

Opgave 1.

- a. Beschrijf kort, maar zo volledig mogelijk, een twee faktor variantie analyse model met interacties.
- b. Welke randvoorwaarden worden meestal op de parameters gelegd? Waarom worden randvoorwaarden op de parameters gelegd?

Opgave 2.

We beschikken over twee niet-lineaire regressiefuncties om onze data mee te beschrijven:

$$f(x, \theta) = \frac{\theta_1 \exp(\theta_2 x)}{1 + \theta_3 \exp(-x)}, \quad \theta = (\theta_1, \theta_2, \theta_3)^T,$$

en

$$f(x, \theta) = \theta_1 \exp(\theta_2 x), \quad \theta = (\theta_1, \theta_2)^T.$$

- a. Geef de algemene vorm van een niet-lineair regressiemodel.
- b. Beschrijf een toets om te bepalen welk van de twee regressiefuncties de data het best beschrijft. Welke aanname(n) heb je hierbij gemaakt?

Opgave 3.

In figuur 1 is een tijdreeks getekend op basis van maandelijkse gegevens uit de jaren 1990 tot 1998.

- a. Beschrijf kort een algemeen model voor deze tijdreeks.
- b. Beschrijf een methode om voor de tijdreeks in figuur 1 de trend en het seizoenseffect te schatten.
- c. Na het verwijderen van de trend en het seizoenseffect, houden we een stationaire tijdreeks over. Voor deze tijdreeks schatten we de autocorrelatiefunctie en de partiële autocorrelatiefunctie, die zijn weergegeven in figuur 2. Welk van de modellen AR, MA en ARMA past op grond van deze plaatjes het best bij deze tijdreeks. Schat ook de orde van het proces. Beargumenteer je keuze.

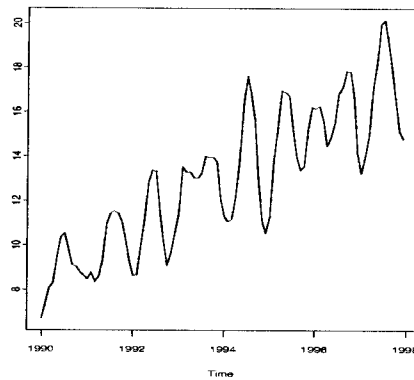


Figure 1: Tijdreeks.

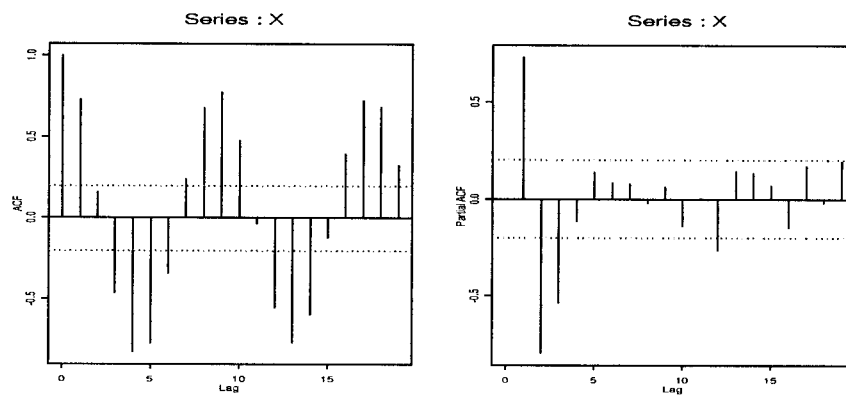


Figure 2: Geschatte autocorrelatie functie (links) en partiële autocorrelatie functie (rechts).

Opgave 4.

Bij onderzoek naar de mate van giftigheid van een bepaalde stof, wordt een test met kevers gedaan. Een kever wordt blootgesteld aan een bepaalde concentratie van de giftige stof. Na vijf minuten wordt gekeken of de kever nog in leven is. Dit experiment wordt bij 481 kevers uitgevoerd; elke keer met een andere concentratie giftige stof. Het doel van het onderzoek is te bepalen hoe de concentratie van de giftige stof de kans dat een kever sterft beïnvloedt.

a. Beschrijf een logistisch regressie model voor de 481 waarnemingen. (Gebruik hierbij de definitie: $Y_i = 1$ als de i -de kever het niet overleefd heeft en $Y_i = 0$ als de i -de kever het wel overleefd heeft.)

b. Het lineaire deel in het logistische regressie model is van de vorm $\eta_i = \beta_0 + \beta_1 x_i$ met x_i de gebruikte concentratie bij kever i . De gegevens in het experiment worden met het statistische pakket S-plus geanalyseerd. S-plus geeft de volgende uitvoer.

	Value	Std. Error	t value
Intercept	-59.21	5.03	-11.77
concentratie	33.42	2.83	11.82

Toets of de concentratie van de giftige stof invloed heeft op de kans dat de kever komt te overlijden. Beschrijf de toets kort, maar nauwkeurig. Neem onbetrouwbaarheid $\alpha = 0.05$. (Het 0.975-kwantiel van de t_{479} -verdeling is gelijk aan 1.96.)

c. Onderzoekers willen de concentratie schatten waarbij 90% van de kevers die worden blootgesteld aan deze concentratie sterft. Geef aan hoe je deze concentratie kunt schatten op basis van bovenstaande gegevens.

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.

- Bij een onderzoek naar geslachts-discriminatie binnen een bedrijf, wordt van elke medewerker gegevens verzameld van het inkomen (gerekend als de persoon 100% zou werken), geslacht, leeftijd en functie (laag, midden, hoog). Het idee leeft dat mannen meer verdienen dan vrouwen met dezelfde leeftijd en met dezelfde functie; dit moet onderzocht worden.
- Weerdeskundigen willen voorspellingen doen over de temperatuur in Nederland voor de komende 50 jaar. Zij hebben de beschikking over temperatuurgegevens in Nederland van de afgelopen 200 jaar.
- Een viskwekerij doet onderzoek naar de invloed van drie soorten voeding op de groeisnelheid van twee vissoorten. Er worden 12 aquaria ingericht; 6 aquaria met elk 10 vissen van vissoort A en 6 aquaria met elk 10 vissen van vissoort B. Alle vissen hebben hetzelfde (begin)gewicht. Per aquarium wordt gekozen met welke voeding de

vissen gevoederd worden. Dit wordt zo gekozen dat steeds elke combinatie vissoort-voedingssoort twee maal voorkomt. Na enige tijd wordt per aquarium het totale gewicht aan vis gemeten. Het doel is te bepalen welke voedingsmiddel en welke vissoort het meeste aantal kilo vis oplevert.

Normering:

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

$$\text{Eindcijfer tentamen} = \text{totaal}/4 + 1$$