Aflevering 42

3.b mat A

Kevin Zhou

3. april 2025

Opgave 1: Normalfordelt BMI

I en by er individernes BMI normalfordelt med middelværdien 23,7 kg/m^2 og spredningen 3,7 kg/m^2.

Løsning:

a. Siden en forskrift for tæthedsfunktionen for en normalfordeling med middelværdi μ og spredning σ er

$$f(x) = \frac{e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\cdot\sigma},$$

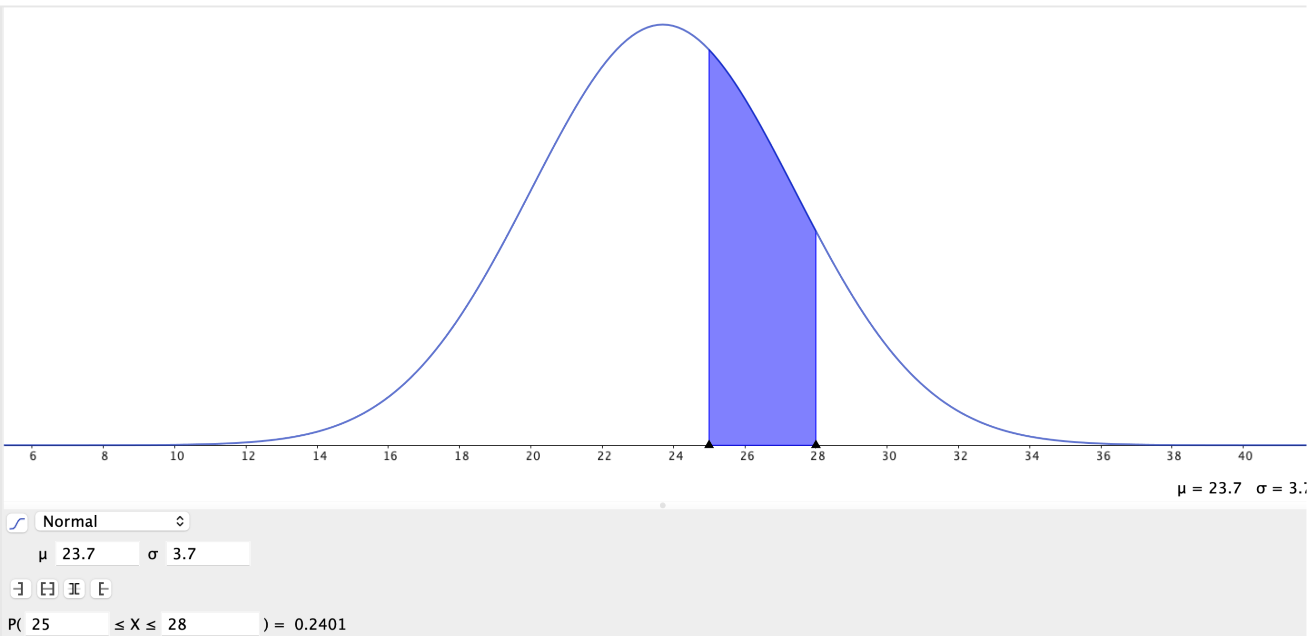
så må en forskrift for tæthedsfunktionen for fordelingen af individernes BMI i populationen være

$$f(x) = \frac{e^{-\frac{1}{2}\cdot\left(\frac{x-23{,}7}{3{,}7}\right)^2}}{\sqrt{2\pi}\cdot 3{,}7},$$

hvor x er målt i kg/m^2.

b. For at bestemme sandsynligheden for, at et individs BMI er mellem 25 kg/m^2 og 28 kg/m^2, benytter vi CAS, hvilket ses i fig. 1.

$$\begin{aligned} P(25 \le X \le 28) &= P(X \le 28) - P(X \le 25) \\ &\approx 0{,}2401 \end{aligned}$$



Figur 1: Sandsynligheden bestemt med CAS

Sandsynligheden for, at et individs BMI er mellem 25 kg/m^2 og 28 kg/m^2 er altså 0,2401.

c. De exceptionelle udfald for individernes BMI er blot udfaldende, der er over tre spredninger fra middelværdien.

$$\begin{aligned}]-\infty;\mu-3\sigma[\cup]\mu+3\sigma;\infty[&=]-\infty;23{,}7-3\cdot 3{,}7[\cup]23{,}7+3\cdot 3{,}7;\infty[\\ &=]-\infty;12{,}6[\cup]34{,}8;\infty[\end{aligned}$$

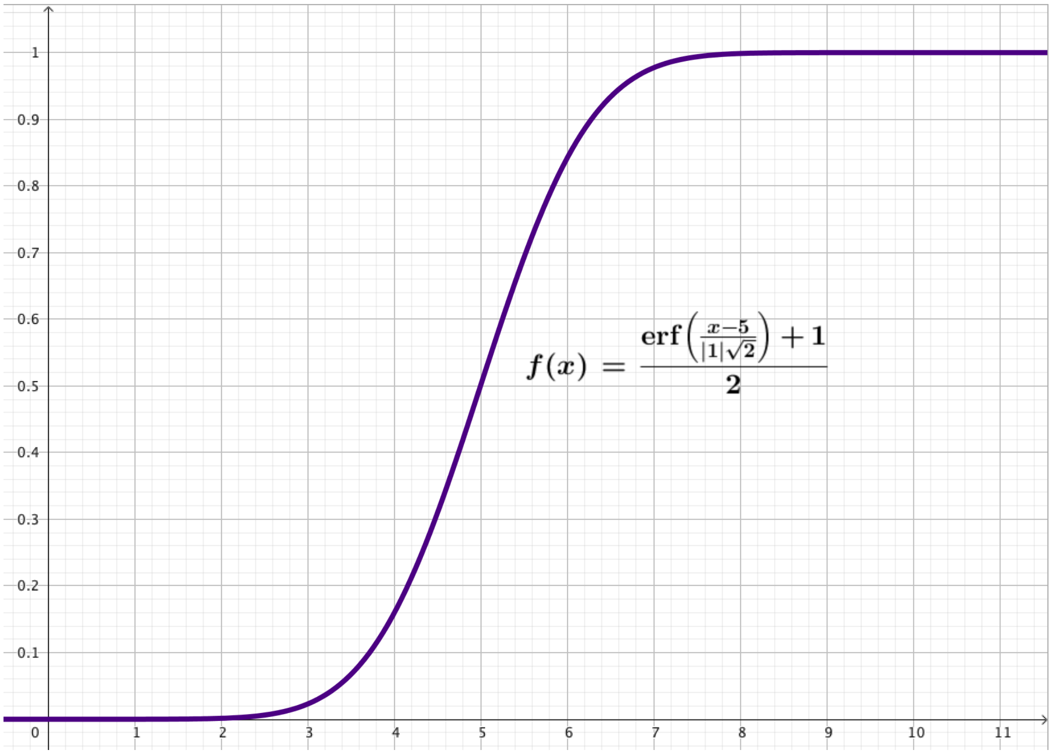
De exceptionelle udfald for individernes BMI målt i kg/m^2 er altså $]-\infty;12{,}6[\cup]34{,}8;\infty[$.

Opgave 2: Normalfordelt stokastisk variabel

En normalfordelt stokastisk variabel X er givet ved $X \sim N(5{,}1)$.

Løsning:

a. Grafen for fordelingsfunktionen hørende til X ses i fig. 2.



Figur 2: Grafen for fordelingsfunktionen hørende til X

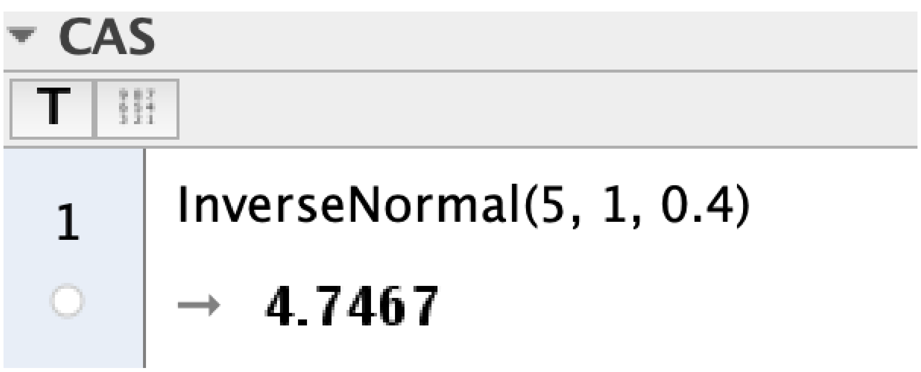
b. For at bestemme k benytter vi fraktilfunktionen. Lad F betegne fordelingsfunktionen hørende til X (bemærk F er injektiv), så gælder

$$\begin{aligned} P(X \le k) = 0{,}4 &\iff F(k) = 0{,}4 \\ &\iff k = F^{-1}(0{,}4). \end{aligned}$$

Vi beregner $F^{-1}(0{,}4)$ med CAS, hvilket ses i fig. 3.

$$\begin{aligned} k &= F^{-1}(0{,}4) \\ &\approx 4{,}7467 \end{aligned}$$

Når $P(X \le k) = 0{,}4$, så er k altså 4,7467.



Figur 3: $F^{-1}(0{,}4)$ udregnet med CAS

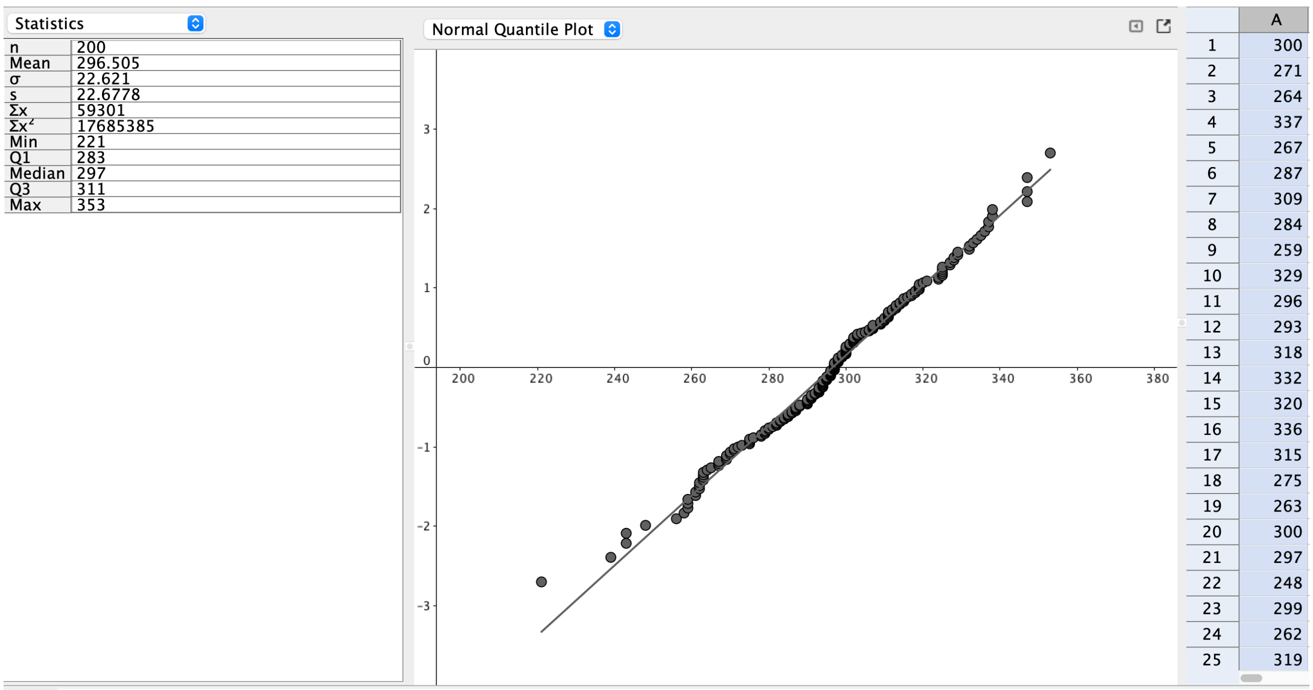
Opgave 3: Normalfordelte tulipanstilke

På et gartneri udvælger man på tilfældig måde 200 fuldt udvoksede tulipaner af en bestemt art. Man måler længden af deres stilke, og udvælger en tilfældig tulipan af denne art.

Løsning:

a. For at redegøre for, at længden af tulipanstilkene med god tilnærmelse kan beskrives ved en normalfordelt stokastisk variabel X, laves et fraktilplot, hvilket ses i fig. 4.

Det ses, at punkterne med god tilnærmelse ligger på linjen i fraktilplottet, hvilket vil sige, at længderne af tulipanstilkene med god tilnærmelse kan beskrives ved en normalfordelt stokastisk variabel X.



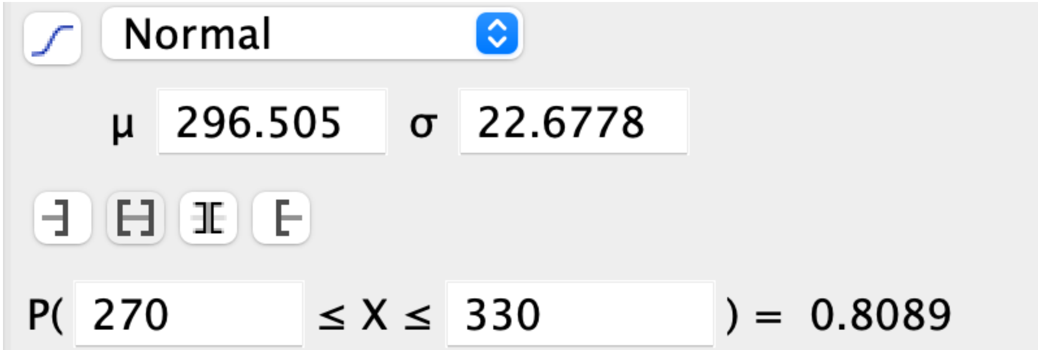
Figur 4: Fraktilplot lavet i GeoGebra

b. Fra fig. 4 har vi, at X har middelværdien (i mm) $\mu = 296{,}505$ og spredningen (i mm) $\sigma = 22{,}6778$.

c. Vi bestemmer sandsynligheden $P(270 \leq X \leq 330)$ med CAS, hvilket ses i fig. 5.

$$\begin{aligned} P(270 \leq X \leq 330) &= P(X \leq 330) - P(X \leq 270) \\ &\approx 0{,}8089 \\ &= 80{,}89\% \end{aligned}$$

Sandsynligheden for, at længden af stilken er mellem 270 mm og 330 mm er altså 80,89%.



Figur 5: Sandsynligheden udregnet med CAS

Opgave 4: Model for benzinforbrug i USA

Tabellen viser udviklingen i benzinforbruget (målt i mia. gallon) i USA i årene 1992-2004. I en model kan udvilklingen beskrives ved

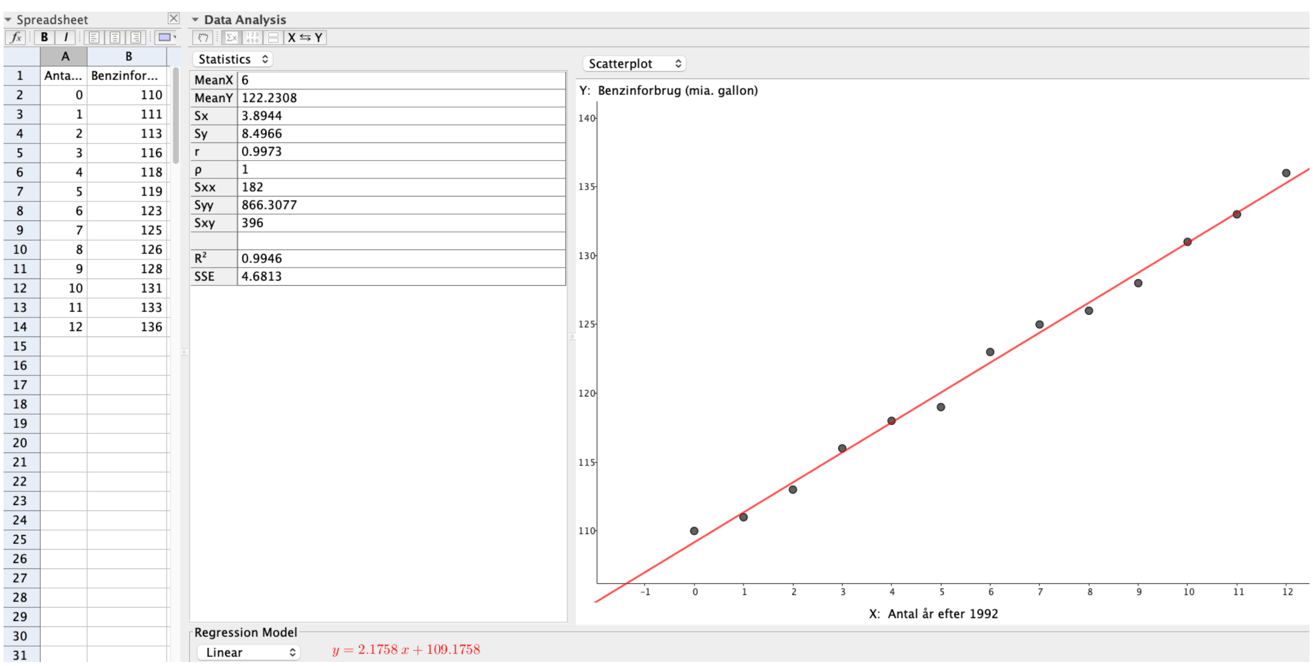
$$f(x) = ax + b,$$

hvor $f(x)$ er benzinforbruget i USA (målt i mia. gallon), og x er altsl år efter 1992. I 2019 var benzinforbruget i USA 142 mia. gallon.

Løsning:

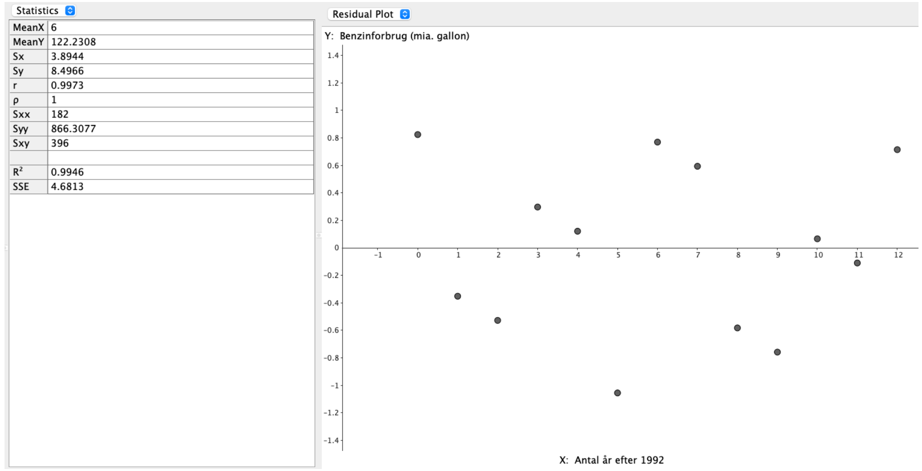
a. Vi bestemmer tallene a og b ved lineær regression på tabellens data, hvilket ses i fig. 6. Fra regressionen har vi, at

$$a = 2{,}1758 \text{ og } b = 109{,}1758.$$



Figur 6: Lineær regression på tabellens data

b. Et residualplot ses tegnet i fig. 7. Det ses, at punkterne ligger nogenlunde tilfældigt fordelt omkring linjen med forskriften $x = 0$. Altså passer modellen tilnærmelsesvist med udviklingen i perioden 1992-2004.



Figur 7: Residualplot

c. For at beregne $f(27)$, opskriver vi først en forskrift for f. Fra regressionen havde vi, at

$$f(x) = 2{,}1758 \cdot x + 109{,}1758.$$

Vi beregner nu $f(27)$.

$$\begin{aligned} f(27) &= 2{,}1758 \cdot 27 + 109{,}1758 \\ &\approx 167{,}92. \end{aligned}$$

Benizinforbruget i USA i 2019 (142 mia. gallon) var altså mindre end ifølge modellen (167,92 mia. gallon).

Opgave 5: Lineær model for blomsterdata

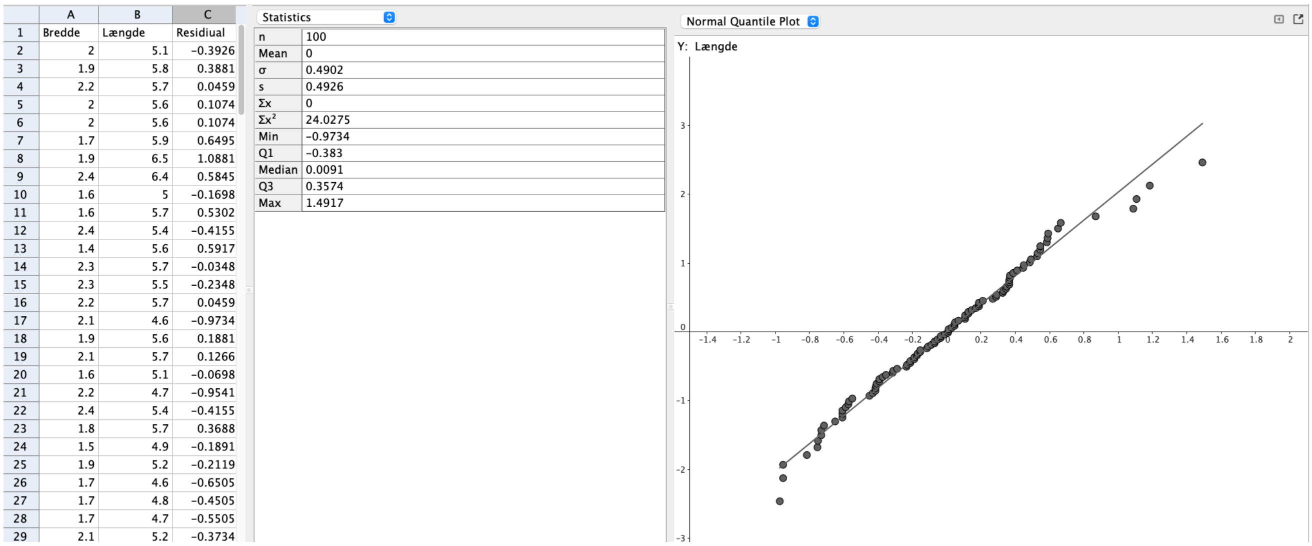
Tabellen viser en række målinger af sammenhørende værdier af nogle kronblades bredde og længde for en bestemt blomsterart. I en lineær regressionsmodel kan sammenhængen mellem bredden og længden af et kronblad beskrives ved

$$f(x) = a \cdot x + b,$$

hvor $f(x)$ betegner længden af et kronblad (målt i cm) med bredden x (målt i cm). En gruppe biologer formoder, at den lineære sammenhæng mellem længden og bredden af et kronblad er voksende.

Løsning:

a. For at redegøre for, at residualerne i modellen med god tilnærmelse er normalfordelte, laver vi et fraktilplot med residualerne, hvilket ses i fig. 8. Det ses, at punkterne med god tilnærmelse ligger på linjen i fraktilplottet. Altså kan residualerne i modellen med god tilnærmelse siges at være normalfordelte.

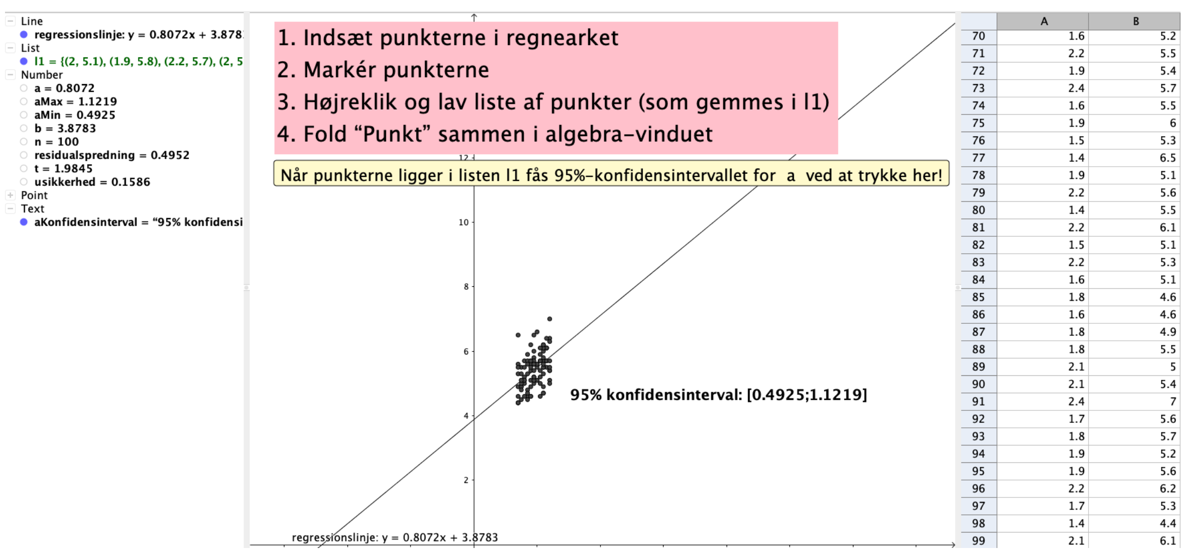


Figur 8: Fraktilplot for residualerne

b. Et 95% konfidensinterval for hældningskoeffienten i modellen findes med GeoGebra, hvilket ses i fig. 9. Fra GeoGebra har vi, at et 95 % konfidensinterval for hældningskoefficienten i modellen er $[0{,}4925; 1{,}1219]$. Siden der gælder, at

$$[0{,}4925; 1{,}1219] \subset \mathbb{R}^+,$$

så er den lineære sammenhæng mellem længden og bredden af et kronblad voksende. Altså er biologernes formodning troværdig.



Figur 9: 95 % konfidensinterval for hældningskoefficiente

Opgave 6: Lineær model for hjernevægt

Tabellen viser sammenhørende værdier for rumfanget af en mands hovede og vægten af hans hjerne. I en model kan sammenhængen beskrives ved

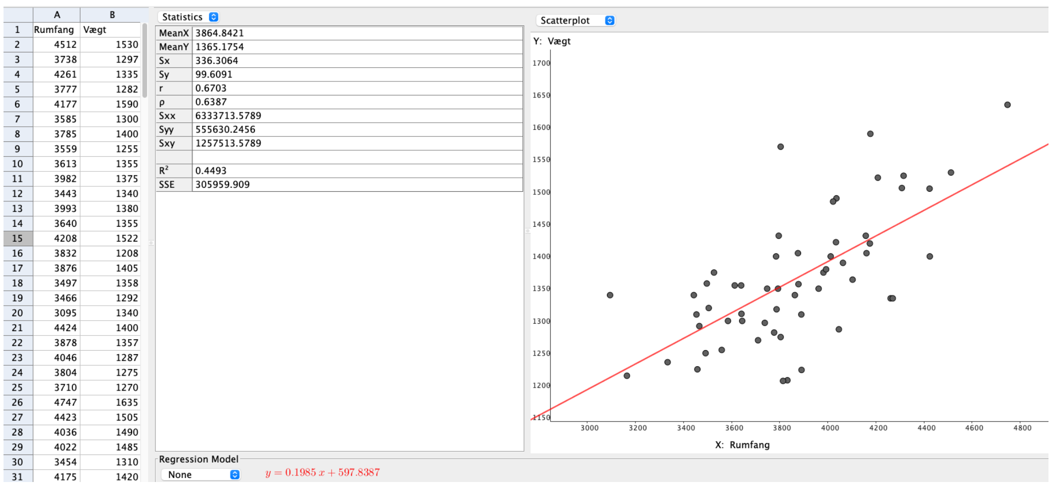
$$f(x) = a \cdot x + b,$$

hvor $f(x)$ betegner vægt (målt i gram), og x betegner rumfang (målt i cm^3).

Løsning:

a. For at bestemme tallene a og b, laver vi en lineær regression på tabellens data, hvilket ses i fig. 10. Fra regressionen har vi, at

$$a = 0{,}1985 \text{ og } b = 597{,}8387.$$

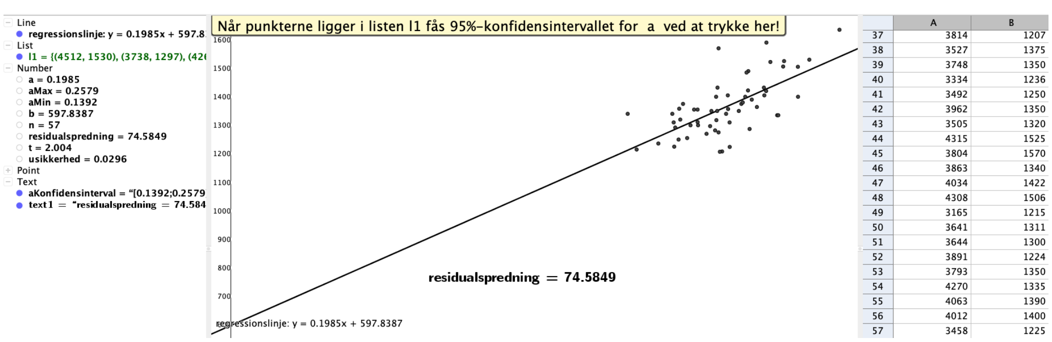


Figur 10: Lineær regression på tabellens data

b. Givet et endeligt antal residualer $r_1, r_2, \ldots, r_n$ er residualspredningen givet ved

$$s = \sqrt{\frac{r_1^2 + r_2^2 + \cdots + r_n^2}{n-2}}.$$

Denne beregner vi med GeoGebra, hvilket ses i fig. 11. Vi har beregnet residualspredningen til at være 74,5849.



Figur 11: Residualspredningen