



Opgave 1.

- a. Beschrijf kort, maar zo volledig mogelijk, een niet-lineair regressiemodel.
- b. Beschrijf, ook kort, een methode waarmee de parameters van een niet-lineair regressiemodel kunnen worden geschat.

Opgave 2.

Veronderstel dat je een tijdreeksanalyse wilt uitvoeren op een gegeven data set $X_1, \dots, X_n$. Uit welke stappen bestaat je analyse? Leg uit hoe en waarom je deze stappen doet.

Opgave 3.

Een kunstmestbedrijf zoekt naar de beste samenstelling voor een van haar producten. Hiertoe wordt het produkt in vier samenstellingen uitgetoetst op drie soorten suikerbieten. Dit gebeurt voor elk combinatie vier keer op stukken grond van gelijke samenstelling en grootte. De opbrengst van de achtenveertig stukken grond wordt genoteerd. Het doel is een zo hoog mogelijke opbrengst te verkrijgen.

- a. Geef een geschikt model voor deze situatie. Vermeld ook welke aannamen hierbij worden gedaan.
- b. Beschrijf een toets om na te gaan of een additief model voldoet. Geef hierbij de nulhypothese, de toetsingsgrootte, de verdeling van de toetsingsgrootte onder de nulhypothese en voor welke waarden van de toetsingsgrootte de toets verworpen moet worden.
- c. De waarnemingen worden geanalyseerd met het software pakket S-plus. De uitvoer is hieronder weergegeven. Welk(e) getal(len) uit de tabel zou je gebruiken om de onder b beschreven toets uit te voeren? Interpreteer de resultaten van deze toets en vermeld tevens welke conclusies nog meer uit de tabel kunnen worden getrokken.

Source	Sum of squares	DF	Mean square	F Value	Pr(F)
biet	185.47	2	92.73	2.16	0.129
mest	449.37	3	149.79	3.49	0.025
biet:mest	573.52	6	95.58	2.23	0.062
Residuals	1541.59	36	42.82		

Opgave 4.

Bij een studie naar de mate van giftigheid van een chemische stof wordt een test gedaan bij 50 proefdieren. Hierbij wordt een proefdier blootgesteld aan een bepaalde concentratie van de chemische stof. Na een vastgestelde tijd wordt gekeken of het proefdier nog in leven is. De test wordt bij elk dier met een andere concentratie uitgevoerd. Doel van het onderzoek is te bepalen hoe de concentratie van de stof de kans dat een dier sterft, beïnvloedt.

- a. Beschrijf een logistisch regressiemodel voor de waarnemingen, waarbij het lineaire deel van het model voor het i -de proefdier wordt gegeven door $\eta_i = \beta_0 + \beta_1 x_i$ met x_i de gebruikte concentratie bij het i -de proefdier.
- b. Voor het model uit a worden de resultaten van de studie met S-plus geanalyseerd. S-plus geeft de volgende uitvoer:

	Value	Std. Error	t value
Intercept	3.11	7.58	0.41
concentratie	-50.66	151.32	-0.33

Toets of de concentratie van de stof invloed heeft op de kans dat een proefdier overlijdt. Beschrijf de toets kort, maar nauwkeurig. Stel een 95% betrouwbaarheidsinterval op voor β_1 . (Het 0.0975-kwantiel van de t_{48} -verdeling is gelijk aan 2.01.)

- c. De onderzoekers willen de minimale concentratie schatten waarbij 90% van de dieren die worden blootgesteld aan de chemische stof, sterft. Geef aan hoe je deze concentratie kunt schatten op basis van bovenstaande gegevens.
- d. Met het oog op efficiëntie besluit men om bij volgende proeven slechts 10 verschillende concentraties te gebruiken, elk op 5 proefdieren. Bij deze proefopzet kunnen de data wederom geanalyseerd worden volgens het model in a. Als alternatief kan bij elke concentratie bepaald worden hoeveel dieren het experiment hebben overleefd. Dit levert 10, in plaats van 50 waarnemingen. Beschrijf een nieuw gegeneraliseerd lineair model voor deze 10 waarnemingen. Denk je dat de conclusies over de mate van invloed van de concentratie van de chemische stof op de kans op overlijden verschillend is bij analyse onder de twee modellen? Waarom (niet)?

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 regressiemodel, 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 tandarts doet onderzoek naar de leeftijd waarop verstandskiezen doorbreken. Daartoe noteert zij een halfjaar lang bij alle patiënten die voor controle komen, of de verstandskiezen al zijn doorgebroken en de leeftijd van de patiënt. De tandarts is geïnteresseerd in de fractie van mensen bij wie de verstandskiezen zijn doorgebroken als functie van leeftijd.
- b. Bij een onderzoek naar drugsgebruik onder jongeren worden 1000 mensen tussen de 12 en 16 jaar ondervraagd. Naast het antwoord op de vraag of en, zo ja, hoeveel van een bepaald verdovend middel (in grammen per maand) ze gebruiken, wordt de leeftijd, het geslacht en de woonplaats genoteerd. In eerste instantie is de onderzoeker geïnteresseerd in de afhankelijkheid tussen het al dan niet gebruiken van het verdovende middel en de leeftijd, het geslacht en de grootte van de woonplaats. De woonplaatsen worden hierbij op grootte ingedeeld (klein, middel, groot).
- c. Tot de schrik van de onderzoeker uit onderdeel b. zijn er meer 12-jarige druggebruikers dan hij had verwacht. Dit is een reden om de groep met 12-jarige druggebruikers nader te bestuderen. Met name wordt nu onderzocht hoe de hoeveelheid van het verdovende middel die een 12-jarigen per maand tot zich neemt, afhangt van het geslacht en de grootte van de woonplaats. (Opnieuw is de grootte van de woonplaats gedefinieerd als in onderdeel b.)

Normering:

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

Eindcijfer tentamen = totaal/4 + 1