

## Multivariate Verfahren (WS 2016/17)

### Hinweise auf zentrale Punkte in der Klausur

Sie finden hier eine Folge von je zwei Aussagen; die zweite, durch eine Leerzeile von der erste getrennte Aussage ist wahr in Bezug auf die erste. Diese Folge soll es Ihnen gestatten, auf die wichtigen Punkte des behandelten Stoffes zu fokussieren.

Diese Zusammenfassung darf *nicht* als Hilfsmittel mit in die Klausur genommen werden. Den Mainzer Regularien entsprechend sind nur handschriftliche Notizen erlaubt!

1. Der erste Punkt kann hier übergangen werden.
2. **Vektortransformation** Linearkombinationen bedeuten die Transformation eines Vektors  $\mathbf{x}$  in einen Vektor  $\mathbf{y}$ , wie anhand der Matrixschreibweise für Linearkombinationen deutlich wird.  $n$ -dimensionale Vektor  $\mathbf{x}$  wird in den  $m$ -dimensionalen Vektor  $\mathbf{y} = A\mathbf{x}$  transformiert. Es gelte  $\|\mathbf{y}\|^2 = \mathbf{x}'A'A\mathbf{x} = \|\mathbf{x}\|^2$ , wobei  $A$  eine  $(n \times n)$ -Matrix ist.  
Die Beziehung  $\|\mathbf{y}\|^2 = \|\mathbf{x}\|^2$  impliziert  $A'A = I$ .
3. **Dyadisches Produkt** Es sei  $\mathbf{x}$  ein  $m$ -dimensionaler Vektor, und  $\mathbf{y}$  sei ein  $n$ -dimensionaler Vektor. Das dyadische Produkt  $\mathbf{x}\mathbf{y}'$  ist wie definiert? Welchen Rang hat die Matrix?  
Die Spaltenvektoren in der entsprechenden Matrix sind alle parallel – also Rang 1
4. **Skalarprodukt 1** Ein weiterer, für die multivariate Analyse zentraler Schlüsselbegriff ist der des Skalarprodukts  $\mathbf{x}'\mathbf{y}$  zweier Vektoren  $\mathbf{x}$  und  $\mathbf{y}$ .  
Summe der Produkte der Komponenten ...
5. **Skalarprodukt 2** Die Komponenten der  $m$ -dimensionalen Vektoren  $\mathbf{x}$  und  $\mathbf{y}$  seien Messwerte für zwei Variable  $X$  und  $Y$ , sind also zufällige Variablen.  
Man erinnere sich: Die Korrelation  $r_{xy}$  zwischen  $X$  und  $Y$  ist nicht gleich dem Skalarprodukt, es gilt vielmehr  $r_{xy} = \cos \theta = \mathbf{x}'\mathbf{y}/(\|\mathbf{x}\|\|\mathbf{y}\|)$ .
6. **Lineare Abhängigkeit 1** Es sei  $\mathbf{x}_1, \dots, \mathbf{x}_n$  eine Menge von  $n$ -dimensionalen Vektoren. Die Frage ist, ob sie linear abhängig oder linear unabhängig sind.  
Es kommt darauf an, die genaue, formale Definition zur Verfügung zu haben.

7. **Gleichungssysteme** Es sei  $A$  eine  $(n \times n)$ -Matrix und es sei das lineare Gleichungssystem  $A\mathbf{x} = \vec{0}$  gegeben; gesucht sind Vektoren  $\mathbf{x}$ , die dieser Gleichung genügen.

Der Vektor  $\mathbf{x} = \vec{0}$  stets eine Lösung, die Frage ist, wann dieser Vektor auch die einzige Lösung ist. Man denke an die Definition der linearen Unabhängigkeit.

8. **Dyadisches Produkt 2** Das dyadische (auch: äußere) Produkt  $\mathbf{x}\mathbf{y}'$  zweier Vektoren  $\mathbf{x}$  und  $\mathbf{y}$  ist ?

Schauen Sie sich noch einmal die genaue Definition an!

9. **Vektorraum 1** Mit dem Begriff des Vektorraums oder des Teilraums eines Vektorraums werden bestimmte Merkmale von Mengen von Vektoren charakterisiert. Dann gilt

$V_n = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}^n\}$  ist ein Vektorraum, wenn irgendwelche Linearkombinationen von Vektoren aus  $V_n$  ... wieder Elemente von ...

10. **Vektorraum 2** Jeder Vektorraum enthält unendlich viele Vektoren, da jede Linearkombination  $\mathbf{x} = \sum_{j=1}^p a_j \mathbf{x}_j$  von endlich vielen Vektoren  $\mathbf{x}_1, \dots, \mathbf{x}_p$  aus dem Raum wieder in diesem Raum liegt und die Koeffizienten  $a_j$  beliebig aus  $\mathbb{R}$  gewählt werden können.

Eine Basis ist eine Menge  $\{\mathbf{b}_1, \dots, \mathbf{b}_r\}$  von linear unabhängigen Vektoren aus dem Vektorraum, die die Erzeugung welcher Vektoren  $\mathbf{x}$  des Raums als Linearkombination  $\mathbf{x} = a_1 \mathbf{b}_1 + \dots + a_r \mathbf{b}_r$  erlaubt?  $r$  ist der Rang des Vektorraums.

11. **Rang einer Matrix 1** Es sei  $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$  eine  $(m \times n)$ -Matrix.

Für den Rang  $r$  von  $X$  gilt ganz allgemein?

12. **Ellipsoide 1**  $M$  sei eine  $(n \times n)$ -Matrix, für die  $M' = M$  gilt.  $M$  definiert eine Menge von Ellipsoiden,

denn  $\mathbf{x}'M\mathbf{x} = k$  ist die Formel für Ellipsoide ( $k$  kann ja irgendwelche Werte aus  $\mathbb{R}$  annehmen).

Anmerkung: die Konfiguration der Fälle muß nicht wie ein Ellipsoid aussehen. Diesen Fall kann man nur erwarten, wenn die Daten multivariat normalverteilt sind. Die durch  $M$  definierte Ellipsoide sind gewissermaßen die am besten passenden Ellipsoide, was die Ausdehnungen (Längen der Hauptachsen und deren Orientierung) der Konfiguration angeht.

13. **Rotation 1** Es gelte  $\mathbf{x}'M\mathbf{x} = k$ ,  $M$  symmetrisch, und  $\mathbf{y}'\Lambda\mathbf{y} = k$ ,  $\Lambda$  eine Diagonalmatrix,  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  und es existiere eine Rotationsmatrix  $T$  mit  $\mathbf{x} = T\mathbf{y}$ .

Es gilt  $\mathbf{x}'M\mathbf{x} = \mathbf{y}'T'MT\mathbf{y} \Rightarrow T'MT = \Lambda$ , und  $T$  ist orthonormal, kann also als was aus  $M$  bestimmt werden?

14. **Rotation 2**  $T'MT = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$  und  $T$  eine Rotation impliziert, dass  $T$  die Matrix der Eigenvektoren von  $M$  ist.
- $X$  sei eine spaltenstandardisierte  $(m \times n)$ -Datenmatrix,  $M = X'X$ . Dann ist  $L = XT$  die orthogonale Matrix, deren Spaltenvektoren die Koordinaten von  $\mathbf{x}_1, \dots, \mathbf{x}_n$  auf hypothetischen latenten Dimensionen enthalten?
15. **Basis** Es sei  $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$  eine  $(m \times n)$ -Matrix. Gesucht ist eine Basis  $\{\mathbf{L}_1, \dots, \mathbf{L}_r\}$  mit  $r \leq \min(m, n)$  für die Spaltenvektoren von  $X$ . Es wird der Ansatz  $X = LT'$  gemacht mit  $L = [\mathbf{L}_1, \dots, \mathbf{L}_r]$  und postuliert, dass die  $\mathbf{L}_k$  orthogonal sind.
- Die Frage ist, was der maximal mögliche Wert von  $\mathbf{L}'_1 \mathbf{L}_1$  ist und welcher Satz (Courant-Fischer) zur Beantwortung dieser Frage herangezogen wird.
16. **SVD** Es sei  $X$  eine  $(m \times n)$ -Datenmatrix, und  $X = Q\Lambda^{1/2}T'$  sei die zugehörige SVD (Singularwertzerlegung).
- Es ist  $XT = Q\Lambda^{1/2} = L$ , – welche Matrix bewirkt also eine Transformation der Hauptachsen der durch  $X$  gegebenen, orientierten Punktekonfiguration der Fälle in eine achsenparallele Konfiguration?
17. **PCA 1** Ein Kunsttheoretiker hat hundert Gemälde nach den Kriterien  $K_1, \dots, K_n$  auf einer Ratingskala jeweils von 1 bis 7 beurteilt und will nun die Struktur seiner Urteile mit einer PCA erkunden. Er berechnet die SVD  $X = Q\Lambda^{1/2}T'$  seiner Datenmatrix (Zeilen: Gemälde, Spalten: Kriterien  $K_j$ ,  $j = 1, \dots, n$ ).
- Man erinnere sich an die Definition der Matrix der Ladungen für die Kriteriumsvariablen!
18. **PCA 2** Es wird noch einmal die Datenmatrix des Kunsttheoretikers betrachtet:
- Man erinnere sich daran, dass Ladungen als Korrelationen zwischen latenten Variablen und gemessenen Variablen gedeutet werden können!
19. **Basisbestimmung** Für die übliche Datenmatrix  $X$  gelte  $XT = L$ .
- Die Spaltenvektoren von  $L$  sind Linearkombinationen von?
20. **Rotation 3** Es sei  $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]$  eine  $(m \times n)$ -Matrix, und die Fälle ( $\rightarrow$  Zeilen von  $X$ ) bilden eine orientierte, dh nicht achsenparallele Punktekonfiguration. Für die Rotation auf ein achsenparalleles Koordinatensystem gilt
- Die Rotation bedeutet den Übergang von  $X$  zu einer Matrix  $?$ , wobei  $?$  orthogonal (oder nicht-orthogonal?) ist.

21. **Rotation 4**  $X$  sei wie in der vorangegangenen Aufgabe definiert.  
 Durch welche Matrix ist die Rotationsmatrix  $T$  gegeben? Man erinnere sich an die Definition der Eigenvektoren von  $X'X$ .
22. **Eigenwerte 2**  $X$  sei wieder eine  $(m \times n)$ -Datenmatrix und es gelte  $XT = L$ ,  $L$  die Matrix der Koordinaten der Fälle auf den latenten Dimensionen.  
 Welcher bekannten Statistik entspricht  $\mathbf{L}'_j \mathbf{L}_j = \lambda_j$ ,  $j = 1, \dots, n$ ? (Mittelwert, oder Varianz, oder Kovarianz?)
23. **Faktorladungen 1** Es sei  $X$  eine spaltenstandardisierte  $(m \times n)$ -Datenmatrix mit der SVD  $X = Q\Lambda^{1/2}T'$ .  $A = T\Lambda^{1/2}$  ist die Matrix der Ladungen der Variablen (Zeilen  $\rightarrow$  Variablen, Spalten  $\rightarrow$  latente Dimensionen).  
 Bilden Sie das Produkt  $AA'$ , d.h.  $r_{jk} = \sum_i a_{ij}a_{ik}$ , – welcher Matrix entspricht also  $AA'$ ?
24. **Rotation 5** Die durch eine PCA bestimmten latenten Dimensionen für eine Datenmatrix  $X$  sind durch Rotation aus den gemessenen Variablen hervorgegangen; die Orthogonalität der latenten Variablen bedeutet, dass sie als unkorreliert interpretiert werden können. Welche Aussage über eine weitere Rotation ist korrekt?  
 Man erinnere sich: mit einer Rotation zB der Variablen-Achsen ist üblicherweise eine dazu korrespondierende Rotation der Achsen für die Fälle verbunden. Man kann aber entweder auf eine Rotation der latenten Dimensionen für die Variablen *oder* auf eine Rotation der latenten Dimensionen für die Fälle fokussieren.
25. **Kommunalität** Für Faktorenanalytiker ist es wichtig, die Kommunalität der untersuchten Variablen zu bestimmen.  
 Zur Erinnerung: Die Kommunalität einer Variablen ist der Anteil der Varianz dieser Variablen, der sich auf die gemeinsamen Faktoren zurückführen lässt.
26. **Eigenwerte-PCA** Für die  $(m \times n)$ -Matrix  $X$  (spaltenstandardisiert) sollen die latenten Dimensionen für die Variablen mittels PCA abgeschätzt werden.  
 Hier sind die großen Eigenwerte wichtig für die bedeutsamen latenten Variablen und die kleinen repräsentieren den Fehler.
27. **Scree 1** Die Matrix  $X$  sei wie in der vorangegangenen Aufgabe definiert. Es wird eine PCA für  $X$  berechnet und dabei der Scree-Test durchgeführt.  
 Man erinnere sich an die Möglichkeit, dass Eigenwerte größer als 1 sind, auch wenn die wahren Werte aller Korrelationen zwischen den gemessenen Variablen gleich Null sind; hier gibt es einen Test, den man vor Betrachtung des Scree-Plots durchführen kann.

28. **Regression 1**  $X$  sein eine  $(m \times n)$ -Prädiktormatrix und eine Variable  $Y$  soll durch die Spaltenvektoren von  $X$  vorhergesagt werden (multiple Regression), d.h. es soll  $\mathbf{y} = X\mathbf{b} + \mathbf{e}$  gelten,  $\mathbf{b}$  der Vektor der Regressionsgewichte. Da  $X'X = T\Lambda T'$ ,  $T$  die Matrix der Eigenvektoren von  $X'X$  und  $\Lambda$  die Diagonalmatrix der zugehörigen Eigenwerte, folgt

$$(X'X)^{-1} = \sum_{k=1}^n \frac{\mathbf{t}_k \mathbf{t}_k'}{\lambda_k}, \quad T = [\mathbf{t}_1, \dots, \mathbf{t}_n]$$

Für die Matrix  $Var(\hat{\mathbf{b}})$  der Varianzen und Kovarianzen zwischen den Schätzungen gilt bekanntlich die Beziehung  $Var(\hat{\mathbf{b}}) = \sigma^2(X'X)^{-1}$ ,  $\sigma^2$  die Varianz der Fehler  $e$ .

Man erinnere sich daran, dass Korrelationen zwischen den Prädiktoren (Spalten von  $X$ ) die Existenz kleiner Eigenwerte bedeuten und kleine Eigenwerte die Varianz und Kovarianz der Schätzungen der Komponenten von  $\mathbf{b}$  erhöhen!

29. **PCA-Regression 1** Bei der multiplen Regression bewirken Korrelationen zwischen den Prädiktoren verzerrende Effekte bei Vorhersagen der abhängigen Variablen (also verzerrende Aussagen zB in der Diagnostik). In welchem Sinne hilft hier die PCA-Regression?

Man erinnere die SVD  $X = Q\Lambda^{1/2}T'$ ; durch Welche Matrix ( $Q$  oder  $T$ ) kann  $X$  ersetzt werden und durch welchen kann der Vektor  $\mathbf{b}$  der Regressionsgewichte ersetzt werden?

30. **PCA-Regression 2** Für die PCA-Regression gilt

Da  $X = Q\Lambda^{1/2}T'$ , kann  $X$  durch  $X_r = Q_r\Lambda_r^{1/2}T_r'$ ,  $r < \min(m, n)$  approximieren, – Reduktion der Anzahl der Prädiktoren durch Approximation von  $X$ !

31. **PCA-Regression 3** Da die PCA-Regression ursprüngliche Prädiktoren durch berechnete Prädiktoren  $\mathbf{q}_1, \dots, \mathbf{q}_n$  (Spaltenvektoren von  $Q$ ) ersetzt, kann man sicher sein, dass man unkorrelierte Prädiktoren erhält etc.

32. **Faktorenanalyse 1** Sie haben 70 klinische Psychologen um eine Evaluation der Wirksamkeit der Verhaltenstherapie bei posttraumatischen Störungen von Bundeswehresoldaten befragt. Die Antworten auf die 30 Items des Fragebogens bestanden aus Ratings auf einer 1 bis 7 Skala. Die PCA der Daten zeigt, dass es drei Items gibt, deren repräsentierende Vektoren paarweise orthogonal sind, und der Scree-Test legt die Existenz dreier latenter Variablen nahe.

Die drei Items erleichtern die Definition von entsprechenden drei latenten Variablen für den Fragebogen etc.

33. **Faktorenanalyse 2** Es werde noch einmal das Ergebnis der in der vorangegangenen Aufgabe beschriebenen Evaluation der Verhaltenstherapie betrachtet. Demnach kann man die Hypothese aufstellen, dass der Therapieerfolg  $Y$  als Funktion der drei Dimensionen des Fragebogens vorhergesagt werden kann, d.h. man kann die lineare Regression  $\mathbf{y} = b_1\mathbf{L}_1 + b_2\mathbf{L}_2 + b_3\mathbf{L}_3 + \mathbf{e}$  betrachten.

Man kann hier zB argumentieren, dass die Beziehung  $\mathbf{y} = b_1\mathbf{L}_1 + b_2\mathbf{L}_2 + b_3\mathbf{L}_3 + \mathbf{e}$  als Validitätstest verwendet werden kann (die drei Prädiktoren sind ja unkorreliert etc, keine Kollinearitäten!)

34. **Binäre Items** Die Items des Fragebogens über die Wirksamkeit der Verhaltenstherapie seien in binäre Items umformuliert worden, d.h. man kann den Aussagen nur zustimmen oder sie ablehnen. Die PCA scheint jetzt für die 30 Items sieben statt nur drei latente Variable zu ergeben.

Hier können zB unterschiedliche Schwierigkeiten der Items und damit verbunden ungleiche Randverteilungen für je zwei Items eine Rolle spielen.

35. **Scree 2** Ein Problem bei der PCA ist die Abschätzung der Anzahl der latenten Dimensionen, die einem Datensatz (einer Matrix  $X$ ) unterliegen.

Noch einmal: die wahre Null-Korrelation zwischen allen Variablen kann im Verbund mit Stichprobenfluktuationen in den Daten Stichprobenkorrelationen größer als 1 erzeugen; Sphärizitätstest!

36. **Klassifikation 1** Es werden  $n$  Variable bei  $m$  Fällen gemessen. Jeder Fall gehört einer und nur einer Klasse  $C_k$  an,  $k = 1, \dots, 3$ . Unter welchen Bedingungen erlaubt eine PCA auch eine Klassifikation der Fälle anhand der Koordinaten  $L$  der Fälle?

Man überlege sich: Wenn z.B. die Koordinaten der Fälle in der Klasse  $C_k$ ,  $k = 1, 2, 3$ , in einem Intervall  $I_{jk}$  der  $j$ -ten latenten Dimension liegen und die Anordnung dieser Intervalle auf den latenten Dimensionen für alle Dimensionen dieselbe ist.

37. **Klassifikation 2** Die gemessenen Vektoren  $\mathbf{x}_1, \dots, \mathbf{x}_p$  liegen in einem  $r$ -dimensionalen Teilraum des  $\mathbb{R}^m$ . Welches Kriterium hat Fisher (1936) formuliert, nach dem sich Vektoren  $\mathbf{y}$  finden lassen derart, dass die Projektion der  $\mathbf{x}_j$  auf  $\mathbf{y}$  eine maximale Trennung der Klassen  $C_k$  ermöglicht?

Offenbar kann man die Relation der Quadratsumme der  $y$ -Werte  $QS_{zw}(\mathbf{u})$  "zwischen" den Kategorien und die Quadratsumme der  $y$ -Werte  $QS_{inn}(\mathbf{u})$  "innerhalb" betrachten, wobei  $\mathbf{y} = X\mathbf{u}$ ,  $\mathbf{u}$  der Vektor, dessen Komponenten die Gewichtung der Prädiktoren ist.

38. **Klassifikation 3** Fishers Lineare Diskriminanzanalyse (LDA) beruht auf dem Ansatz, einen Teilraum des von den Prädiktorvektoren  $\mathbf{x}_1, \dots, \mathbf{x}_p$  aufgespannten Vektorraums (der ist durch die lineare Hülle  $\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_p)$  ge-

geben) zu bestimmen, in dem die Fälle als Gruppen mit minimaler Überlappung repräsentiert werden.

Hier erübrigt sich zB die Frage nach einer weiteren Rotation der gefundenen latenten Achsen, – die wurden ja so berechnet, dass die optimale Orientierung bestimmt wird.

39. **Klassifikation 4** Es sei  $W$  die Matrix der Varianzen und Kovarianzen "innerhalb" der Gruppen,  $B$  die entsprechende Matrix "zwischen" den Gruppen. Die dem in der vorangegangenen Aufgabe entsprechenden Vektoren  $\mathbf{u}$  ergeben sich als Lösung der Eigenvektorgleichung  $W^{-1}B\mathbf{u} = \lambda\mathbf{u}$ .

Hier gibt ja keine weiteren einschränkenden Bedingungen, außer der, dass  $\mathbf{u}$  auch tatsächlich berechenbar sein muß, d.h.  $W^{-1}$  und  $B$  müssen existieren.

40. **Korrespondenzanalyse** Abhängigkeiten zwischen Zeilen- und Spaltenkategorien einer Kontingenztafel werden üblicherweise durch einen  $\chi^2$ -Test erfasst, der aber die Abhängigkeiten nicht im Detail reflektiert. Dieser Nachteil kann durch die Korrespondenzanalyse erfasst werden:

Man erinnere sich an die Definition der Matrix, auf die die SVD angewendet wird!.

41. **Korrespondenzanalyse 2** Viele Häufigkeitstabellen sind mehr als 2-dimensional; so hat man die Selbstmorde, die während einer bestimmten Anzahl von Jahren registriert wurden, nach Maßgabe (i) des Geschlechts, (ii) des Alters der Personen und (iii) nach Art der Methode klassifiziert. Die Korrespondenzanalyse geht aber von 2-dimensionalen Häufigkeitstabellen aus.

Man darf sich hier nicht täuschen lassen: es besteht ja immer die Möglichkeit, mehrdimensionale Tabellen durch geschicktes Aneinanderreihen zu 2-dimensionalen Tabellen zu machen! Bestimmte Kategorien tauchen dann mehrfach auf und werden von der Korrespondenzanalyse formal als 'verschieden' interpretiert, – zB die Altersgruppen im Selbstmordbeispiel, oder die Kategorien 'männlich' und 'weiblich'.

42. **Grundsätzliches** Viele der standardmäßig in der Psychologie verwendeten multivariaten Verfahren basieren auf Resultaten der *linearen Algebra*, d.h. es wird implizit angenommen, dass Kovarianzen bzw Korrelationen durch ein rein additives Zusammenwirken latenter Variablen erklärt werden können.

Die meisten nichtlinearen Zusammenhänge können recht gut durch lineare Modelle approximiert werden.