

Transformation af data

Linearisering af data

Har vi data, vi forventer kan beskrives ved en eksponentiel model vil det være en fordel at kunne transformere datasættet, så det kan beskrives ved en lineær model. Dette gælder tilsvarende hvis vi har data, der kan beskrives ved en potensmodel. Vi kan anvende følgende sætninger, når vi skal transformere data.

Sætning 1.1. *Hvis der er en eksponentiel sammenhæng mellem to variable x og y , altså at*

$$y = b \cdot a^x,$$

så er der en lineær sammenhæng mellem x og $\ln(y)$.

Bevis. Vi har, at

$$y = b \cdot a^x,$$

så derfor gælder der, at

$$\begin{aligned}\ln(y) &= \ln(b \cdot a^x) \\ &= \ln(b) + \ln(a^x) \\ &= \ln(b) + x \ln(a).\end{aligned}\tag{1.1}$$

■

Sætning 1.2. *Hvis der er en potenssammenhæng mellem to variable x og y , altså at*

$$y = b \cdot x^a,$$

så er der en lineær sammenhæng mellem $\ln(x)$ og $\ln(y)$.

Bevis. Vi har, at

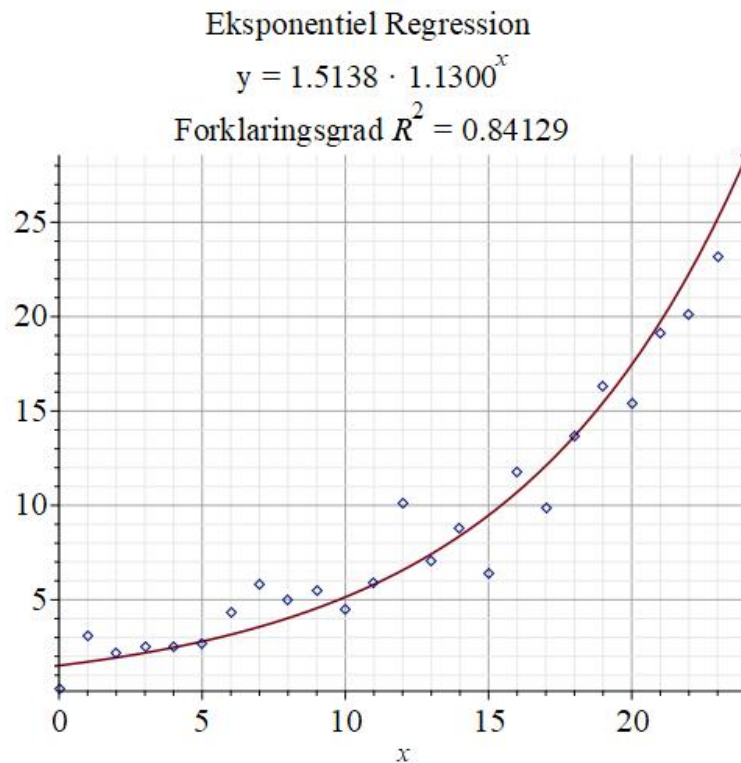
$$y = b \cdot x^a.$$

Som før fås nu, at

$$\begin{aligned}\ln(y) &= \ln(b \cdot x^a) \\ &= \ln(b) + \ln(x^a) \\ &= \ln(b) + a \ln(x).\end{aligned}$$

■

Eksempel 1.3. Vi har følgende [datasæt](#), der beskriver sammenhængen mellem forløbne timer og antallet af bakterier i en opløsning. Vi udfører eksponentiel regression i Maple og får regressionslinjen, der kan ses af Fig. 1.

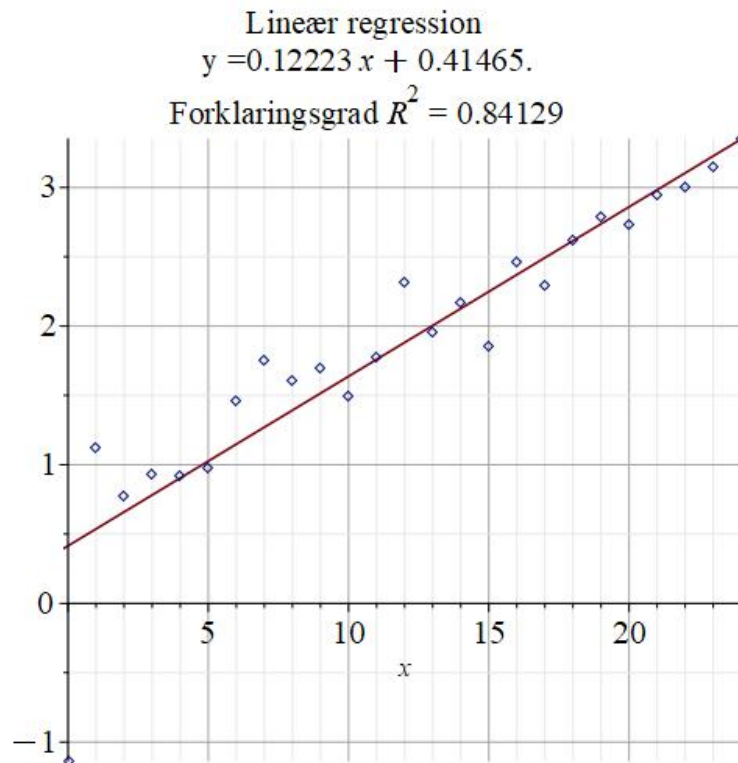


Figur 1: Eksponentiel regression på bakteriedata

Vi får følgende sammenhæng mellem antallet af bakterier y (i mia.) og forløbne timer x .

$$y = 1.5138 \cdot 1.13^x.$$

Vi tager \ln af alle y -værdierne og laver i stedet lineær regression. Dette giver regressionslinjen, der kan ses af Fig. 2. At transformere alle y -værdierne ved at tage logaritmen af dem giver os et *enkeltlogaritmisk koordinatsystem*.



Figur 2: Lineær regression i enkeltlogaritmisk koordinatsystem

Den lineære model lyder

$$\ln(y) = 0.12223x + 0.41465.$$

Ønsker vi at bestemme den eksponentielle sammenhæng ud fra dette data, så skal vi transformere koefficienterne tilbage som det fremgår af (1.1). Vi får derfor, at

$$b = e^{0.41465} = 1.5138$$

og

$$a = e^{0.12223} = 1.1300,$$

og den eksponentielle sammenhæng mellem den forløbne tid x og antallet af bakterier y er derfor

$$y = 1.5138 \cdot 1.13^x,$$

hvilket er nøjagtigt hvad vi bestemte ved den eksponentielle regression. Der er dog værd at notere, at det formentlig er præcist sådan at Maple laver eksponentiel regression, og derfor er det heller ikke mærkeligt, at vi opnår det samme resultat.

Opgave 1

I [dette datasæt](#) fremgår sammenhængen mellem antal forløbne minutter x og $\ln(y)$, hvor y er antallet af bakterier i en bestemt opløsning. Det antages, at der er lineær sammenhæng mellem x og $\ln(y)$.

- i) Brug datasættet til at bestemme en sammenhæng

$$\ln(y) = ax + b.$$

- ii) Afgør, om residualerne er normalfordelte og brug dette til at afgøre, om modellen er god.
- iii) Brug din model til at bestemme en ligning for den eksponentielle sammenhæng

$$y = \beta \cdot \alpha^x.$$

- iv) Brug din model til at afgøre, hvornår der vil være 1 bio. bakterier i opløsningen.

Opgave 2

Man har målt 10 forskellige bremselængder for 10 forskellige biler med omtrent samme vægt ved tilfældigt udvalgte hastigheder. I [dette datasæt](#) kan $\ln(x)$ og $\ln(y)$ findes, hvor x er hastigheden af bilen (i m/s) og y er bremselængden (i meter). Vi antager, at der er en potenssammenhæng mellem x og y .

- i) Brug datasættet til at bestemme en sammenhæng

$$\ln(y) = \ln(b) + a \ln(x).$$

- ii) Afgør, om residualerne er normalfordelte og brug dette til at afgøre, om modellen er god.
- iii) Brug din model til at bestemme en ligning for potenssammenhængen

$$y = \beta \cdot x^\alpha.$$

- iv) Brug din model til at bestemme bremselængden for en bil, der kører 100m/s.