OPGAVE 2

Figur 1 viser antal nye computervirus på verdensplan i de sidste 10 år. Tallene stammer fra en artikel i Berlingske Tidende. Figuren viser en brat stigning, der kan tolkes som eksponentiel vækst. Figur 2 viser derfor logaritmen til antal virus. En eksponentiel vækst på figur 1,

(1) $y \sim \alpha\beta^t$,

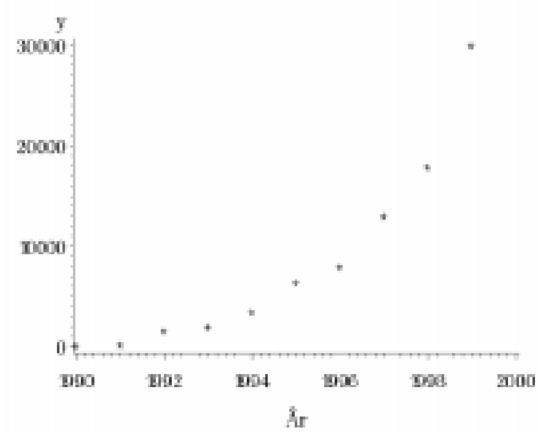
svarer på figur 2 til en lineær vækst

(2) $\log(y) \sim a + bt$

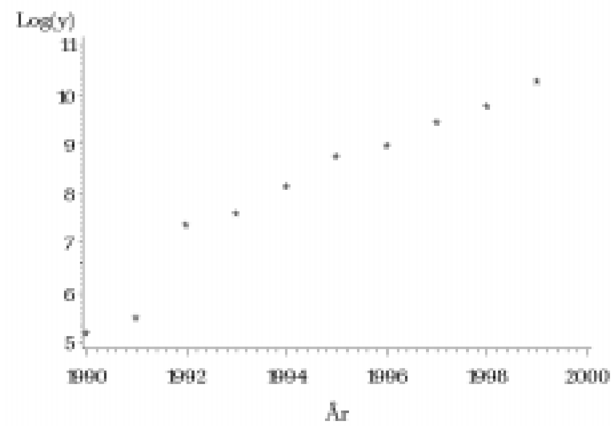
med

(3) $a = \log(\alpha)$ og $b = \log(\beta)$.

Ofte ser man imidlertid, at eksponentiel vækst er for kraftig, således at kurven på figur 2 krummer lidt nedad som $a + bt^{\lambda}$, med λ mindre end 1. Opgaven går bl.a. ud på at teste, om det er tilfældet.



Figur 1 Antal computervirus y.



Figur 2 Antal computervirus logaritmisk transformeret log(y).

Som model for eksponentiel vækst i antal computervirus y_t i det t-te år anvendes Poissonfordelingen med intensitet givet ved (1) altså $\alpha\beta^t$. Idet det antages, at der foreligger n observationer fra årene t_1 , .. , t_n , bliver punktsandsynligheden dermed

MODEL 1

$$\prod_{i=1}^{n} \frac{(\alpha\beta^{t_i})^{y_{t_i}}}{y_{t_i}!}\exp(-\alpha\beta^{t_i}) \ .$$

Antages i stedet, at væksten er svagere end eksponentiel, som beskrevet ovenfor, bliver intensiteten $\alpha\beta^{t^{\lambda}}$, så den simultane punktsandsynlighed er

MODEL 2

$$\prod_{i=1}^{n} \frac{(\alpha\beta^{t_i^{\lambda}})^{y_{t_i}}}{y_{t_i}!}\exp(-\alpha\beta^{t_i^{\lambda}}) \ .$$

For $\lambda = 1$ reducerer MODEL 2 til MODEL 1 . I SAS kørslerne sættes $t_1 = 1$, .. , $t_{10} = 10$ svarende til årene t = 1990 - 1999.

I MODEL 1 estimeres de to parametre ved et kald af PROC NLP og de tilhørende standardafvigelser til

Parameter	Estimat	Standardafvigelse
α	352.6	5.4
β	1.56	0.003

Kovariansmatricen for de to estimatorer beregnes til

$$-\mathbf{Q}^{-1} = \begin{pmatrix} 29.1 & -0.015 \\ -0.015 & 0.00000777 \end{pmatrix}$$

I MODEL 2 estimeres de tre parametre og de tilsvarende standardafvigelser, atter ved et kald af PROC NLP, til

Parameter	Estimat	Standardafvigelse
α	27.8	5.1
β	7.99	1.12
λ	0.52	0.02

a)

Parametrene a og b i (2), estimeres i MODEL 1, jf. (3) ved

$\hat{a} = \log(352.6) = 5.87$ og $\hat{b} = \log(1.56) = 0.44$.

Angiv den approksimative variansmatrix for disse estimatorer for a og b.

b)

Vis at scorevektoren i MODEL 2 er givet ved

$$\frac{\partial}{\partial \alpha} \log L = \sum_{i=1}^{n} \frac{y_i}{\alpha} - \sum_{i=1}^{n} \beta^{t_i^{\lambda}}$$

$$\frac{\partial}{\partial \beta} \log L = \sum_{i=1}^{n} \frac{y_i t_i^{\lambda}}{\beta} - \sum_{i=1}^{n} \alpha \beta^{t_i^{\lambda} - 1} t_i^{\lambda}$$

og

$$\frac{\partial}{\partial \lambda} \log L = \sum_{i=1}^{n} \left(y_i - \alpha \beta^{t_i^{\lambda}} \right) t_i^{\lambda} \log(\beta) \log(t_i)$$

Husk at $\frac{d}{d\gamma} x^{\gamma} = \log(x) x^{\gamma}$.

I MODEL 1 er maksimum for loglikelihoodfunktionen - 791.1 , mens den i MODEL 2 er - 463.2.

c)

Test hypotesen λ = 1 ved et Wald test.

d)

Beregn Likelihood Ratio teststørrelsen for hypotesen λ = 1.

Idet både middelværdien og variansen i en Poissonfordeling er lig med dens intensitet, er i MODEL 2

$$E[y_i] = \mathrm{var}(y_i) = \alpha\beta^{t_i^\lambda}$$

og altså i MODEL 1

$$E[y_i] = \mathrm{var}(y_i) = \alpha\beta^{t_i}$$

e)

Vis at middelværdien af $\frac{\partial}{\partial\lambda}\log L$ med λ = 1 indsat er 0 (nul) under antagelse af, at MODEL 1 er sand. Vis ligeledes at variansen af $\frac{\partial}{\partial\lambda}\log L$, atter under antagelse af at MODEL 1 er sand og med λ = 1 indsat, er

$$\mathrm{var}\left(\frac{\partial}{\partial\lambda}\log L\right) = \sum_{i=1}^{n} \alpha\beta^{t_i} t_i^2 \log(\beta)^2 \log(t_i)^2 \ .$$

Idet parametrene estimeret i MODEL 1 og λ = 1 indsættes i $\frac{\partial}{\partial\lambda}\log L$, bliver værdien

(4)
$$\frac{\partial}{\partial\lambda}\log L = -167 \ ,$$

mens variansen beregnes til

(5)

$$\mathrm{var}\left(\frac{\partial}{\partial\lambda}\log L\right) = \sum_{i=1}^{n} \alpha\beta^{t_i}\log(\beta)^2\log(t_i)^2 = 5788995 \ .$$

Indsættes estimaterne for parametrene i MODEL 1 , α = 352.6 og β = 1.56 samt λ = 1, i scorevektoren for MODEL 2, fås dermed numerisk

$$\mathbf{q} = (\ 0 \ \ 0 \ 167\ )\ .$$

f)

Beregn ud fra tallene (4) og (5) et test for hypotesen λ = 1 og redegør for dette tests forbindelse til LM testet.