

LØSNING, noen oppgaver fra kap. 7.9 i Løvås: Statistikk.

Oppgave 1

Det anbefales å håndregne oppgaven for å få inn forståelsen av hvordan formlene virker, og deretter kontrollregne med kalkulatorens innebygde rutiner for regresjon, for å finne ut hvordan kalkulatoren gjør dette.

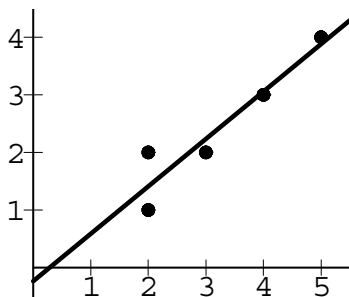
Regner først ut hjelpestørrelsene

$$\begin{aligned}\bar{x} &= \sum x_i/n = (4 + 2 + 3 + 5 + 2)/5 = 3.2 \\ \bar{y} &= \sum y_i/n = (3 + 2 + 2 + 4 + 1)/5 = 2.4 \\ \sum x_i^2 &= 4^2 + 2^2 + 3^2 + 5^2 + 2^2 = 58 \\ \sum x_i^2 &= 3^2 + 2^2 + 2^2 + 4^2 + 1^2 = 34 \\ \sum x_i y_i &= 4 \cdot 3 + 2 \cdot 2 + 3 \cdot 2 + 5 \cdot 4 + 2 \cdot 1 = 44 \\ s_{xx} &= \sum x_i^2 - n \bar{x}^2 = 6.8 \\ s_{yy} &= \sum y_i^2 - n \bar{y}^2 = 5.2 \\ s_{xy} &= \sum x_i y_i - n \bar{x} \bar{y} = 5.6\end{aligned}$$

Parametrene regnes ut (estimeres) da med

$$\begin{aligned}r &= s_{xy}/\sqrt{s_{xx}s_{yy}} = 5.6/\sqrt{6.8 \cdot 5.2} = 0.9417 \\ b &= s_{xy}/s_{xx} = 5.6/6.8 = 0.8235 \\ a &= \bar{y} - b\bar{x} = 2.4 - 0.8235 \cdot 3.2 = -0.2352\end{aligned}$$

Det er tryggest å ha med en del desimaler i mellomregningene, men i sluttsvaret begrenser vi det som oftes til ett gjeldende siffer mer enn i dataene. Det vil si at korrelasjonen er $r = 0.94$ og regresjonslikningen er $y = -0.24 + 0.82x$.



Oppgave 4

Oppgaven betyr at vi skal regne ut korrelasjonen og regresjonskurven for datasettet

Motorstørrelse x	75	145	55	88	122
Bensinforbruk y	0.48	0.52	0.53	0.97	0.78

Vi får da hjelpestørrelsene $n = 5$, $\bar{x} = 97$, $\bar{y} = 0.656$, $\sum x_i^2 = 52303$, $\sum y_i^2 = 2.332$, $\sum x_i y_i = 321.07$, $s_{xx} = 5258$, $s_{yy} = 0.1793$ og $s_{xy} = 2.91$, som gir

$$r = 2.91/\sqrt{5258 \cdot 0.1793} = \underline{\underline{0.095}}, b = 2.91/5258 = \underline{\underline{0.001}} \text{ og } a = 0.656 - 0.0006 \cdot 97 = \underline{\underline{0.602}}.$$

I dette tilfellet er det (nesten) ingen korrelasjon, og nyhetsoppslaget har ingen berettigelse. Denne ene feilregistreringen ødelegger konklusjonen helt. At n er liten (i forhold til standardavvikene på x og y) gir for det første lavt signifikansnivå, og for det andre stor følsomhet overfor feilregistreringer. Med $n = 500$ ville dette problemet antagelig forsvinne, og finner vi en signifikant økning i bensinforbruk med motorstørrelse vil denne

konklusjonen nok være uendret selv om det var en feilregistrering blant observasjonene. Regresjonslikningen og korrelasjonen ville bare endres ubetydelig. Om feilregistreringen var ekstrem (f.eks. 48 isteden for 0.48) kunne det bety noe, men ved en rask kikk på dataene (f.eks. i et spredningsplott) ville vi fort oppdage at noe var galt.

Oppgave 7

$\bar{x} = 173.67, \bar{y} = 76.00, s_{xx} = 875.33, s_{yy} = 770.00, s_{xy} = 751.00$.

$b = \underline{0.8580}$ og $a = \underline{-73.00}$, så regresjonslikningen er $y = -73.0 + 0.858x$.

Vi estimerer σ^2 ved s_e^2 , som vi får ved å sette inn estimerte størrelser fra formel 45 i formelsamlinga:

$$s^2 = \frac{s_{yy} - b^2 s_{xx}}{n - 1} = \frac{770.00 - 0.8580^2 \cdot 751.00}{6 - 1} = 31.42$$

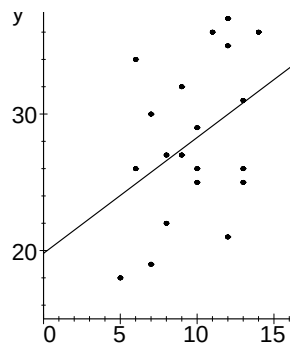
Estimatet for σ er dermed $\hat{\sigma} = \sqrt{31.42} = \underline{5.6}$.

$r = 751/\sqrt{875.33 \cdot 770.00} = 0.915$

Oppgave 9

Vi har $n = 20$ og kan regne ut hjelpestørrelsene $\bar{x} = 9.25, \bar{y} = 28.1, s_{xx} = 139.75, s_{yy} = 621.8$ og $s_{xy} = 118.5$, og fra disse få at regresjonskoeffisientene er $\hat{\beta} = b = 0.848$ og $\hat{\alpha} = a = 19.8$, så regresjonslikningen er $\underline{y = 19.8 + 0.848x}$.

Spredningsplott med regresjonslinjen tegnet inn;



Spørsmålet om avlingen øker med gjødselsmengden er om vi kan slutte at β er positiv. Det kan jo tenkes at verdien $b = 0.848$ er såpass liten at det bare er snakk om et tilfeldig utslag, selv om $\beta \leq 0$. Vi setter derfor opp

$$H_0 : \beta \leq 0 \quad \text{mot} \quad H_1 : \beta > 0$$

Fra formelsamlingen (formel (45)) har vi at $T = (B - \beta)/\sqrt{S_e^2/s_{xx}}$ har en Student's t -fordeling med $n - 2 = 18$ frihetsgrader. Hvis H_0 er sann bruker vi den β -verdien som er nærmest H_1 , altså $\beta = 0$. Om dette er tilfellet er $T = B/\sqrt{S_e^2/s_{xx}}$ Student's t -fordelt med 18 frihetsgrader. Hvis H_1 er sann får vi store observasjoner av denne, og prosedyren er å forkaste H_0 om denne blir observert til en verdi større enn 5%-fraktilen i en T_{18} -fordeling, som fra tabell 4.3 er $k = 1.734$.

Vi må regne ut estimatet for variansen σ^2 : $s_e^2 = (s_{yy} - b^2 s_{xx})/(n - 2) = (621.8 - 0.848^2 \cdot 139.75)/18 = 28.96$. Ved å sette disse verdiene inn i testobservatoren får vi da

$$t = \frac{0.848}{\sqrt{28.96/139.75}} = 1.86 \text{ som er større enn den kritiske verdien } 1.734, \text{ forkast } H_0$$

Dermed kan vi konkludere med at økt gjødselmengde gir økt avling (på 5%-nivået, men ikke på 1%-nivået der $k = 2.552$).

Ved å spre 180 tonn jevnt utover de 20 målene blir det 9 tonn per mål, som i følge regresjonslikningen gir en forventet avling på $19.8 + 0.848 \cdot 9 = 27.4$ tonn per mål, og dermed $20 \cdot 27.4 \approx \underline{550}$ tonn på hele åkeren.