Stat100 vår 2018

**Løsning til Uke 17**

**Oppgave 1 (se scannet løsning)**

**Oppgave 2**

Vekt er respons, høyde er forklaringsvariabel

Har ingen tolkning fordi vi ikke kan snakke om menn med høyde 0 cm

: Gjennomsnittlig økning i vekt ved å vokse 1 cm i høyde.

Spredning (som standardavvik) i vekt for alle med samme høyde.

i Forskjell i observert vekt for person nr. i og gjennomsnittlig vekt for alle med samme høyde som person nr. i.

Høyde er respons, alder er forklaringsvariabel (gjelder for barn opp til 14 år)

Gjennomsnittlig høyde (lengde for nyfødte)

: Gjennomsnittlig økning i høyde ved å for barn som blir ett år eldre.

Spredning (som standardavvik) i høyde for alle barn som er like gamle.

i Forskjell i observert høyde for barn nr. i og gjennomsnittlighøyde for alle barn som har samme alder som barn nr. i.

Bensinforbruk (for bil) er respons, motorstørrelse er forklaringsvariabel

Har ingen tolkning fordi vi ikke kan snakke om biler med motor uten hestekrefter.

: Gjennomsnittlig økning i bensinforbruk ved å ved å øke motorytelsen med 1 hestekraft.

Spredning (som standardavvik) bensinforbruk for alle biler med samme motorstørrelse.

i Forskjell i observert bensinforbruk for bil nr. i og gjennomsnittlig bensinforbruk for alle biler med motorstørrelse som bil nr. i.

Eksamenspoeng i stat100 er respons, ukentlig timer nedlagt i stat100 forberedelser gjennom semesteret er forklaringsvariabel.

Gjennomsnittlig poeng for de som ikke leser

: Gjennomsnittlig økning i poeng ved å øke lesetid med 1 time i uka.

Spredning (som standardavvik) i eksamenspoeng for alle som leser like mye.

i Forskjell i observert poengsum for student nr. i og gjennomsnittlig poengsum for alle som leser like mye som han/hun.

**Oppgave 3**

Plot selv, Graph- Scatterpolot, gå til OPTION og fjern alle haker.

Med hensyn på bruktbilpris ser det ut som om alder betyr mer enn kjørelengde.

**b** Estimate Std. Error t value Pr(>|t|)

(Intercept) **74.142** 10.298 7.200 9.25e-05 \*\*\*

aar **13.271** 2.948 4.501 0.002 \*\*

s: **19.47** on 8 degrees of freedom

Multiple R-squared: 0.7169,

Estimate Std. Error t value Pr(>|t|)

(Intercept) **157.9189** 17.5528 8.997 1.86e-05 \*\*\*

tusenkm **-0.6150** 0.2064 -2.980 0.0176 \*

s: **25.19** on 8 degrees of freedom

Multiple R-squared: 0.5261.

**Estimater med fet rød skrift.**

Vi anslår at en bil fra år 2000 koster 74 142 kroner.

Vi anslår at årlig prisfall er 13 271 kroner i gjennomsnitt.

Standardavvik i pris for biler fra samme årgang anslåes til 19470 kroner.

Vi anslår at nybilprisen er 157 918 (litt søkt når dette er en kjent størrelse)

Vi anslår at gjennomsnittlig prisfallet er 615 kroner for hver kjørte 1000 kilometer eller 6,15 kroner mila.

Standardavvik i pris for biler med samme kjørelengde anslåes til 25 190 kroner.

**c)** 61,5 kroner

**d)** 6 635 kroner

**e)** Confint(RegModel.1, level=0.98)

Estimate 1 % 99 %

(Intercept) 74.14220 44.315674 103.96873

aar 13.27064 **4.731332 21.80995**

Vi er rimelig sikre på at intervallet (4731, 21809) dekker det årlige prisfallet.

**f)** Alder: 74 142 + 13271\*1 = 87 413

Kjørelengde: 157 919 - 615\*85 = 105 644.

Relativt forskjellig svar.

fit lwr.CI upr.CI lwr.PI upr.PI

Alder: 87.413 60.14 114.68 16.62 158.19

Kjørelengde: 105.644 78.16 133.11 16.776 194.51

Breie intervaller. Ubrukelige???

**g)** Alder: Ny bil 140 495. Mellomlegg: 53 082

Kjørelengde: Ny bil: 136.393 Mellomlegg: 30 749

**h**) Skal teste H0:  = 10 mot H1:  > 10

Under null hypotesen er  er t- fordelt med 8 frihetsgrader.

Vi finner T = 1.109

Vi kan ikke forkaste nullhypotesen. Bilselgers påstand er ikke bevist.

**i)** Spør foreleser om hjelp.

**Oppgave 4**

Yi =  + xi + i, der Yi er Ph i slakt nr. i og xi er tid slakt nr. i har hengt. , i = 1, 2, . . .10.

Vi antar at i -ene er uavhengige, normalfordelte og med forventning 0 og standardavvik .

 er forventet (gjennomsnittlig for all slakt) PH når tid er null (nyslaktet)

 er forventet endring i PH pr time.

 er spredning (standardavvik) i PH for all slakt som har hengt like lenge.

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.99649 0.09691 72.20 1.51e-12 \*\*\*

Tid -0.20869 0.01970 -10.59 5.51e-06 \*\*\*

s: 0.1595 on 8 degrees of freedom

Multiple R-squared: 0.9335,

**PH i nyslaktet dyr anslås til ca. 7.**

**b)** Confint(RegModel.1, level=0.95)

Estimate 2.5 % 97.5 %

(Intercept) 6.996494 6.7730286 7.2199592

Tid -0.208689 **-0.2541148 -0.1632632**

For hver time slakt henger kan vi med rimelig sikkerhet si at reduksjon i ph vil i gjennomsnitt for all slakt være mellom 0.16 og 0.25.

H0: Tid har ingen effekt med hensyn på på ph.  = 0

H1: Tid har effekt med hensyn på ph ≠

Vi forkaster på 5 % signifikansnivå fordi et 95 % konfidensintervall ikke dekker 0.

Merk at den svært lave p-verdien til denne testen selvsagt gir samme konklusjon.

**c)** Residual er avstand mellom observert PH og den PH verdien som anslås av regresjonslinja

ei = Yi – (6,99 – 0,21\*xi)

De to første observasjonen har residual 0,23 og 0,14. (Model- Add observation statistics to data)

**d) **

**e)** Bruk: Data – Manage variables in active dataset – compute new variables. Merk at for logaritme med naturlig grunntall skriv: log(), der du klikker tid inn i parentesen.

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.98363 0.04853 143.90 6.08e-15 \*\*\*

lntid -0.72566 0.03443 -21.08 2.70e-08 \*\*\*

s: 0.08226 on 8 degrees of freedom

**Multiple R-squared: 0.9823**,

**e)** For spesielt interessert, spør foreleser om hjelp.

**Løsning oppgave 5:**

\beta_0: Gjennomsnittlig levealder dersom alle land valgte å ikke gi penger til helsebudsjettet.

\beta_1: Forventet økning i levealder dersom et land øker bevilgning med en dollar pr innbygger.

\sigma: Spredning (som standardavvik) i levealder dersom alle land la like mye i helsebudsjettet.

Estimater: Henholdsvis:  44.2   og 0.086  og 9.41.

**b)** \hat{\beta}_1 \pm t_{0.005, 11}SE(\hat{\beta}_1) = 0,086 ±3,106\*0,02 =(0,024;   0,148)

Dersom vi øker helseutgift med en dollar er vi 99 % sikre på at økning i levealder ligger mellom 0,024 og 0,148 år

**c)** Modellen passer dårlig fordi vi ser en sterk vekst i levealder ved lave helseutgifter og en avflatende vekst når utgiftene øker. Dette samsvarer ikke med hva en rett linje kan beskrive.

**d)** Ser ut til å ikke ha noen informasjon.

**Løsning oppgave 6:**

**a)** Y_i=\alpha + \beta x_i+\epsilon_i,\:\:der\:\epsilon_i\sim N(0, \sigma)\:\:og\:\epsilon_i -ane er uavhengige, *i* = 1, ..., *n* der *n* = 5.

*\alpha* – forventa pris for eit piggfritt dekk.

\beta – forventa prisforskjell for to dekk med forskjell i antal piggar lik ein, evt. forventa endring i pris viss ein aukar talet på piggar med ein.

\sigma – populasjonsstandardavviket til prisen på dekk med same antal piggar.

\hat{\alpha}=208.72

\hat{\beta}=0.8681

\hat{\sigma}=42.9248

**b)**Testar hypotesa *H*0: \beta = 0 mot *H*1: \beta > 0 med *\alpha* = 0.05.

T=\frac{\hat{\beta}}{SE(\hat{\beta})}=\frac{0.8681}{0.1934}=4.49

Forkastar *H*0 viss *T* > *ta*.

*ta* = *t*0.05 = 2.353 med *n* – 2 = 5 – 2 = 3 friheitsgrader.

*T* = 4.49 > 2.353 = *t*0.05  ==> Forkast *H*0. Me trur at det er ein positiv samanheng mellom talet på piggar og pris.

**c)**R^2=\frac{SS_R}{SS_T}=\frac{37113}{42641}=0.870

87.0 % av variasjonen i pris kan forklarast av modellen.

**d)** . Sett inn verdier og du får x = 160 (pigger)

**Løsning oppgave 7:**

**a)** Modell:   Yi = \alpha + \betaxi + \epsilon_i  der \epsilon_i-ene er uavhengige og N(0, \sigma).

Yi er avlingnr. i, og xi er såtid nr i.        i = 1, 2, . . . .14.

Estimater (fra utskrift)  \hat{\alpha} = 554.5 \hat{\beta} = -2.82 \hat{\sigma} = 21.8

Dersom vi sår 1. april estimerer vi gjennomsnittsavling til 554,5 kg. Forventet tap i avling pr sådag utsatt estimeres til 2,82 kg. Spredning (standardavvik) for avling med samme såtid estimeres til 21,8 kg.

**b**) Modellen passer relativt bra, selv om vi kan se en svak krumming på residualplottet. R2er relativt stor. Andel variasjon i avling forklart av såtid er 68,9 %.

**c)** H0: \beta = 0. (Tidlig) Såtid har ingen betydning H1: \beta < 0 Tidlig såtid har positiv betydning.

Merk nå x øker så betyr dette at du sår seinere på våren.

T = \frac{\hat{\beta}}{SE(\hat{\beta})} er t- fordelt med 12 frihetsgrader under H0.

Dermed forkast H0 hvis T < -t\alpha,12, er \alpha  er valgt nivå.

Vi får at T = -5.13, dermed kan vi forkaste på 1 % nivå, (der t t0,01, 12 = 2,681). Vi påstår at jo tidligere du sår, jo større vil forventet avling bli.

**d)** Prediksjonsintervall: Se utskrift. Derom du sår 6. mai, er vi rimelig sikre (i betydning 95 % sikre) på at avling vil ligge mellom 397 kg og 503 kg.