

## VERVOLGDOCTORAAL FYSISCH GEOGRAFIE

Tentamen: MUVS

Datum: vrijdag 26 oktober 2001 Tijd: 14-17 uur

Docent: Pebesma

TENTAMENAANWIJZINGEN: NIET INLEVEREN!

### LEES DIT EERST!

1. Op je tafel ligt een bruin formuliertje, voorzien van een nummer. Dit nummer dien je te vermelden op ALLE op het tentamen uitgereikte tentamenformulieren, en eventueel bijgevoegde kaartjes, tabellen etc.
2. Op dit bruine formuliertje vermeld je je studentnummer en de naam met voorletters.
3. Let op: voor het verlaten van de zaal alle tentamenvellen (incl. figuren) inleveren. De vragen mag je behouden.
4. Lees eerst rustig alle vragen door. Beantwoord de vragen eerst puntsgewijs op het kladvel en werk het dan pas uit op het tentamenblad (de achterkant mag gebruikt worden). Houd het kort en duidelijk gestructureerd.
5. Indien veel doorhalingen e.d. de leesbaarheid sterk verminderen kan nieuw tentamenpapier worden verstrekt. Schrijf s.v.p. duidelijk.
6. Na het verlaten van de zaal niet in de buurt samenscholen! De geluidsoverlast is hinderlijk voor de resterende deelnemers!
7. Dit blad dient verder als KLADPAPIER. Extra kladpapier kan worden verstrekt. Kladpapier NIET inleveren!
8. **Je mag de tentamenzaal op zijn vroegst een half uur na de aanvang van het tentamen verlaten!**

De uitslag van dit tentamen wordt vanaf 16 november 2001 opgehangen op het prikbord bij het secretariaat FG, in de Jan Zonneveldvleugel.

Vervolgdoctoraal Fysische Geografie,  
Tentamen Multivariate Statistiek en Matrix Algebra (RW-MVSI)  
Vrijdag 26 oktober 2001, 14-17 uur

Maak alle vragen (gesloten boek).

VRAAG 1.

Gegeven de matrices  $A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 1 \\ 3 & 2 & 2 \end{bmatrix}$ ,  $b = \begin{bmatrix} 1 \\ 10 \\ 11 \end{bmatrix}$ ,  $c = \begin{bmatrix} 0.71 \\ 0.71 \\ 0 \end{bmatrix}$ .

- a Bereken de determinant van  $A$ .
- b Is  $A$  singulier? Licht toe.
- c Los  $Ax = b$  op voor  $x$ .
- d Bereken de inverse van  $A$ .
- e Stel dat  $c$  de eerste eigenvector is van de correlatiematrix van een datamatrix, en dat  $A$  deel uitmaakt van deze data matrix. Bereken nu voor de data records in  $A$  de scores op de eerste hoofdcomponent.

VRAAG 2. Er is een hoofdcomponentenanalyse uitgevoerd op oxiden en spore-elementen in de maas afzettingen, verzameld door Leo Tebbens (dataset computerpracticum). Hierbij is uitgegaan van de correlatiematrix. Oxiden werden uitgedrukt in gewichtspercentage, spore-elementen in ppm (parts per million). Zand bestaat nagenoeg uit  $\text{SiO}_2$ , klei voor het belangrijkste deel uit  $\text{Al}_2\text{O}_3$ .

Hieronder staan standaard deviaties (d.w.z. wortel van eigenwaarden) van alle hoofdcomponenten, en vervolgens de loadingen voor de eerste drie hoofdcomponenten.

Standard deviations:

Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6	Comp. 7	Comp. 8
3.915543	2.071144	1.370075	1.059248	0.9770295	0.8634739	0.834752	0.7671491

Comp. 9	Comp. 10	Comp. 11	Comp. 12	Comp. 13	Comp. 14	Comp. 15
0.6682423	0.6300045	0.5957847	0.4839472	0.3900084	0.3860264	0.3720318

Comp. 16	Comp. 17	Comp. 18	Comp. 19	Comp. 20	Comp. 21	Comp. 22
0.3435076	0.3149465	0.2615525	0.2354402	0.2110172	0.1833138	0.1659259

Comp. 23	Comp. 24	Comp. 25	Comp. 26	Comp. 27	Comp. 28
0.1574528	0.1335607	0.114701	0.1021311	0.08420635	0.04605262

Loadingen:

	Comp. 1	Comp. 2	Comp. 3
SiO2	-0.1766	-0.32860	0.0255
TiO2	0.2408	-0.10014	-0.1018
Al2O3	0.2449	0.00181	0.1517
Fe2O3	0.0753	0.39095	-0.1606
MnO	0.0383	0.36331	-0.2887
MgO	0.2381	0.04747	-0.0829
CaO	0.0571	0.27932	-0.3731
Na2O	0.0907	-0.17507	-0.2773
K2O	0.2421	-0.05947	0.0111
P2O5	0.0330	0.33572	-0.1144
LOI	0.1143	0.32871	0.0832
Ba	0.2347	-0.04694	0.1234
Ce	0.2406	-0.04584	0.0185
Co	0.2013	0.03476	-0.0324
Cr	0.2471	-0.05435	0.0178
Cu	0.1170	-0.08909	0.2334
Ga	0.2406	0.00124	0.1779
La	0.2221	-0.02582	0.0504
Nb	0.1499	-0.28547	-0.2445
Nd	0.2212	-0.14998	-0.1807
Ni	0.2284	0.04224	0.0861
Pb	0.0885	-0.06905	0.0118
Rb	0.2418	-0.03619	0.1536
Sr	0.2032	0.08084	-0.2508
V	0.2440	0.01196	0.1436
Y	0.1668	-0.18557	-0.2016
Zn	0.2195	0.09525	0.1835
Zr	0.0247	-0.28779	-0.4762

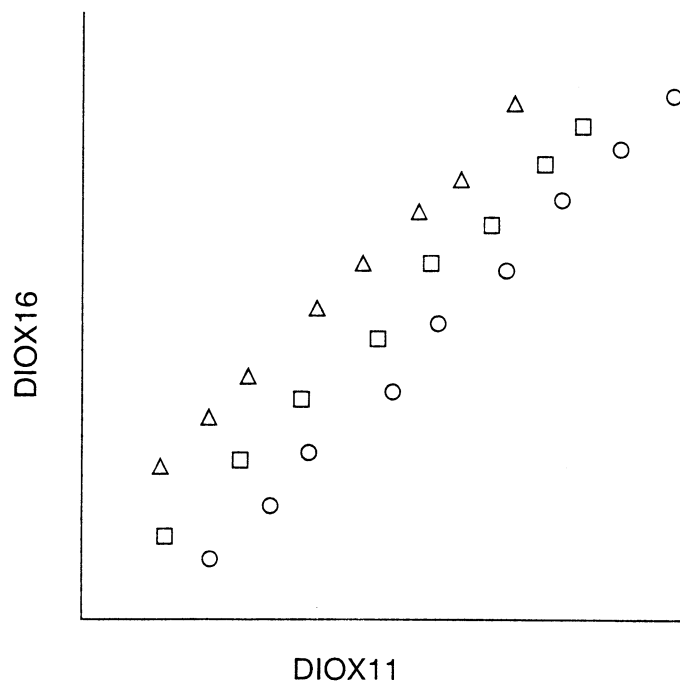
- geef de reden waarom het voor deze data beter is om met de correlatie matrix in plaat van met de covariantiematrix te werken.
- geef een interpretatie van de eerste, en van de tweede hoofdcomponent.
- hoeveel procent van de totale *variantie* wordt hierboven door de eerste en tweede hoofdcomponent tezamen verklaard?

- d stel, we willen volstaan met twee assen om het onder 3. genoemde percentage van de variantie te verklaren. Wat kun je nu doen om de interpretatie van deze assen in termen van de onderliggende variabelen te verbeteren? Leg uit.

VRAAG 3. Een promovenda (Karin Pfeffer) doet onderzoek naar vegetatietypen in een skigebied in de Alpen. Hiertoe heeft ze 150 vegetatieopnamen gemaakt waarbij in totaal 200 plantensoorten zijn opgenomen. In elke cel van de kruistabel van de opnames staat de abundantie, een maat voor frequentie van een plantensoort in het opname-gebiedje. Zij wil deze data samenvatten door middel van een of enkele *continue* (d.w.z. interval/ratio) variabelen.

- a Noem de multivariaat statistische techniek die ze hiervoor dient te gebruiken.
- b Hoe wordt bij deze techniek gecorrigeerd voor verschillen in voorkomen c.q. verschillen in soortrijkdom tussen locaties (m.a.w., hoe worden de data gecentreerd)?
- c Hoe wordt bij deze techniek gecorrigeerd voor verschillen in variabiliteit na het centreren (m.a.w., hoe worden de data gestandaardiseerd)?
- d Hoe kan de opnametabel het beste worden geordend?
- e Wat is de waarde van een geordende opnametabel ten opzichte van een willekeurig geordende tabel?
- f Op grond van de met onder 1. genoemde techniek gevonden continue variabelen wil de promovenda de data indelen in groepen (vegetatietypen). Noem de multivariaat statistische techniek die hiervoor geschikt is, en beschrijf kort hoe deze werkt.

VRAAG 4. Onderstaand figuur geeft, gestandaardiseerd, de spreiding van een aantal waarnemingen weer voor twee dioxines (DIOX11 en DIOX16). De drie verschillende symbolen duiden de herkomst van de monsters ( $\square$ ,  $\circ$ ,  $\triangle$  zijn resp. havens A, B of C).



- Geef de loadings van de eerste hoofdcomponent voor deze data, waarbij herkomstinformatie (haven) wordt genegeerd; teken de hoofdcomponent in de figuur.
- Hoe verandert de relatieve belangrijkheid van de eerste hoofdcomponent t.o.v. de situatie onder a wanneer niet over alle waarnemingen gezamenlijk, maar per haven een hoofdcomponentenanalyse wordt uitgevoerd.
- Geef de loadings van de (eerste) discriminantfunctie, die bedoeld is om de drie havens zo goed mogelijk te scheiden; teken deze functie in de figuur.
- Hoe ligt de tweede discriminantfunctie? (geef loadings en/of schets in figuur).