

Tentamen Statistiek 1 voor Scheikunde (2DS00) op dinsdag 16 december 2003, 9.00-12.00 uur

De uitwerkingen van de opgaven dienen duidelijk geformuleerd en overzichtelijk opgeschreven te worden. De numerieke waarden in de datasets zijn weergegeven in de Engelse decimale notatie.

De benodigde data sets (met decimale punten i.p.v. komma's) van de tentamenopgaven zijn beschikbaar via <http://www.win.tue.nl/~sandro/2DS00/tentamens.html>.

Uitwerkingen van het tentamen zullen zo spoedig mogelijk in te zien zijn op het WWW via URL:

<http://www.win.tue.nl/~sandro/2DS00/>

1. Bij deze opgave gebruiken we de data set **bromide.sf3**. Deze dataset bevat 15 metingen van het bromidegehalte (uitgedrukt in $\mu\text{g/g}$) in tomaten bepaald via een chromatografische methode.
 - a) Toets of het verwachte bromidegehalte het wettelijk voorgeschreven maximum van $765 \mu\text{g/g}$ overschrijdt. Geef duidelijk de details van een handmatige berekening of de keuzes binnen Statgraphics.
 - b) De toets bij a) gaat uit van bepaalde veronderstellingen. Onderzoek via geschikte grafieken en toetsen of de meetgegevens aan deze veronderstellingen voldoen. Verandert Uw conclusie bij a)?
2. Volgens de ideale gaswet geldt dat $pV = mRT$, waarbij p de druk is, V het volume, m de massa van het gas, R de gasconstante en T de temperatuur. We nemen aan dat de gasconstante R exact bekend is en de waarde 0,287 heeft. Gegeven zijn de volgende waarden van een bepaald gas en de meetnauwkeurigheid die een bepaald laboratorium kan halen:

$$p = 820 \text{ Pa}, \sigma_p = 7,50 \text{ Pa}$$

$$m = 2,00 \text{ kg}, \sigma_m = 0,01 \text{ kg}$$

$$T = 305 \text{ K}, \sigma_T = 1,50 \text{ K}.$$

- a) Bepaal de standaardafwijking in het volume van dit gas als we de ideale gaswet toepassen.
- b) Het laboratorium wil graag de nauwkeurigheid van de volumebepaling vergroten. Vanwege budgetbeperkingen dient men te kiezen tussen een nieuwe drukmeter of een nieuwe temperatuurmeter. In beide gevallen kan men een halvering van de huidige standaardafwijking bereiken. Beargumenteer wat de beste keuze is voor het laboratorium (de drukmeter en de temperatuurmeter zijn alleen inzetbaar voor volumemetingen van gassen).

3. Bij deze opgave gebruiken we de dataset **colburn.sf3**. Bij een bepaald type reactor wordt er stof getransporteerd van een deeltjesoppervlak naar fluïdum dat door het reactorbed stroomt. De Colburnfactor voor stofoverdracht is een dimensieloze coëfficiënt. In sommige gevallen blijkt deze factor alleen af te hangen van het zogenaamde Reynoldsgetal Re . De dataset colburn.sf3 bevat metingen van de Colburnfactor voor bolvormige deeltjes in een zogenaamde gepakt reactorbed. Voer regressie-analyses uit met enkele eenvoudige modellen. Kies een goed passend model en voer daarop een volledige regressie-analyse uit. Verwijder eventuele verdachte waarnemingen. Geef duidelijk aan waarom U kiest voor het door U gekozen model.
4. Bij deze opgaven gebruiken we de dataset **linthurst.sf3**. Deze dataset bevat metingen van biochemische experimenten uitgevoerd door de biochemicus Linthurst. Het betreft metingen gedaan aan een moerasgrassoort *Spartina alterniflora* die voorkomt in North Carolina. Het doel van de metingen was inzicht te krijgen in de gevolgen van de bodemkwaliteit op de biomassa die door deze grassoort wordt geproduceerd. De afhankelijke variabele is biomassa (g/m^2). De onafhankelijke variabelen zijn zout (zoutgehalte, %), zuurgraad (pH), kalium (ppm), natrium (ppm) en zink (ppm). Voer een regressie-analyse uit met de 5 onafhankelijke variabelen, waarbij U geen modelcontroles e.d. uit hoeft te voeren. **Om de hoeveelheid werk beperkt te houden, hoeven in deze opgave geen interactietermen (zoals bijv. zout*kalium) meegenomen te worden.**
- a) Laat zien dat het model significant is. Hoeveel procent van de variantie in de data wordt verklaard door dit model?
- b) Welke onafhankelijke variabelen zijn niet significant? Hoe verklaart U dat het model significant is, terwijl de meeste onafhankelijke variabelen niet significant zijn (liefst met een korte kwantitatieve onderbouwing)?
- c) Bepaal een beter model dan het model met alle 5 onafhankelijke variabelen door minstens één onafhankelijke variabele weg te laten. Geef duidelijke argumenten aan waarop U Uw keuze baseert.
- d) Voer een uitgebreide regressie-analyse uit met het door U bepaalde regressiemodel in c). Indien U bij c) geen model hebt kunnen bepalen, neemt U het volledige model met alle 5 onafhankelijke variabelen.
5. Bij deze opgave gebruiken we de dataset **puromycin.sf3**. De galactosyltransferase van Golgi membranen (een enzymatische reactie) kan mogelijk versneld worden door toevoeging van puromycine. De snelheid van de reactie wordt gemeten via het aantal radioactieve deeltjes per minuut. Om de onderzoeksvraag te beantwoorden voert men experimenten uit met verschillende substraatconcentraties puromycine. Op grond van kinetische overwegingen kiest men voor het volgende model: $v(x) = \frac{\theta_1 x}{\theta_2 + x}$, waarbij v de reactiesnelheid is (uitgedrukt in deeltjes/ min^2) en x de concentratie puromycine (uitgedrukt in ppm) is. Vanwege het feit dat reactiesnelheid en concentratie positieve getallen zijn, dienen de parameters θ_1 en θ_2 beide positief te zijn.
- a) Laat zien dat het model geschreven kan worden als een lineair model. Geef hierbij duidelijk de parameters aan.
- b) Toon via een korte analyse aan dat de data niet goed passen bij het lineaire

model van a).

c) Geef voor elk van de parameters θ_1 en θ_2 een interpretatie in termen van de grafiek van het model. Aanwijzing: vind eerst een interpretatie voor θ_1 . Beargumenteer daarna waarom θ_2 in dit geval niet in termen van een verticale asymptoot geïnterpreteerd kan worden. Vind een alternatieve interpretatie door $v(\theta_2)$ te berekenen.

d) Voer een uitgebreide niet-lineaire regressie uit. Beargumenteer of dit een goed passend model oplevert.