SSresidual/22)

b) Modell 1: SSTotal = 113940. $R^2 = SSREG/SSTotal = 0,91$.

Modell 2: Samme SSTotal som i Modell 1. $R^2 = SSREG/SSTotal = 0,85$

R^2 måler andel variasjon i vekt som er forklart av henholdsvis variasjon i brystomkrets og variasjon i nakkeomkrets.

c) Et 95 % KI for β er gitt ved: $\hat{\beta} \pm t_{0,025,22} S.E.(\hat{\beta})$. Dette gir:

Modell 1: $2,73 \pm 2,074 * 0,18$ eller (2,36, 3,10)

Alternativt: Modell 2: $4,25 \pm 2,074 * 0,38$ eller (3,46, 5,04)

Siden ingen av intervallene dekker 0, kan vi forkaste hypotesen $H_0: \beta = 0$, og tro på alternativet $H_1: \beta \neq 0$ på 5 % signifikansnivå.

Med andre ord vi har gjennom KI påvist sammenheng mellom vekt og brystomkrets og alternativt mellom vekt og nakkeomkrets.

d) $\hat{Y} = \hat{\alpha} + \hat{\beta}x$. Dermed:

Modell 1: $\hat{Y} = -133,52 + 2,73 * 90 = 112,18$.

Modell 2: $\hat{Y} = -118,23 + 4,25 * 54,3 = 112,5$.

Merker oss at dette er bjørn med forklaringsvariable identisk med senter som vi brukte for å tilpasse modellen. Dermed blir det inne i rottegnet lik 25/24 for begge modeller.

Dermed blir intervallene:

$$\hat{Y} \pm t_{0,025,22} \hat{\sigma} \sqrt{\frac{25}{24}}$$

Modell1: $112,18 \pm 2,074 * 21,12 * 1,02$ eller (51.9, 173.4)

Modell2: $112,5 \pm 2,074 * 28,01 * 1,02$ eller (31.9, 193.1)

Bredde er $2t_{0,005,22} * \hat{\sigma} * 1,02 = 5,75 \hat{\sigma}$.

Siden $\hat{\sigma}_1 = 21.12$, og $\hat{\sigma}_2 = 28.01$, og dette er bjørn med forklaringsvariable identisk med senter som vi brukte for å tilpasse modellen, blir PI smalest for Modell 1.

e) Vi antar at $\varepsilon_i \sim N(0, \sigma)$, der $\varepsilon_i = Y_i - (\alpha + \beta x_i)$,

Dermed bør residualene $\hat{\varepsilon}_i = Y_i - (\hat{\alpha} + \hat{\beta}x)$ oppføre seg omtrent likedan.

Residualplottet bør ikke vise noen form for mønster, verken i form av kurver, eller ved at for eksempel variansen til disse øker med økende x-verdi (eventuelt ved økende y-hatt-verdi). I tillegg bør et histogram over residualer ha en tilnærmet klokkeform. Det er ikke lett å se fra et residualplot, men en kan undersøke om medianen er langt fra 0 ved å se om det er omtrent like mange positive som negative residualer.

Vi kan skimte en kurveform, ser også at små bjørner får positive residualer, middels får negative og store får positive residualer.

Det å avgjøre konstant varians kan være litt verre, men det er kanskje mulig å se økende variasjon i residualene ved økt brystomfang (her representert ved økt tilpasset vekt, y -hatt).

Nr 14, får en negativ tilpasset verdi. Dette er en lett bjørn med et lite brystomfang (bjørnunge), og det viser at modellen er ubruktbar for denne.

Nr 18 har en vekt omtrent som forventet. Denne er en slags gjennomsnittsbjørn blant de med samme brystomfang som denne.

Nr 8 er mye lettere enn det brystomkretsen skulle tilsi.

Nr 22 er mye tyngre enn det brystomkretsen skulle tilsi.

Det er en tendens til at hanner får positive residualer, og hunner negative residualer. Hanner er egentlig tyngre enn det modellen skulle tilsi. Det er kanskje en ide og lage en modell for hannbjørner og en for binner (hunn-bjørner). Eller kanskje enda bedre: Lage en egen modell for bjørnunger i tillegg.

Oppgave 3

a) Dette er en enveis variansanalysemodell.

$$Y_{ij} \sim N(\mu_i, \sigma), i = 1, 2, 3. j = 1, 2, \dots, 20.$$

Alle observasjoner er uavhengige.

Y_{ij} er skalltjukkelse på egg nr j for pelikan som lever på sted nr. i .

$$\text{Estimater: } \hat{\mu}_1 = 0.288, \quad \hat{\mu}_2 = 0.317 \quad \hat{\mu}_3 = 0.325. \quad \hat{\sigma} = \sqrt{0.0021} = 0.05$$

$H_0: \mu_1 = \mu_2 = \mu_3$. H_1 : Minst to μ -er er forskjellige.

$F = 0.0075/0.0021 = 3.57$. Denne må sammenlignes med 3.16 som er kritisk verdi ved 5 % signifikansnivå. F er større, vi forkaster H_0 , og påstår at skalltjukkelse avhenger av sted.

P -verdi = 0.034. Det betyr at det er ca 3,4 % sannsynlig å oppsere så store forskjeller i skalltjukkelse mellom steder i utvalget, selv om det ikke er forskjell i populasjonen.

b) La $\theta = \mu_1 - \mu_2$. Da er $\hat{\theta} = -0,029$, og $SE(\hat{\theta}) = \sqrt{\frac{MSE * 2}{20}} = 0,015$.

$H_0: \theta = 0$ mot $H_1: \theta \neq 0$

Under H_0 så er $T = \frac{\hat{\theta}}{SE(\hat{\theta})}$ t, fordelt med 57 frihetsgrader.

Vi får $T = -1.93$. Forkaster på 10 % nivå hvis $|T| > 1,67$ (litt avrundet). Dette skjer, og vi kan påstå at det er forskjell i forventet skalltjukkelse mellom California og Florida.