

Ruimtelijke Analyse en GIS 1, Tussentoets, 14 maart 2008, 9-12 uur

Elke deelvraag (1a, 1b, 2a, etc.) heeft bij benadering hetzelfde gewicht in de berekening van je cijfer.

Vraag 1

Geef een definitie van:

- a) Point-in-polygon search;
- b) Quadtree (in de context van het opslaan van raster gegevens);
- c) Hierarchical database structure.

Vraag 2

Er moet op basis van veldgegevens een kaart (of kaarten) worden gemaakt met bodemgegevens van een deel van de Ardennen, België.

- a) Er bestaan twee typen conceptuele modellen om ruimtelijke informatie te representeren. Beschrijf elk van deze twee conceptuele modellen en illustreer deze beschrijvingen aan de hand van het voorbeeld van de bodemkaart.

Er bestaan twee typen data modellen om ruimtelijke informatie te representeren: het vector data model en het raster data model.

- b) Let uit hoe wegen worden gerepresenteerd op een kaart met behulp van 1. het vector data model en 2. het raster data model.
- c) Er bevinden zich wegen van verschillende breedte in het gebied. Leg uit hoe deze informatie wordt opgeslagen in 1. het vector data model en 2. het raster data model van vraag 2b.

Vraag 3

Iemand heeft een remote sensing beeld waarmee een ruimtelijk beeld van vegetatie bedekking kan worden verkregen. Daarnaast is er een papieren geologische kaart beschikbaar. Er dient een GIS data base te worden gemaakt bestaande uit een vegetatiekaart en een geologische kaart, beiden in raster formaat. De vegetatiekaart dient de vegetatie bedekking weer te geven, relevant voor het modelleren van interceptie, niet de soorten vegetatie die gevonden worden.

- a) Leg uit welke stappen moeten worden ondernomen om te komen tot een raster kaart van de geologie, uitgaande van de hierboven gegeven informatie.
- b) Leg uit welke stappen moeten worden ondernomen om te komen tot een raster kaart van de vegetatie, uitgaande van de hierboven gegeven informatie..

Vraag 4

Een onderzoeker veronderstelt dat bodemtype van invloed is op de neerslag. Op 107 plaatsen in Nederland is op 20 juni 2007 zowel neerslag (in mm) als bodemtype (zand, zeeklei, rivierklei, veen, löss) bepaald. De onderzoeker presenteert als bewijs de volgende ANOVA tabel.

Response: neerslag				
	Df	Sum Sq	Mean Sq	F
bodemtype	5	490,2	98,04	2,6040
residual	101	4706,3	46,59	
Totaal	106	5196,5		

a) Stel dat de tabel juist is. Leg uit: Wat is H_0 in dit geval? Wordt H_0 verworpen bij $\alpha = 0,05$? Wat is de verklaarde variantie? Zijn de gegevens gebalanceerd? Wat is je conclusie uit deze analyse?

De wetenschapper heeft aantoonbaar fouten gemaakt in de ANOVA tabel.

b) Welke fouten? Geef de verbeterde tabel, ervan uitgaande dat n , $SS_{\text{bodemtype}}$, en SS_{residual} juist zijn. Wat verandert er in je antwoorden bij a)?

c) Geef twee argumenten tegen de geldigheid van dit onderzoek.

Vraag 5

Maak bij het beantwoorden gebruik van de bijlage.

Een onderzoeker is geïnteresseerd in de invloed van abiotische eigenschappen van eilanden op het voorkomen van endemische diersoorten (endemisch: soorten die nergens anders voorkomen). Een dataset geeft van alle eilanden in de Galapagos archipel de grootte (*Area*), de hoogte (*Elevation*), de grootte van het naburige eiland (*Adjacent*), het aantal aanwezige soorten schildpadden (*Species*), en natuurlijk het aantal endemische soorten (*Endemic*).

Allereerst (**model A**) zou je kunnen zeggen: *Endemic* hangt af van *Species*, want hoe meer soorten, hoe meer er eventueel endemisch kunnen zijn.

a) Schrijf hiervoor het statistische model op, en de nulhypothese.

Vervolgens (**model B**) veronderstelt de onderzoeker dat *Endemic* behalve van *Species* ook van de andere variabelen afhangt: hier is zij eigenlijk in geïnteresseerd.

b) Waarom blijft *Species* als onafhankelijke variabele in het model?

c) Leg uit: welke rij diagnostic plots hoort bij model A? En welke bij model B?

Na bestudering van de diagnostic plots stelt de onderzoeker het derde model op (**model C**).

d) Leg uit wat zij doet, en waarom. Gebruik alle informatie die je hebt.

e) Welk model is het best? Wat zegt dat over de veronderstellingen van de onderzoeker?

Bijlage bij vraag 5: Computer-uitvoer van de modellen

Model A

Call:

```
lm(formula = Endemics ~ Species, data = gala)
```

Residuals:

Min	1Q	Median	3Q	Max
-13.8439	-4.1038	-0.7102	2.8645	19.6577

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.95644	1.48938	4.671	7.38e-05
Species	0.22948	0.01037	22.133	< 2e-16

Residual standard error: 6.368 on 27 degrees of freedom
Multiple R-squared: 0.9478, Adjusted R-squared: 0.9458
F-statistic: 489.9 on 1 and 27 DF, p-value: < 2.2e-16

Analysis of Variance Table

Response: Endemics

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Species	1	19863.2	19863.2	489.88	< 2.2e-16
Residuals	27	1094.8	40.5		

Model B

Call:

```
lm(formula = Endemics ~ Species + Area + Elevation + Adjacent, data = gala)
```

Residuals:

Min	1Q	Median	3Q	Max
-7.4004	-3.0767	-0.6265	1.5222	13.9819

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.954544	1.530061	2.585	0.01626
Species	0.181930	0.017271	10.534	1.76e-10
Area	-0.003823	0.001957	-1.954	0.06250
Elevation	0.024451	0.007139	3.425	0.00222
Adjacent	-0.003854	0.001982	-1.945	0.06364

Residual standard error: 5.373 on 24 degrees of freedom
Multiple R-squared: 0.9669, Adjusted R-squared: 0.9614
F-statistic: 175.5 on 4 and 24 DF, p-value: < 2.2e-16

Analysis of Variance Table

Response: Endemics

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Species	1	19863.2	19863.2	688.1223	< 2.2e-16
Area	1	10.8	10.8	0.3742	0.546459
Elevation	1	282.0	282.0	9.7707	0.004594
Adjacent	1	109.1	109.1	3.7812	0.063640
Residuals	24	692.8	28.9		

(model C op het volgende blad)

Model C

Call:
lm(formula = ln(Endemics) ~ ln(Species), data = gala)

Residuals:
Min 1Q Median 3Q Max
-0.60039 -0.10569 0.07028 0.12958 0.38487

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.07961 0.09569 0.832 0.413
ln(Species) 0.75133 0.02485 30.229 <2e-16

Residual standard error: 0.2091 on 27 degrees of freedom
Multiple R-squared: 0.9713, Adjusted R-squared: 0.9702
F-statistic: 913.8 on 1 and 27 DF, p-value: < 2.2e-16

Analysis of Variance Table

Response: ln(Endemics)
Df Sum Sq Mean Sq F value Pr(>F)
ln(Species) 1 39.951 39.951 913.79 < 2.2e-16
Residuals 27 1.180 0.044

