

## Statistische gegevensanalyse: Examenproject

### Opgaven

1. Een onderzoeker wil meer inzicht krijgen in factoren die het groeiproces beïnvloeden van de zwarte spar, een veel voorkomende soort in Noord-Amerika die economisch belangrijk is voor de bosbouw. Hiervoor heeft hij een experiment opgezet met 72 random geselecteerde zaailingen die willekeurig onder verschillende condities geplant worden op verschillende percelen. De twee condities die onderzocht worden, zijn het effect van toegevoegde meststoffen (fertilizer) en het effect van het verwijderen van andere planten in de omgeving van de spar (competition). Om het groeiproces te evalueren worden de hoogte en de diameter van de bomen gemeten bij de start (tijdstip 0) en 5 jaar later (tijdstip 5). De dataset in file “ZwarteSpar.csv” bevat dan ook de volgende variabelen:

**Competition:** NC = geen competitie, C = competitie

**Fertilizer:** NF: geen meststoffen, F = meststoffen

**Height0:** Hoogte bij start (in inch)

**Height5:** Hoogte na 5 jaar (in inch)

**Diameter0:** Diameter bij start (in inch)

**Diameter5:** Diameter na 5 jaar (in inch)

Op basis van de verzamelde gegevens wenst de bioloog de volgende vragen te onderzoeken:

- (a) Wat is het effect van het toevoegen van meststoffen op de hoogte van de bomen? Neemt de hoogte van de bomen hierdoor sneller toe of niet?
- (b) Heeft de aanwezigheid van andere planten in de omgeving van de spar een belangrijk effect op de toename van de diameter van de bomen of niet?

- (c) Hoe sterk is de associatie tussen de toename van de lengte en de toename van de diameter? Hoe precies kan deze associatie ingeschat worden?
- (d) Stel een model op dat de hoogte van de spar na 5 jaar verklaart aan de hand van zowel het al of niet toevoegen van meststoffen alsook het al of niet verwijderen van andere planten in de omgeving van de spar. Geef de kwaliteit van het model weer en de nauwkeurigheid van de schattingen. Hebben de beide condities een significant effect op de hoogte van de spar?
- (e) Kan bovenstaand model verbeterd worden door de hoogte bij de start in rekening te brengen?
- (f) Bepaal op basis van het beste model een nuttige inschatting van de gemiddelde hoogte van de sparren na 5 jaar onder elk van de vier mogelijke combinaties van de 2 condities.

2. De voortplanting van degenkrabben is nog steeds gedeeltelijk een mysterie. Tijdens de paringsperiode trekken de krabben naar modderige stranden waar ze paren vormen. De vrouwelijke krabben maken een nest in het zand en hun eitjes worden bevrucht door het mannetje op hun rug. Daarnaast zijn er een aantal mannetjes zonder vrouwtje die in de onmiddellijke omgeving van een paar verblijven. De rol van deze zogenaamde ‘satellieten’ is nog niet uitgeklaard. Er is wel een vermoeden dat het al of niet aantrekken van ‘satellieten’ afhangt van bepaalde kenmerken van het vrouwtje. Een wetenschapper heeft informatie verzameld over vrouwelijke krabben tijdens de paartijd. Voor elk van de vrouwtjes werd haar gewicht (in gram) en de grootte van haar schild (in cm) opgemeten. Daarnaast werd vastgesteld of er minstens één satelliet in de buurt was (TRUE) of niet (FALSE). De gegevens van 173 krabben werden verzameld in file “Krabben.csv”. Op basis van deze gegevens wenst de wetenschapper de volgende vragen te onderzoeken:

- (a) Welk percentage van de vrouwelijke degenkrabben in de populatie beschikt tijdens de paringsperiode over minstens één satelliet?
- (b) Hangt de aanwezigheid van satellieten af van de grootte van het schild van het vrouwtje? Zo ja, in hoeverre is de schildgrootte van de vrouwtjes met satelliet enerzijds en die zonder satelliet anderzijds verschillend?
- (c) Stel een model op dat toelaat om de aanwezigheid van minstens één satelliet te voorspellen aan de hand van het gewicht en de schildgrootte van het vrouwtje. Hoe betrouwbaar zijn de voorspellingen op basis van dit model?

# Taken

- Rapporteer de resultaten van je analyses in een beknopt en duidelijk rapport van maximaal 4 pagina's.
- Beschrijf in je rapport de beschikbare gegevens op een relevante wijze voor elke analyse.
- Voeg een appendix toe met de R commando's die je gebruikt hebt om de analyse resultaten in je rapport te bekomen. Deze commando's moeten volledig zijn en voorzien van commentaar zodat het duidelijk is wat het doel is van elk stukje R code.
- Dien je rapport ten laatste in op 31 mei. Het mondelinge examen gaat door op donderdag 5 en vrijdag 6 juni 2014.

Veel plezier en succes!