ISBN 9788770668781

Projekt 9.7 Ikke-lineær regression

De fleste modeller er naturligvis ikke-lineære. Men det lineære tilfælde rummer så mange tekniske fordele, at vi typisk forsøger at transformere andre modeller over i lineære modeller. Vi har i HEM2, kapitel 4 omtalt dette nærmere. Det er en klassisk metode, når man undersøger eksponentielle eller potenssammenhænge. Vi vil her illustrere sådanne transformationer med to typiske eksempler fra andre modeller: Forskudt eksponentiel vækst $y = b \cdot \mathrm{e}^{-k \cdot x} + c$ og logistisk vækst $y = \dfrac{M}{1 + c \cdot \mathrm{e}^{-b \cdot x}}$.

Eksempel 1. Forskudt eksponentiel vækst

I den forskudte eksponentielle model er *b* og *c* "lineære parametre". Så for en fast værdi af *k*, kan vi bestemme *b* og *c* ved en lineær regression. Det gør det forholdsvis nemt at indføre en skyder for *k* og se grafisk/numerisk, hvornår den lineære regressionsmodel har den højeste forklaringsgrad og dermed den mindste sum af kvadratiske afvigelser. Det er altså forholdsvis nemt selv at finde parametrene *b*, *c* og *k* ved hjælp af mindste kvadraters metode, fordi der kun er én ikke-lineær parameter.

Øvelse 1. Newtons afkølingslov

Vi vil undersøge følgende data for afkøling af en kop the:

Tid i minutter	0	5	10	15	20	25	30
Temperatur i °C	88.0	85.0	82.2	79.5	77.0	74.5	72.2

a) Afbild **temperaturen** som funktion af **tiden**. Indfør en skyder for den ikke-lineære parameter *k*, hvor *k* løber fra 0 til 0.025 i trin af 0.0001.

b) Indfør hjælpevariablen $x\_\mathrm{var} = \exp(-k \cdot \boldsymbol{tid})$ og udfør en lineær regression af **temperaturen** som funktion af hjælpevariablen **x_var**.

c) Afbild grafen for den tilhørende modelfunktion: $f(t) = a \cdot \exp(-k \cdot t) + b$, hvor *a* og *b* hentes fra den lineære regression, i samme diagram som punktplottet for dine datapunkter.

d) Bestem nu, ved at trække i skyderen, den værdi af den ikke-lineære parameter *k*, der fører til den mindste sum af de kvadratiske afvigelser.

e) Hvad var stuetemperaturen?

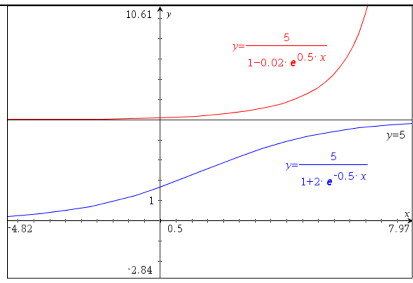
Eksempel 2. Logistisk regression

Modellering af data med den logistiske vækstmodel indeholder nogle større udfordringer. Det er slet ikke sikkert, at der er nogen løsning, og hvis der er løsninger, er det ikke sikkert, at løsningen er entydig. Værktøjsprogrammer har en kommando, der hedder logistisk regression, men man kan blive slemt skuffet ved en ukritisk anvendelse af denne. Meget afhænger af kvaliteten af data:

a) Data skal være tydeligt *voksende*.

b) Man skal tydeligt kunne skelne den logistiske vækstmodel fra fx den lineære og den eksponentielle vækstmodel.

c) Man skal kunne skelne *standardmodellen* for en logistisk vækst, hvor alle konstanterne er positive, og hvor grafen vokser op mod den vandrette asymptote $y = M$, fra den *alternative* logistiske vækst med *c* og *b* negative, hvor grafen vokser op fra en vandet asymptote $y = M$ og i stedet nærmer sig en lodret asymptote.



Logistiske funktioner: Grafen for en standardmodel for logistisk vækst (blå) sammen med grafen for et eksempel på den alternative model (rød)

ISBN 9788770668781

Øvelse 2. Beregning af en logistisk graf gennem tre punkter
Et værktøjsprogram beregner en logistisk regression ved at arbejde sig trin for trin gennem datamaterialet, men dette kræver et passende udgangspunkt. Og et sådant kræver typisk, at man kan finde den logistiske funktions graf gennem tre punkter. Kan det løses, så kan det løses i hånden. Du kan ***her*** **(Mangler – På vej)** prøve kræfter med denne udfordring

Eksempel 3. reciproktransformationen

Hvis man skal finde ligningen for en logistisk funktions graf gennem et udvalgt datasæt, kan man tilsvarende forsøge sig med en passende transformation af data. To typiske transformationer, der kan kaste lys over den logistiske vækst, er *reciproktransformationen* $\frac{1}{y}$ og *den logaritmiske transformation* $\ln(y)$. Vi vil illustrere metoderne ved hjælp af de klassiske data for en solsikkeplantes vækst, som vi allerede undersøgte i B-bogens kapitel 6.

Dag	7	14	21	28	35	42	49	56	63	70	77	84
Højde i cm	17.93	36.36	67.76	98.10	131.00	169.5	205.5	228.3	247.1	250.5	253.8	254.5

Øvelse 3 Reciproktransformation

a) Gør rent teoretisk, dvs. ud fra formlerne, rede for, at reciproktransformationen fører en logistisk vækst over i en forskudt eksponentiel vækst.
b) Gør rent teoretisk, dvs. ud fra formlerne, rede for, at grafen for den forskudte eksponentielle vækst er aftagende og opad hul (”glad”).
c) Opret grafer for henholdsvis **højde** som funktion af **dag**, og den **reciprokke højde** som funktion af **dag**.
d) Bestem herved ved samme metode som i øvelse 9.16 den bedste forskudte eksponentielle vækst ved mindste kvadraters metode, og giv ud fra dette et bud på den logistiske vækst.
e) Sammenhold det med den indbyggede logistiske vækst i dit værktøjsprogram. Konklusion?

Eksempel 4. Den logaritmiske transformation

Det er nemmest at kontrollere kvaliteten af data grafisk, og det er specielt nemt ud fra den logaritmiske transformation, der ydermere har den fordel at den giver anledning til en grafisk logistisk regression, der i princippet kan udføres i hånden, selv om vi her vil udføre den i et værktøjsprogram.

Øvelse 4. Logaritmisk transformation

a) Gør rent teoretisk, dvs. ud fra formlerne, rede for, at hvis y afhænger logistisk af x med positive modelparametre b, c og M, så vil den transformerede funktion $z(x)=\ln(y)$ have følgende pæne egenskaber:
   1. Den er voksende og nedad hul (”sur”)
   2. Grafen har en skrå asymptote med hældningskoefficient b for $x \to -\infty$, og en vandret asymptote med ligning $y=\ln(M)$ for $x \to +\infty$
   3. De to asymptoter skærer hinanden i $x=\frac{1}{b}\cdot\ln\left(\frac{c}{M}\right)$.
b) Afbild nu den logaritmiske højde ln(**højde**) som funktion af **dag**. Giv et grafisk skøn over de to asymptoter og bestem derved de tre modelparametre.
c) Afbild **højde** som funktion af **dag** sammen med den derved fundne logistiske vækstmodel. Sammenhold med den indbyggede logistiske vækst i dit værktøjsprogram. Konklusion?