

**Opgave 1.**

- Beschrijf kort, maar zo volledig mogelijk, een niet-lineair regressie model.
- Beschrijf kort een methode waarmee de parameters in een niet-lineair regressie model geschat kunnen worden.

**Opgave 2.**

Bij onderzoek naar de genezende werking van een medicijn wordt een test met muizen uitgevoerd. In totaal worden 8 verschillende concentraties van het medicijn getest. Voor elk van de 8 verschillende concentraties krijgt een aantal zieke muizen het medicijn met deze concentratie twee maal daags toegediend. Na vijf dagen wordt het aantal muizen dat genezen is, geteld. Een muis kan slechts één maal deelnemen aan het experiment. In totaal deden 481 muizen aan de test mee. Het doel van het onderzoek is te bepalen hoe de concentratie van het medicijn de kans dat een muis geneest beïnvloedt. De gegevens staan in onderstaande tabel.

| concentratie | aantal geteste<br>muizen | aantal muizen<br>genezen |
|--------------|--------------------------|--------------------------|
| 1.69         | 59                       | 6                        |
| 1.72         | 60                       | 13                       |
| 1.76         | 62                       | 18                       |
| 1.78         | 56                       | 28                       |
| 1.81         | 63                       | 52                       |
| 1.84         | 59                       | 52                       |
| 1.86         | 62                       | 61                       |
| 1.88         | 60                       | 60                       |

- Beschrijf een gegeneraliseerd lineair model voor dit experiment. Geef duidelijk aan wat de respons variabele is en welke link functie je gebruikt.
- Beschrijf hoe je kunt toetsen of de concentratie van het medicijn invloed heeft op de kans dat de muis geneest. Geef de nul en de alternatieve hypothese, de toetsingsgrootheid,

de verdeling van de toetsingsgrootte onder de nulhypothese en voor welke waarden van de toetsingsgrootte de toets verworpen moet worden.

c. De gevonden gegevens zijn met S-plus geanalyseerd. De uitvoer staat in onderstaande tabel. Toets of de concentratie van het medicijn invloed heeft op de kans dat de muis

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

geneest. Neem onbetrouwbaarheid  $\alpha = 0.05$ . (Het 0.975-kwantiel van de  $t_{479}$ -verdeling is gelijk aan 1.96.)

### Opgave 3.

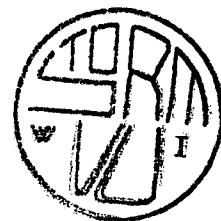
- Wanneer is een tijdreeks  $\{X_t\}$  (zwak) stationair?
- Veronderstel dat je een tijdreeksanalyse uit moet voeren op een gegeven data set  $X_1, \dots, X_n$ . Leg duidelijk uit welke stappen je gaat uitvoeren.

### Opgave 4.

Een aardappelteler doet onderzoek naar de invloed van drie soorten kunstmest op de groeisnelheid van drie aardappelsoorten. Elke combinatie wordt vier keer uitgetoet op stukken grond van gelijke grootte. De opbrengst voor alle 36 stukken grond wordt genoteerd. Het doel is een zo hoog mogelijke opbrengst te krijgen.

De waarnemingen worden geanalyseerd met S-plus. De uitvoer is hieronder weergegeven.

| Source         | sum of squares | DF | Mean square | Test Statistic (T) | Sig. of T (p-value) |
|----------------|----------------|----|-------------|--------------------|---------------------|
| Main Effects   |                |    |             |                    |                     |
| aardappelsoort | 29.67          | 2  | 14.83       | 3.749              | 0.037               |
| mestsoort      | 5.05           | 2  | 2.53        | 0.638              | 0.536               |
| Interaction    |                |    |             |                    |                     |
| aardappel-mest | 61.35          | 4  | 15.34       | 3.89               | 0.0129              |
| Residual       | 106.81         | 27 | 3.96        |                    |                     |



- a. Beschrijf een variantie analyse model voor dit probleem.
- b. Beschrijf een toets om na te gaan of een additief model voldoet. Geef de nulhypothese, de toetsingsgrootte, de verdeling van de toetsingsgrootte onder de nulhypothese en voor welke waarden van de toetsingsgrootte de toets verworpen moet worden. Voer de toets uit met onbetrouwbaarheid  $\alpha = 0.05$ .
- c. Beschrijf een toets om na te gaan of de hoofdeffecten in het model moeten worden opgenomen. Geef de nulhypothese, de toetsingsgrootte, de verdeling van de toetsingsgrootte onder de nulhypothese en voor welke waarden van de toetsingsgrootte de toets verworpen moet worden. Voer de toets uit met onbetrouwbaarheid  $\alpha = 0.05$ .
- d. Hoe moet de aardappelteler de statistische resultaten interpreteren?

### 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 tandarts doet onderzoek naar de leeftijd waarop verstandskiezen doorbreken. Daartoe noteert hij een half jaar lang bij al zijn patiënten die langs komen voor controle of er al verstandskiezen zijn doorgebroken en de leeftijd van de patiënt. De tandarts is geïnteresseerd in de fractie van mensen bij wie verstandskiezen zijn doorgebroken als functie van leeftijd.
- b. Over enkele weken zijn de landelijke verkiezingen. Enkele politieke partijen maken zich zorgen over het aantal zetels dat zij dan zullen winnen. Deze zorgen zijn gebaseerd op de resultaten van opiniepeilingen die regelmatig gedaan worden. De opiniepeilingen geven het aantal zetels dat de politieke partijen zouden winnen als op het moment van de peiling verkiezingen gehouden zouden worden. Aan de hand van deze gegevens wil een politicus een voorspelling doen over de mate van verlies of winst van het aantal zetels van zijn partij.
- c. Een verzekeraar wil de premie van de autoverzekering laten afhangen van de leeftijd en het geslacht van de verzekerde en of hij/zij in de randstad woont. De verzekeraar heeft namelijk het vermoeden dat de frequentie waarmee de verzekerde brokken maakt en kosten declareert afhangt van deze drie factoren. Om dit te onderzoeken zijn gegevens van alle personen die de afgelopen vijf jaar bij hem een verzekering hadden lopen beschikbaar.

Normering:

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

Eindcijfer tentamen = totaal/4 + 1