Stat100

Løsning eksamen vår 2016

Oppgave 1

a) μ_d er gjennomsnittlig differanse mellom sann temperatur og varslet temperatur over en lang tidsperiode.

$\hat{\mu} = \overline{D} = 0{,}5$ (grader). Gjennomsnittlig differanse for de 10 målingene.

$SE(\hat{\mu}) = \frac{s}{\sqrt{10}} = 0{,}62.$

b) H_0: $\mu_d = 0$ mot H_1: $\mu_d \neq 0$. $T = \frac{\overline{D}}{\frac{s}{\sqrt{10}}} = 0{,}81$

Sammenlignes med $t_{0,05,\,9}$ som er 1,833. Vi kan ikke forkaste H_0, og har dermed ikke påvist systematiske forskjeller mellom varslet og sann temperatur.

c)
```
95 percent confidence interval:
 -0.90  1.90
```
Vi er 95 % sikre på at dette intervallet dekker den gjennomsnittlige differansen mellom målt og sann temperatur.

d) Ikke parvise data krever uavhengighet mellom alle observasjoner. Det er utenkelig at varslet temperatur for en dag er helt uavhengig av den sanne temperaturen, i så fall er temperaturvarsling helt meningsløs.

Oppgave 2

a) La Y_i være tilvekst og x_i være proteinprosent i foret for fisk nr. i
Vi antar følgende modell:
$Y_i = \alpha + \beta x_i + \varepsilon_i$ for i = 1, 2, . . .50.
Alle ε_i er uavhengige og normalfordelte med forventning 0 og standardavvik σ.

Det er 3 parametere, disse er α, β og σ.
Disse estimeres med:
$\hat{\alpha} = 211.8 \quad \hat{\beta} = 0{,}93 \quad \hat{\sigma} = 7.1$ (alt målt i gram)

$\hat{\alpha} = 211.8$
Dette kunne tolkes som at et anslag på gjennomsnittlig tilvekst dersom du gir for helt uten protein ville være 211,8 gram. Men dette har vi ikke data i nærheten av, slik at vi ikke vet om modellen er gyldig for slike proteinverdier.

$\hat{\beta} = 0{,}93$
For hver økning av proteininnholdet med en prosent anslår vi at gjennomsnittlig økning i tilvekst blir på 0,93 gram.

$\hat{\sigma} = 7.1$:
For all fisk som får like stor mengde protein i foret, anslår vi at spredning i vekt (i form av et standardavvik) er på 7.1 gram
$R^2 =$ 0,75.
Dette betyr at 75 % av den observerte variasjon i tilvekst kan forklares gjennom variasjon i proteininnhold i foret.

b) H_0: $\beta = 0$ mot H_1: $\beta > 0$. $T = \dfrac{0{,}93}{0{,}078} = 11{,}9$

Forkastes hvis T er større enn T tabellverdi i t-fordeling med 48 frihetsgrader.
Kan forkastes på alle signifikansnivå 0,5 %. Vi er sikre på at det er en positiv sammenheng mellom proteinmengde og tilvekst.

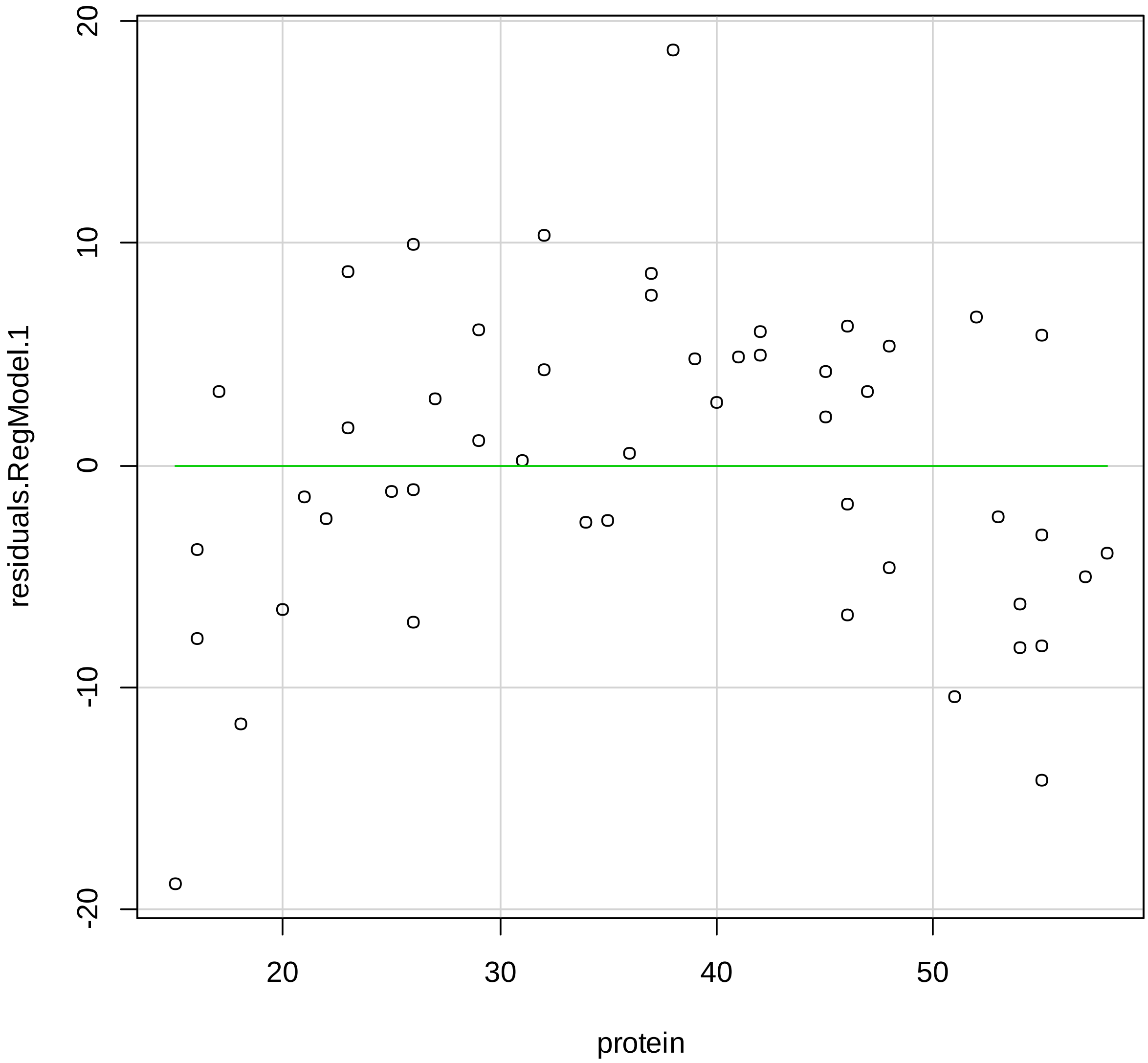
c) Predikert tilvekst: $\hat{Y} \mid (x = 40)$ = 211.8 + 0,93·40 = 249 (gram)
99 % prediksjonsintervall: (229.9; 268.4) (fra utskrift).
Vi er 99 % sikre på at en ørret som får 40 % protein i foret vil ha en tilvekst på mellom 229 og 268 gram.
Intervallet blir bredt hvis $\hat{\sigma}$ er stor, noe som igjen (sannsynligvis) betyr at σ er stor. Det vil si at det er mye uforklart variasjon.
Intervallet blir smalt hvis antallet fisk som er med i forsøket er stort (stor n)
Intervallet blir bredt hvis proteinmengden til den fisken du vil predikere vekta på ligger langt fra gjennomsnittsverdien av den proteinmengdene du hadde i forsøket.
Intervallet blir bredt, hvis de proteinmengdene du hadde i forsøket ligger svært nære hverandre.

Generelt:
Velg en god x (en som forklarer mye av Y)
Store datamengder
Spenn ut x-ene mest mulig,
Prediker i Y for x-verdier i nærheten av snittet for de x-verdien som ble brukt til å estimere linja.

d) Hver observasjon har sitt eget residual.
Dette er observert Y verdi – tilpasset Y-verdi
Eller $e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\alpha} + \hat{\beta} x_i)$
I dette tilfelle: Avstanden mellom tilveksten til en enkelt fisk på en spesiell proteinmengde, og det vi estimerer gjennomsnittstilveksten for all fisk på denne proteinmengden.

Residual for ørret nr. 1: 207 – (211.8 + 0,93 ·15) = -18,75.

Det er ikke problemer med normalfordeling, men store problemer med lineariteten. Residualplottet ser ut som en liggende halvmåne.


residuals.RegModel.1
protein
20
10
0
-10
-20
20
30
40
50

Flervalg

Oppgave	A	B	C	D	E	F
1						o
2		o				
3		o				
4	o					
5		o				
6	o					
7	o					
8					o	
9						o
10	o					
11		o				
12		o				
13					o	
14		o				
15						o
16					o	
17	o					
18			o			
19			o			
20					o	