(1).

Opps 2 Eksamen H05

Vekten til rotter
20 rotter fra en populasjon ... veies
så veies leveren.
X_i – kroppsvekt til rotte i
Y_i – levervekt — " —

$\bar{X} = 573$ g, empirisk standardavvik $\underline{S_x} = 87$ g
$\bar{Y} = 7.0$ g — " — $\underline{S_y} = 2.1$ g
Empirisk korrelasjon $\underline{r}$ mellom kroppsvekt og levervekt 0.9

a) $\hat{y} = a + bx$ (skal finne)

$\boxed{\text{har fra s. 137}}$ $b = r\dfrac{S_y}{S_x} = 0.9 \cdot \dfrac{2.1}{87} = 0.02172$

$$a = \bar{y} - b\bar{x} = 7.0 - 0.02172 \cdot 573 = -5.4456$$

Anslag for en rotte som veier 650 g:
sette inn 650 for x =>

avrundet

$$\hat{Y}_{650} = (-5.45 + 0.02\cancel{2} \cdot 650 = 7.55)$$
$$= -5.4456 + 0.02172 \cdot 650 = 8.6624$$

b)

fra side 172

$$r^2 = \frac{\text{varians av predikert } \hat{Y}}{\text{varians av observert } Y}$$

$r = 0.9 \Rightarrow r^2 = 0.81$

så 81% av variansen i levevekt
(lineært)
har forklares ~~lineært~~ fra varians i hoppevekt

$$\left(r^2 = b^2 \frac{s_X^2}{s_Y^2} = \frac{s_{\hat{Y}}^2}{s_Y^2} \right)$$

$$\left(s_{\hat{Y}}^2 = b^2 s_X^2 \right)$$

$$\boxed{\text{s. } 202 \Rightarrow \quad Y = a + bX \quad \Rightarrow \quad \sigma_{a+bX}^2 = b^2 \sigma_X^2}$$

$$r^2 = \frac{s_{\hat{Y}}^2}{s_Y^2} = \frac{\frac{1}{n-1} \sum_{i=1}^{n} (\hat{Y}_i - \bar{\hat{Y}})^2}{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

$$= \frac{\sum_{i=1}^{n} (\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

(2)

Opps 3 STK1000 eksamen H06

Sirissenes sang!

Vingen gnisses mot hverandre.
Se på gnisse frekvens og hvordan den varier med luft temperatur.

15 temperatur obs. og frekvens obs.

a) Sett opp en lineær regresjonsmodell.

s. 638: Vi må huske at vi har ett utvalg av en populasjon så vi må sette opp en regresjons linje for den sanne populasjonen ikke for utvalget!

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma)$$

estimater for β_0, β_1 er b_0 og b_1

hvor $b_1 = r \frac{s_y}{s_x}$ og $b_0 = \bar{y} - b_1 \bar{x}$

Her har minitab funnet alt så vi leser bare av

$b_0 = 6.47$, $b_1 = 0.381$

Estimatet for σ er s hvor

$$s = \frac{\sum (y_i - \hat{y}_i)^2}{n-2} \Rightarrow 0.971518 \text{ fra minitab}$$

b)

Se på

r^2 som gir hvor godt en lineær regresjon passer. 1 er perfekt

$(0 \le r^2 \le 1)$

leser fra minitab

$= (0.697)$

$r^2 = 69.7\%$, så 69.7% ~~av Y~~ ~~forklares~~ av andelen av variasjonen i Y forklares fra en lineær sammenheng.

Residualplott - kan gi en mistanke at det ikke er lineært

normalfordelingsplott - ok

c) 95% KI for forventet forekomst

lese rett av!!

95% KI = [17.452, 19.143]

Frihetsgrader

s. 6~~4~~9:

$\hat{y} \pm t^* SE_{\hat{y}}$

t^* er verdien for $t(n-2)$

$n-2 = 15-2 = 13$, så t^* for 13 frihetsgrader og $C = 0.95$ er 2.16

③

10.1

Finn 95% KI for stigning β_1.

Bruker DF = n - 2 (s. 649) og informasjon $t^* SE_{b_1} = 8.05 \cdot t^*$

(a) $n = 25$, $\hat{y} = 12.3 + 16.10x$, ~~$SE_{b_1} = 2.011$~~

$DF = 23$, $b_1 = 16.1$, $t^* = 2.069$

$b_1 \pm t^* SE_{b_1} = 16.1 \pm 2.069 \cdot 8.05$

$\Rightarrow [-0.5555, 32.7555]$

b) $n = 25$, $\hat{y} = 1.3 + 6.10x$,

$DF = 23$. $b_1 = 6.10$, $t^* = 2.069$

$6.10 \pm 2.069 \cdot 8.05 \Rightarrow$

95% KI = $[-10.5555, 22.7555]$

c) $n = 123$, $\hat{y} = 12.3 + 16.10x$

$DF = 100 \ (123) \Rightarrow t^* = \underline{1.984}$

$16.10 \pm 1.984 \cdot 8.05 \Rightarrow$

95% KI = $[0.1288, 32.0712]$

10.2

for 10.1 test $H_0: b_1 = 0$ mot $H_1: b_1 \neq 0$

$$t = \frac{b_1}{SE_{b_1}} = \frac{b_1}{8.05}, \quad DF = n-2$$

a) $DF = 23$, $b_1 = 16.1$

$$t = \frac{16.1}{8.05} = 2$$

$$P(|T| \geq 2) = 2P(t > 2) \in \cancel{2[0.010, 0.025]} \cancel{= [0.04, 0}$$

$$\in 2[0.025, 0.05]$$

$$= [0.05, 0.1]$$

b) $DF = 23$, $b_1 = 6.10$

$$t = \frac{6.10}{8.05} = 0.76$$

$$P(|T| \geq 0.76) = 2P(T > 0.76) \in [0.40, 0.50]$$

c) $DF = 100 (123)$ $b_1 = 16.1$

$$t = \frac{16.1}{8.05} = 2$$

$$P(|T| \geq 2) = 2P(T > 2) \in [0.04, 0.0.05]$$

Her er $P < 0.005$ forkast ved $\alpha \geq 0.005$

(4)

10.3

tabell hva som er feil og hvorfor.

a) parameter til enkel lineær regresjons modellen er b_0, b_1 og s

Svar: parameterne er β_0, β_1 og σ

b_0, b_1 og s er estimatorer for disse parameter

b) for å teste $H_0: b_1 = 0$ bruk t-test

svar: Vi tester $H_0: \beta_1 = 0$

c) for en verdi av forklarende variabel vil KI for gjennomsnitt respons være bredere enn prediksjons intervallet for en observasjon

svar: det vil være smallere siden

KI tar bare høyde for usikkerheten i estimatet av gjennomsnittlig respons

mens prediksjons intervallet tar høyde for tilfeldig feil til en individuell respons

(se side 649 og 677)

10.4

Målinger av vann kvalitet

bruke stemplots

område (Anal)

```
0 | 2
0 | 5 6 8 8 9 9 9
1 | 0 0 2 4
1 | 6 6 8 8 9
2 | 1 1 1 1 3 3
2 | 6 6 6 6 7 8 8 9
3 | 1 1 2 2 4 4
3 | 9
4 |
4 | 7 9 9
5 | 2 4 4
5 | 7 8 9
6 |
6 | 9
7 | 0
```

FBI

```
2 | 9 9
3 | 2 3 3
3 | 9
4 | 1 3
4 | 6 7
5 | 3 4
5 | 5 5 6 8 9 9
6 | 0 1 2 4
6 | 7
7 | 1 1 1 2 4
8 | 0 0 1 2 2 2 3 4 4
8 | 5 5 6 8 9 9
9 | 1
```

Anal er høyre skjev, $\bar{x} = 28.2851$

$s_x = 17.7192$

FBI er venstre skjev, $\bar{y} = 65.9388$

$s_y = 18.2796$

9

Forts 10.4

b) Lag skatterplot (se vedlegg)
viser en svak positiv assosiasjon
og mere scatter i y for liten x

c) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, 2, \ldots, 49$
$\varepsilon_i \sim N(0, \sigma)$

d) $H_0 : \beta_1 = 0$ mot $H_a : \beta_1 \neq 0$

e) (se vedlegg)

$FBI = 52.92 + 0.4602$ oarea1, $s = 16.53$

Hypotesen fra d) gir $t = 3.42$ og $P = 0.001$

f) se residualplott.
mer variasjon for liten x

g) venstre-skjev residualer men eller-
ser de normalfordelt.

10.6

referer til 10.4

a) Finn 95% KI for areal på 30 km^2

$$\hat{IBI} = 52.92 + 0.4602 \cdot x$$

La $x = 30 \Rightarrow$

$$\hat{IBI}_{30} = 66.73$$

Minitab gir 95% KI = (61.95, 71.50)

b) Minitab gir 95% PI = (33.12, 100.34)

c) For mange vann med areal 30 så estimerer vi gjennomsnittet IBI til å være mellom 61.95 og 71.50

For et individuell vann med areal 30 ~~estimer~~ forventer vi IBI til å være mellom 33.12 og 100.34

d) ~~ve.~~ Sannsynligvis region spesifikke data!