

Tussentoets Ruimtelijke Analyse en GIS 1, geo2-4206

9 maart 2007, 9-12h

AW C008/KCZ

Voor het RA gedeelte is de toets open boek: een statistiek boek (W&W, Freund) mag worden gebruikt, alsmede eventuele kopieën uit het W&W boek.

Wanneer berekeningen worden gevraagd, laat dan met volledige of gedeeltelijke tussenberekeningen zien hoe het antwoord tot stand is gekomen.

Alle vragen dienen te worden beantwoord.

Vraag 1)

Een meervoudige regressie-analyse met S-Plus, waarbij log-zinc wordt gemodelleerd als functie van hoogte in het terrein (alt) en afstand tot de rivier (dist) geeft de volgende uitvoer:

```
Call: lm(formula = log.zn ~ alt + dist, data = maas)
```

```
Residuals:
```

Min	1Q	Median	3Q	Max
-1.013	-0.2485	0.008049	0.2434	1.089

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t)
(Intercept)	8.5493	0.2815	30.3662	0.0000
alt	-0.2639	0.0369	-7.1444	0.0000
dist	-0.0018	0.0002	-10.1578	0.0000

```
Residual standard error: 0.4179 on 152 degrees of freedom
```

```
Multiple R-Squared: 0.6692
```

```
F-statistic: 153.7 on 2 and 152 degrees of freedom, the  
p-value is 0
```

Beantwoord de volgende vragen precies en kort.

- Hoe ziet de regressievergelijking er uit voor de afzonderlijke waarnemingen? Vul de berekende (geschatte) coëfficiënten in.
- Wat wordt bedoeld met "the p-value is 0", en welke conclusie mogen we hieruit trekken? (noem evt. ter verduidelijking de nulhypothese)
- Hoe moet de regressiecoëfficiënt voor alt worden geïnterpreteerd?
- Bereken de "corrected R-Squared" voor dit regressiemodel.

In dezelfde data set wordt een twee-weg variantie analyse gedaan naar de relatie van log-zinc met twee factoren: kalkgehalte (limecls: kalkrijk of kalkarm) en overstromingsfrequentie (flfreq: nooit, soms, vaak):

```
> summary(aov(log.zn~limecls*flfreq, maas))
              Df Sum of Sq  Mean Sq  F Value    Pr(F)
limecls      1  29.40236  29.40236  109.0390 0.000000000
flfreq      2   7.97904   3.98952   14.7952 0.000001367
limecls:flfreq 1   2.42240   2.42240    8.9835 0.003189301
Residuals 150  40.44748   0.26965
```

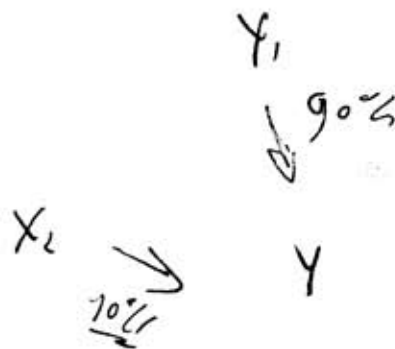
Residual standard error: 0.5192782
 1 out of 6 effects not estimable
 Estimated effects may be unbalanced

- e) Beoordeel of de interactie (limecls:flfreq) significant is.
- f) Leg in woorden uit wat het betekent wanneer deze interactie significant is.
- g) 1 out of 6 effects not estimable (*niet schatbaar*). Leg uit hoe dit komt (NB het betreft hier 1 van de 2 interactietermen).

Vraag 2)

- a) Geef een argument voor en een argument tegen het gebruik van stapsgewijze variabeleselectie, wanneer er een groot aantal sterk gecorreleerde predictoren beschikbaar zijn.
- b) Leg uit hoe je partiële correlaties kunt gebruiken voor het beoordelen of een effect direct dan wel indirect is (eventueel met voorbeeld).

bokend.
 ↓↓
 Nik
 ↓↓
 Plank.



$$r_{X_1} = 0,9$$

$$r_{X_2} = 0,1$$

$$r_{X_1, X_2} = 0,1$$

Vraag 3)

In het boek van Burrough & McDonnell worden twee verschillende conceptuele modellen gegeven van geografische fenomenen.

- a) Welke twee modellen zijn dit? Leg voor elk conceptueel model uit wat de uitgangspunten zijn van dit model. Illustreer je antwoord met een voorbeeld.

Raster gegevens kunnen op verschillende manieren worden opgeslagen in een Geografisch Informatie Systeem (GIS). Bekende methoden zijn run length codering en het gebruik van quadrees.

- b) Leg uit wat een quadtree is (in de context van de opslag van raster gegevens in een GIS). Illustreer indien dit je antwoord verduidelijkt met een tekening.
- c) Leg uit wat run length codering is (in de context van de opslag van raster gegevens in een GIS). Illustreer indien dit je antwoord verduidelijkt met een tekening.

Een GIS visualisatie programma voor het afbeelden van kaarten op het scherm dient vaak snel, door in- en uitzoomen, op verschillende resoluties gegevens te kunnen afbeelden.

- d) Leg uit wanneer een dergelijk visualisatie programma sneller zal werken: in het geval dat de raster gegevens worden opgeslagen als een quadtree of in het geval dat deze gegevens worden opgeslagen met run length codering.

Vraag 4)

Er is een papieren geologische kaart beschikbaar van een gebied dat geheel bestaat uit Holocene afzettingen van rivieren. De kaart is gemaakt in 1970. Doordat er sinds 1970 een aanzienlijke hoeveelheid nieuwe gegevens is verzameld en beschikbaar is gekomen, kan nu een betere, en meer gedetailleerde kaart worden gemaakt.

- Deze nieuwe kaart kan gebaseerd worden op een groot aantal beschikbare gegevens: veldobservaties (met name boringen van voor en na 1970) in het gebied, topografische informatie, hoogtemetingen, de geologische kaart 1970 op papier. Een deel van dit soort gegevens is digitaal beschikbaar.
- Delen van de kaart zullen hetzelfde blijven als geïnterpreteerd in 1970, maar grote delen zullen veranderen.
- De nieuwe kaart zal worden gemaakt in een GIS.

De eerste stap is het invoeren van deze ruimtelijke gegevens in het GIS systeem ('data import').

- a) Geef voor elk van de bronnen van gegevens aan hoe deze kunnen worden ingevoerd in een GIS systeem.
- b) Geef voor elk van de bronnen van gegevens aan hoe deze beschikbaar zijn in het GIS direct na data import: als 'vector data structure' of als 'raster data structure'. Leg uit.

Na deze fase van 'data import' zijn de gegevens beschikbaar in het GIS, en kan een nieuwe geologische kaart gemaakt gaan worden met behulp van het GIS.

- De opdrachtgever wil uiteindelijk een kaart ontvangen in een vector formaat. Hier moet vooral op staan: 1. waar rivierlopen liggen en 2. hoe oud iedere rivierloop is.
 - De opdrachtgever voorziet in maximaal 2 tussenronden waarbij geologische experts feedback geven op voorlopige versies van de geologische kaart. De experts weten alles van riviergeologie maar weten niks van GIS-techniek.
- c) Geef aan op welke wijze de verschillende geïmporteerde gegevens kunnen worden gecombineerd tot een nieuw GIS data product: een up-to-date digitale geologische kaart in 'vector data structure'. Onderscheid tenminste 3 stappen in de combinatie procedure.

Tijdens de procedure van het maken van de nieuwe geologische kaart kan feedback worden gevraagd aan geologische experts.

- d) Wat voor feedback vraag je aan de geologische experts en hoe gebruik je het GIS bij het vragen van deze feedback en bij het verwerken van deze feedback?