

Tentamen Statistische Modellen

22 december 2004



1. Beschrijf kort, maar zo volledig mogelijk, de algemene vorm van een gegeneraliseerd lineair model en geef een concreet voorbeeld van zo'n model.
2.
 - a) Geef de definitie van een (zwak) stationaire tijdreeks.
 - b) Beschrijf het MA(q) model.
 - c) Geef de autocovariantiefunctie voor een MA(2) model.
 - d) Leg uit wat er met het 'cut-off verschijnsel' wordt bedoeld en hoe dit verschijnsel kan worden uitgebuit bij het selecteren van een model voor een stationaire tijdreeks.
3. Een arts wil onderzoeken of er een verschil bestaat tussen de verwachte bloeddruk van personen in drie verschillende regio's. Ze heeft de beschikking over bloeddrukmetingen van tien aselect gekozen personen uit de verschillende gebieden. De metingen werden geanalyseerd met het software pakket *R*. De uitvoer is hieronder weergegeven. Enkele getallen zijn door vraagtekens vervangen.

```
> dat.aov<-aov(y~indic,data=dat)
> summary(dat.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
indic	2	5.1119	2.5559	?	0.05439
Residuals	?	21.2381	0.7866		
Total	29	26.3500			

- a) Beschrijf het variantie analyse model voor dit probleem
 - b) Beschrijf een toets om na te gaan of de verwachte bloeddruk voor de groepen verschilt. Vermeld hiertoe in ieder geval de nulhypothese, de toetsingsgrootte, de verdeling van de toetsingsgrootte onder de nulhypothese en voor welke waarden van de toetsingsgrootte de nulhypothese verworpen dient te worden.
 - c) Welke getallen horen op de plaatsen van de vraagtekens te staan? U mag niet verder uitgewerkte producten, machten of quotiënten in het antwoord laten staan. Beargumenteer uw keuze.
 - d) Voer de toets uit bij significantieniveau $\alpha = 0.05$ en geef de correcte formulering van de conclusie.
4. Zij

$$Y_i = f(x_i, \theta) + \epsilon_i, \quad i = 1, \dots, n$$

een niet-lineair regressiemodel waar $x_1, \dots, x_n \in \mathbb{R}^k$ vectoren van verklarende variabelen zijn, $\epsilon_1, \dots, \epsilon_n$ i.i.d. gecentreerde normale variabelen en $Y_1, \dots, Y_n$ responsvariabelen. De parametervector θ behoort toe aan de parameterruimte $\mathbb{R}^p$.

- a) Geef de definitie van de kleinste kwadratenschatter $\hat{\theta}$ in dit model.
 - b) Zij $\hat{\sigma}^2 (\hat{V}^T \hat{V})^{-1}$ een schatter voor de covariantiematrix van de schatter $\hat{\theta}$. Geef op basis hiervan (en van de schatter $\hat{\theta}$) een, bij benadering, 95% betrouwbaarheidsinterval voor de componenten θ_j van de parametervector θ volgens de klassieke theorie.
 - c) Beschrijf de stappen die moeten worden gezet om tot een bootstrap betrouwbaarheidsinterval te komen voor θ_j .
5. Hieronder worden enkele praktijksituaties geschetst. Geef bij deze situaties één of meerdere geschikte modellen, waarbij gekozen mag worden uit lineair regressiemodel (SDA), het variantie-analyse model, niet-lineair model, gegeneraliseerd lineair model en een tijdreeksmodel. Beargumenteer hierbij de keuze en beschrijf de bijbehorende modellen zo volledig mogelijk.
- a) Een verzekeringsmaatschappij is geïnteresseerd in de samenhang tussen het aantal schadeclaims dat wordt ingediend door gasinstallatie bedrijven, afhankelijk van het aantal werknemers, de omzet en de regio waar wordt gewerkt. Het land is opgedeeld in drie regio's
 - b) Voor Nederlandse kampioenschappen schaatsen op de ijsbaan van Heerenveen worden altijd oranje petjes met daarop het jaar gedrukt, verkocht. Het is voor de organisatie van belang te weten hoeveel petjes er naar verwachting verkocht zullen worden. Van de afgelopen 14 Nederlandse kampioenschappen zijn hiervoor gegevens bekend.

Normering

1	2	3	4	5
5	a:3	a:2	a:3	a:3
	b:3	b:4	b:3	b:3
	c:2	c:2	c:3	
	d:2	d:1		

Cijfer tentamen: $1+3(\text{aantal punten})/13$