

Ruimtelijke Analyse en GIS 1 (AW2-2006); Tussentoets 2004
Bacheloropleiding Aardwetenschappen

Donderdag 4 maart 2004; tijd: 14.00 - 17.00.

NB: U moet alle vragen beantwoorden.

Tekeningen en grafieken kunnen uw antwoorden verhelderen.

1. Geef antwoord op de volgende:
 - Geef twee definities en een korte toelichting van het begrip 'Geographical Information System'
 - Beschrijf in het kort de 5 belangrijkste hardware componenten van een GIS en hun functies
 - Leg uit wat de basis elementen (geographical primitives) zijn van de geografisch informatie
 - Verklaar het gebruik van 'data planes' ('overlays') in GIS
 - Leg uit wanneer en waarom men interpolatiemethoden moet gebruiken in GIS.
2. Beschrijf de basis-principes, aannamen en data structuren van Vector- en Raster-GIS. Voor welke toepassingen is de ene vorm boven de andere te prefereren? Voor welke toepassing is een gecombineerde aanpak het meest geschikt? (Denk aan toepassingen binnen de fysische geografie).
3. Met behulp van remote sensing, GPS, fotogrammetrie, digitizing en scanning kan de geometrie van het fysisch aardoppervlak worden verzameld vooruitlopend op het gebruik ervan binnen een GIS.
 - a. Behandel deze vijf technieken in het kort en geef aan of het om raster- of vector-data gaat en aan welke ruimtelijke nauwkeurigheid we ongeveer moeten denken.
 - b. Geef voorbeelden van andere toepassingen voor GPS aan de input-kant van GIS naast het verzamelen van geometrische data?

6.
 Vraag 4. Hieronder volgt de S-Plus uitvoer van een tweetal lineaire modellen.
 De variabele year is een numerieke variabele, niet een factor.

*** Linear Model ***

Call: lm(formula = amm ~ year, data = lmr, na.action = na.exclude)

Residuals:

Min	1Q	Median	3Q	Max
-22.29	-11.99	-6.295	8.686	49.01

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-6595.5897	7133.4475	-0.9246	0.3612
year	3.3462	3.5775	0.9353	0.3557

Residual standard error: 18.24 on 37 degrees of freedom

Multiple R-Squared: 0.0231

F-statistic: 0.8749 on 1 and 37 degrees of freedom, the p-value is 0.3557

*** Linear Model ***

Call: lm(formula = amm ~ year + station, data = lmr, na.action = na.exclude)

Residuals:

Min	1Q	Median	3Q	Max
-12	-2.667	0.6667	3.179	9.654

Coefficients:

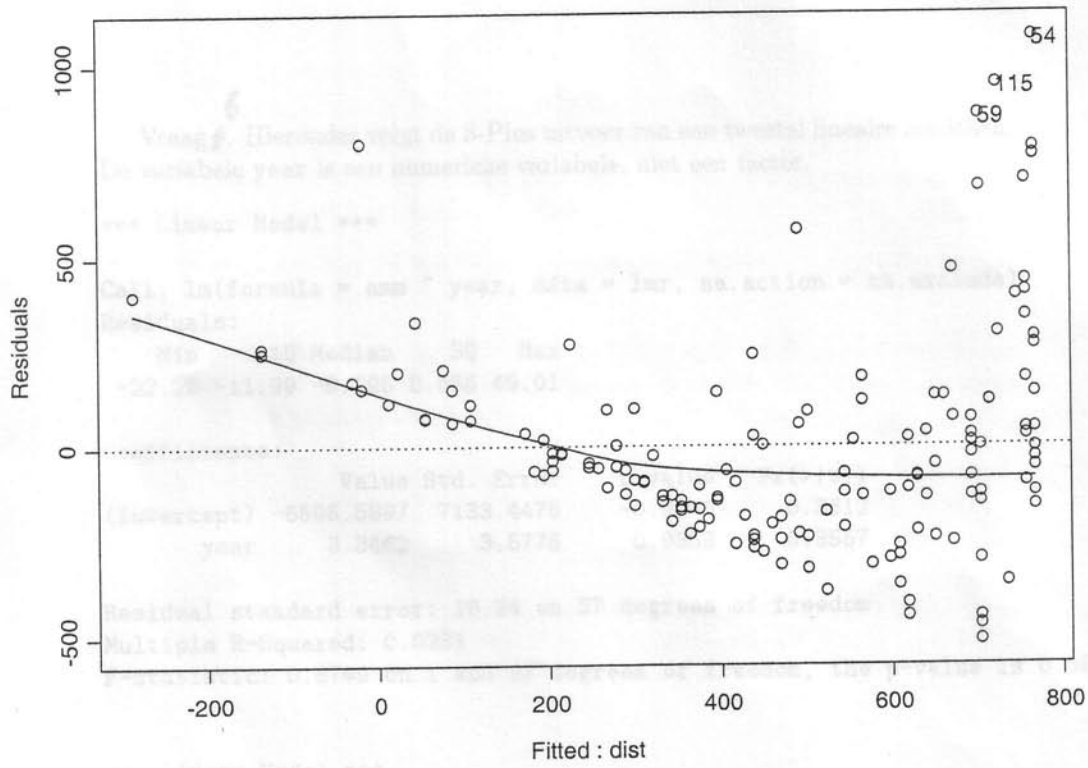
	Value	Std. Error	t value	Pr(> t)
(Intercept)	-6595.5897	2189.2846	-3.0127	0.0059
year	3.3462	1.0979	3.0477	0.0054
station1	-6.8333	2.2855	-2.9898	0.0062
station2	0.5000	1.3196	0.3789	0.7079
station3	-0.8333	0.9331	-0.8931	0.3803
station4	6.2333	0.7227	8.6245	0.0000
station5	0.6556	0.5901	1.1109	0.2772
station6	-2.1508	0.4987	-4.3124	0.0002
station7	-1.4464	0.4319	-3.3488	0.0026
station8	-2.1991	0.3809	-5.7730	0.0000
station9	4.4074	0.3407	12.9360	0.0000
station10	0.8485	0.3082	2.7532	0.0108
station11	-1.3207	0.2813	-4.6945	0.0001
station12	-1.2201	0.2588	-4.7147	0.0001

Residual standard error: 5.598 on 25 degrees of freedom

Multiple R-Squared: 0.9378

F-statistic: 29.01 on 13 and 25 degrees of freedom, the p-value is 7.806e-012

a geef de vergelijking van de twee lineaire modellen die zijn geanalyseerd;
 geef duidelijk aan waar elke term voor staat.



- Vraag 5. Bovenstaand figuur is een zgn. Diagnostic plot van een regressie waarbij zingehalte als afhankelijke variabele is gekozen (Maas data set).
- Hoe worden de waarden op de x -as berekend?
 - Wat was/waren de onafhankelijke variabele(n)?
 - Geef een interpretatie van de diagnostic plot. Wat scheelt er aan, en wat zijn mogelijke oplossingen?

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-5805.5897	7133.4478	-0.8127	0.4204
year	3.3462	3.5778	0.9354	0.3524
station1	-0.8333	2.2855	-0.3648	0.7163
station2	0.5002	1.3198	0.3789	0.7079
station3	-0.8333	0.9331	-0.8931	0.3767
station4	4.2333	0.7227	5.8585	0.0000
station5	0.6556	0.2601	2.5208	0.0172
station6	-2.1808	0.4057	-5.3754	0.0000
station7	-1.4465	0.4319	-3.3488	0.0026
station8	-2.1991	0.3809	-5.7730	0.0000
station9	4.4074	0.3407	12.9360	0.0000
station10	0.8421	0.3082	2.7322	0.0108
station11	-1.3267	0.2813	-4.7147	0.0001
station12	-1.2301	0.2508	-4.9047	0.0001

Residual standard error: 16.24 on 27 degrees of freedom
 Multiple R-Squared: 0.0281

F-statistic: 29.01 on 12 and 27 degrees of freedom, the p-value is 7.206e-012

geef de vergelijking van de twee laatste modellen die zijn gekozen
 geef duidelijk aan waar elke term voor staat

b wanneer je geïnteresseerd bent in verandering van ammoniumgehalte in regenwater, als functie van de tijd, welk van de twee regressiemodellen is dan het meest interessant? Leg uit.

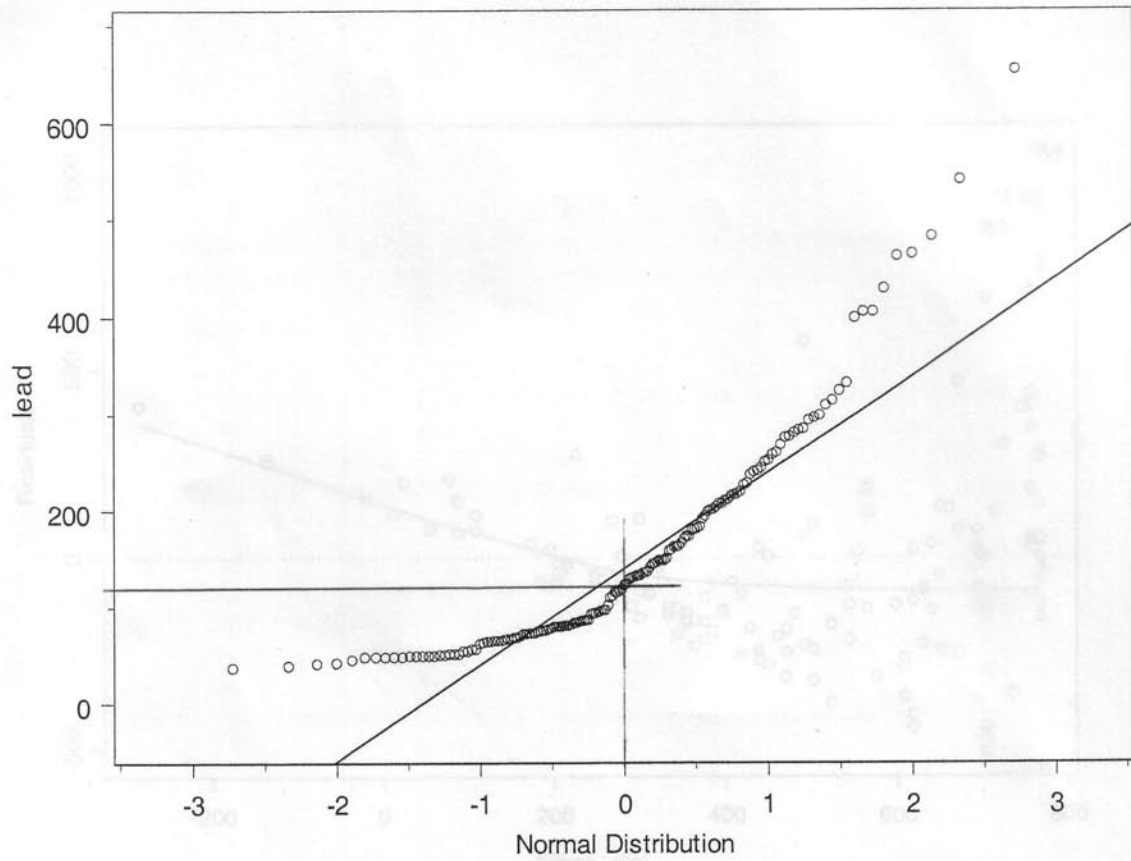
c Er is gemeten op 13 stations. Waarom zijn er slechts 12 coëfficiënten voor stations gegeven?

d de geschatte coëfficiënt voor year is gelijk voor beide regressiemodellen. Wat kun je hieruit concluderen ten aanzien van de relatie (en b.v. correlatie) van year en station?

e geef voor elk van de twee modellen aan wat het is: (i) een regressie-model, (ii) een ANOVA model, of (iii) een gemengd regressie-ANOVA model.

f bereken de gecorrigeerde R^2 , \bar{R}^2 voor beide modellen.

g welk van de maten is geschikt voor het vergelijken van de kwaliteit van deze twee modellen: (1) residual standard error, (2) R^2 , (3) \bar{R}^2 , (4) F, (5) p-value van F, (6) residual degrees of freedom.



Vraag 4.

Beoordeel bovenstaand figuur.

- a. kan de verdeling van loodgehaltes als normaal worden beschouwd?
- b. Wat is het mediane loodgehalte?
- c. Hoe kun je uit deze figuur aflezen tussen welke grenzen 95% van de loodgehaltes ligt?