

Opgave 1.

- a. Beschrijf kort, maar zo volledig mogelijk, een logistisch regressiemodel.
- b. In principe zijn de verklarende variabelen in een logistisch regressiemodel continue variabelen gemeten op een ratio schaal. Geef aan hoe een nominale variabele met k categorieën ($k > 2$) als verklarende variabele in een logistisch regressiemodel kan worden opgenomen.

Opgave 2.

Een arts wil onderzoeken of er een significant verschil bestaat tussen de verwachte bloeddruk van personen in drie verschillende regionale gebieden. Ze heeft de beschikking over bloeddrukmetingen van een aantal personen uit de verschillende regio's: 8 uit regio I, 10 uit regio II en 12 uit regio III. De metingen werden geanalyseerd met het software pakket S-plus. De uitvoer is hieronder weergegeven. Enkele getallen zijn door vraagtekens vervangen.

Source	Sum of squares	DF	Mean square	Test statistic (T)	Sign. of T (p-value)
Between groups	1103.5	2	551.8	?	0.86
Within groups	98674.2	?	3654.6		
Total	99777.7	29			

- a. Beschrijf een 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 elk geval de nulhypothese, de toetsingsgrootheid, de verdeling van de toetsingsgrootheid onder de nulhypothese en voor welke waarden van de toetsingsgrootheid de nulhypothese verworpen dient te worden.
- c. Welke getallen horen op de plaatsen van de vraagtekens te staan? Beargumenteer je

antwoord.

d. Voer de toets uit met onbetrouwbaarheid $\alpha = 0.05$ en geef je conclusie.

Opgave 3.

In Figuur 1 op de volgende bladzijde is voor de periode van 1 januari 1993 tot en met 1 december 2002 een tijdreeks van maandelijkse gegevens weergegeven.

a. Geef een geschikt model voor deze gegevens. Vermeld ook welke aannamen hierbij worden gedaan.

b. Beschrijf een methode om voor de tijdreeks in Figuur 1 het seizoenseffect te schatten.

Opgave 4.

Figuur 2 toont twee series metingen. Serie 1, weergegeven door o, werd verkregen onder koudere omstandigheden dan Serie 2, weergegeven door x. Voor elke serie kunnen de data worden beschreven met een niet-lineair model van de vorm

$$Y_i = \theta_1 + \frac{\theta_2 x}{\theta_3 + x} + \epsilon_i, \quad i = 1, \dots, 20,$$

waarbij $\epsilon_1, \dots, \epsilon_{20}$ onderling onafhankelijke meetfouten zijn die normaal verdeeld worden verondersteld met verwachting 0 en variantie σ^2 . De waarden van $\theta_1, \theta_2, \theta_3$ en σ^2 zijn onbekend. Het vermoeden is, dat de waarde van θ_2 wordt beïnvloed door de temperatuur en dat de andere parameterwaarden temperatuurgevoelig zijn.

a. Hoe is de kleinste kwadratenschatter $\hat{\theta}$ van $\theta = (\theta_1, \theta_2, \theta_3)^T$ gedefinieerd?

b. Zij $\widehat{\text{Cov}}(\hat{\theta}) = \hat{\sigma}^2(\hat{V}^T \hat{V})^{-1}$ een schatter voor de covariantiematrix van $\hat{\theta}$. Geef een, bij benadering, 95% betrouwbaarheidsinterval voor θ_l , $l = 1, 2, 3$.

c. In Tabel 1 staan de parameterschattingen met bijbehorende 95% betrouwbaarheidsintervallen voor de twee series. Wat concludeer je op grond van deze resultaten met betrekking tot het vermoeden omtrent de temperatuurgevoeligheid van θ_1, θ_2 en θ_3 ?

d. Beschrijf hoe de nulhypothese dat de waarden van θ_2 voor de twee series gelijk zijn, getoetst kan worden.

parameter	Serie 1		Serie 2	
	schatting	interval	schatting	interval
θ_1	5.54	(5.414, 5.657)	5.52	(5.383, 5.648)
θ_2	2.94	(2.813, 3.068)	3.11	(2.967, 3.244)
θ_3	11.96	(10.016, 13.899)	10.96	(9.154, 12.762)

Tabel 1.

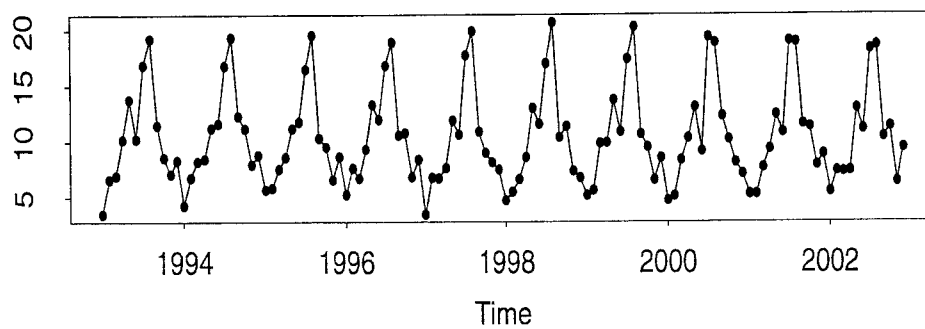


Figure 1: Een tijdreeks.

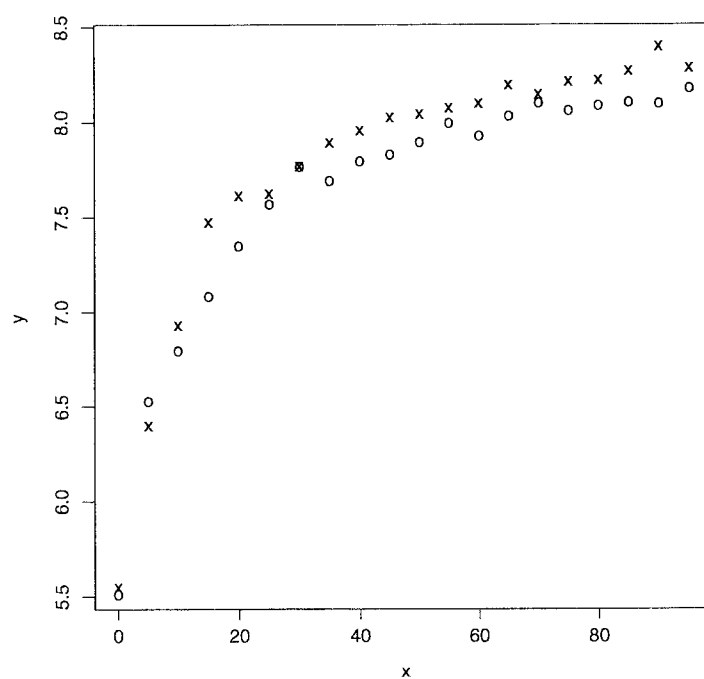


Figure 2: Metingen Serie 1 ($\circ$), Serie 2 ($\times$).

Opgave 5.

Hieronder worden enkele praktijksituaties geschetst. Kies voor elke situatie één of meer geschikte modellen. Kies hierbij uit: lineair regressiemodel, variantie-analyse model, niet-lineair model, gegeneraliseerd lineair model, tijdreeksmodel. Beargumenteer je antwoord. Beschrijf het gekozen model zo volledig mogelijk. Vermeld hierbij in elk geval wat de respons- en verklarende variabelen zijn.

- a. Een fabriek van glazen microchips wil onderzoeken of de kwaliteit van de chips afhangt van de omgevingstemperatuur tijdens het productieproces, van de afkoe-lingstijd van het glas tijdens de produktie en van de viscositeit van de gebruikte grondstof. Hiertoe worden 30 partijen van 100 chips gekeurd. De chips van eenzelfde partij hebben dezelfde waarden voor de verklarende variabelen. Die van verschillende partijen kunnen andere waarden hebben. Voor elke partij wordt genoteerd hoeveel van de chips voldoen aan de standaard kwaliteitseisen.
- b. In een onderzoek bij diabetes patiënten werd onderzocht of de volgende factoren verband houden met de mate van controle die de patient heeft over zijn of haar glucosegehalte: kennis omtrent de ziekte gemeten op een schaal van 0 tot 100, atti-tude ten aanzien van de ziekte gemeten als hoeveelheid fatalisme op een schaal van 0 tot 100, duur van de ziekte in maanden en aantal jaren scholing. Van elk van de patiënten werd tevens eenmalig het glucosegehalte gemeten.
- c. Bij een verzekeringsmaatschappij komen op de werkdagen maandag tot en met vrij-dag honderden telefoontjes binnen. Er zijn klachten dat de gemiddelde afhandeltijd per telefoontje, en daarmee de wachttijd voor een beller, op maandagochtend langer is dan op de andere dagdelen. Hoewel het vermoeden bestaat, dat niet een lan-gere afhandeltijd, maar een groter aantal telefoontjes de langere wachttijd op de maandagochtend veroorzaakt, wordt toch onderzocht of het inderdaad zo is dat de medewerkers op maandagochtend trager werken. Drie maanden lang wordt de afhan-deltijd van elk telefoontje dat op maandag tot en met vrijdag binnenkomt, genoteerd.

Normering:

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

Eindcijfer tentamen = totaal/4 + 1