

13.7.05		
Matrikelnummer	Name	Unterschrift

**Aufgabe 1 (Wahrscheinlichkeitsrechnung) 4 Punkte**

Es sei  $\Omega_1 = \{0, 1\}$  mit  $P(0) = 0.4$ ,  $P(1) = 0.6$  und  $\Omega_2 = \{1, 2\}$  mit  $P(1) = 0.3$ ,  $P(2) = 0.7$ . Geben Sie  $\Omega_1 \bullet \Omega_2$  explizit an und bestimmen Sie die Wahrscheinlichkeiten der Elementarereignisse in diesem Wahrscheinlichkeitsraum.

---

---

**Aufgabe 2 (Wahrscheinlichkeitsrechnung) 5 Punkte**

Bei 1000 untersuchten Zellen wurden bei 5 Zellen Chromosomenaberrationen festgestellt.

Welcher Typ von Wahrscheinlichkeitsverteilungen ist geeignet für die Zahl der Aberrationen, wenn Sie vereinfacht davon ausgehen, dass die Schädigungen unabhängig erfolgen?

---

Welchen numerischen Wert für den Parameter würden Sie schätzen? \_\_\_\_\_

Geben Sie mit Hilfe des geschätzten Parameters den natürlichen Logarithmus der Wahrscheinlichkeit dafür an, keine Chromosomenaberration bei den nächsten 1000 Zellen zu finden.

---

### Aufgabe 3 (Wahrscheinlichkeitsrechnung) 5 Punkte

In einer klinischen Studie werden 10 Patienten mit Kolonkarzinom operativ behandelt. Die Wahrscheinlichkeit für eine erfolgreiche Operation (kein Tumorrest mehr nach Operation) ist unbekannt und wird mit  $p$  bezeichnet. Geben Sie die Likelihoodfunktion für den Fall an, dass 8 Patienten geheilt werden.

---

In 25% aller Patienten, die als erfolgreich operiert gelten, wurde in Wirklichkeit Tumorrest übersehen. Wenn dagegen bei einem Patienten Tumorrest gefunden wurde, geht man davon aus, dass dieser Befund fehlerfrei ist, der Patient also nicht erfolgreich operiert wurde. Geben Sie nun die Likelihoodfunktion dafür an, bei 8 von 10 Patienten keinen Tumorrest zu finden.

---

### Aufgabe 4 (Prinzip des Statistischen Tests) 2 Punkte

In zwei verschiedenen Studien (gleiches Signifikanzniveau), die zwei Medikamente verglichen, wurden folgende Heilungsraten beobachtet:

	Studie 1	Studie 2
Medikament A	50%	50%
Medikament B	55%	60%

Das Ergebnis in Studie 2 war mit dem Chi-Quadrat Test nicht signifikant, das Ergebnis in Studie 1 war signifikant. Was muss der Grund gewesen sein?

---

### Aufgabe 5 (Allgemeines Lineares Modell) 2 Punkte

Fügen Sie die Zeichen „+“, „-“, und/oder „=“ so ein, dass die Gleichung korrekt ist:

$$\sum \left( \hat{y}_i - \bar{y} \right)^2 \quad \sum \left( y_i - \bar{y} \right)^2 \quad \sum \left( y_i - \hat{y}_i \right)^2$$

### Aufgabe 6 (Allgemeines Lineares Modell) 2 Punkte

Formulieren Sie das folgende lineare Gleichungssystem in Matrixschreibweise

$$2b_1 + b_2 + b_3 + b_4 = 4$$

$$b_1 + 2b_2 + b_3 + b_4 = 5$$

$$b_1 + b_2 + 2b_3 + b_4 = 6$$

$$b_1 + b_2 + b_3 + 2b_4 = 7$$

### Aufgabe 7 (Allgemeines Lineares Modell / Nichtlineare Regression) 2 Punkte

Das Regressionsmodell

$$Y = \beta_0 + \beta_1 X + \sin(\beta_3 X + \beta_4 X^2) + \varepsilon.$$

ist nicht linear, weil

---

ist linear, weil

---

(Genau eine von beiden Aussagen trifft zu).

### Aufgabe 8 (Allgemeines Lineares Modell) 5 Punkte

Beim Vergleich vier verschiedener Subtypen einer Erkrankung mittels eines linearen Modells wurden die folgenden Kovariablen gewählt:

	$X_0$	$X_1$	$X_2$
Subtyp 1	1	1	0
Subtyp 2	1	1	0
Subtyp 3	1	0	1
Subtyp 4	1	0	0

Formulieren Sie das entsprechende lineare Modell und geben Sie die Interpretation der Parameter an. Bei gleichen Anzahlen in den Gruppen sind welche Kovariablenvektoren paarweise orthogonal?

---

---

---

### Aufgabe 9 (Allgemeines Lineares Modell) 5 Punkte

In einem linearen Modell sollte die Expression eines bestimmten Gens durch die Faktoren Geschlecht, Tumortyp (I; II, III) und Erkrankungsdauer (Monate) erklärt werden.

Die folgenden Modelle wurden betrachtet:

- M0: Nur Intercept
- M1: Intercept, Geschlecht, Tumortyp
- M2: Intercept, Erkrankungsdauer, Tumortyp
- M3: Intercept, Tumortyp, Geschlecht, Wechselwirkung Tumortyp Geschlecht
- M4: Intercept, Tumortyp, Geschlecht, Wechselwirkung Tumortyp Geschlecht, Erkrankungsdauer

Welche Modelle stehen nicht in einem hierarchischen Verhältnis zueinander?

---

Die folgenden Quadratsummen wurden gefunden:

M0:	SSQ (Modell):	8000
	SSQ (Residuen):	4000
M1:	SSQ (Modell, korrigiert)	3000
	SSQ (Residuen)	_____

Geben Sie den Wert für SSQ (Residuen) an.

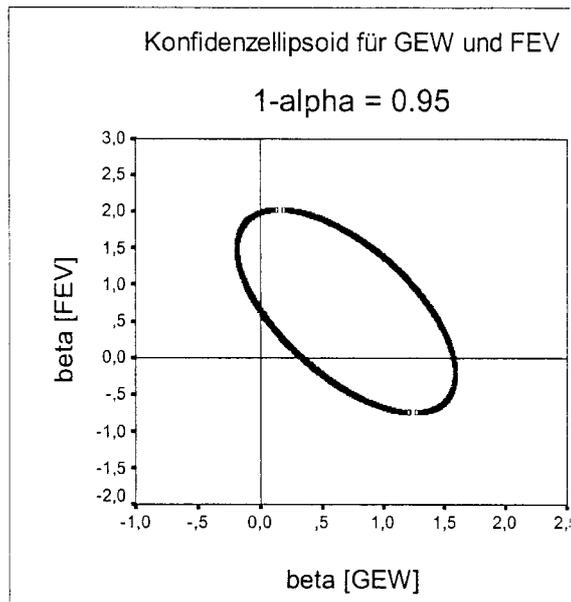
Beim F-Test von Modell M4 gegen Modell M3 wurde ein p-Wert von 0.001 gefunden.

Welcher t-Test in welchem Modell führt exakt zum selben p-Wert?

---

**Aufgabe 10 (Allgemeines Lineares Modell) 2 Punkte**

Welche der genannten Alternativen ist in einem Modell mit dem Konfidenzellipsoid



anzunehmen:

- $H_1: \beta_1 \neq 0 \text{ und } \beta_2 \neq 0$
- $H_1: \beta_1 \neq 0 \text{ oder } \beta_2 \neq 0$

**Aufgabe 11 (Allgemeines Lineares Modell) 5 Punkte**

Für zwei Parameter  $\beta_1 = 30, \beta_2 = 43$  wurde die folgende Varianz-Kovarianzmatrix gefunden:

$$\begin{pmatrix} 25 & -1 \\ -1 & 9 \end{pmatrix}$$

Bestimmen Sie die Standardfehler von  $\beta_1$  und  $\beta_2$ .  $SE(\beta_1) =$   $SE(\beta_2) =$

Bestimmen Sie den Standardfehler von  $\beta_1 - \beta_2$   $SE(\beta_1 - \beta_2) =$

Nennen Sie Nullhypothese und Alternative für den t-Test auf Unterschiedlichkeit von  $\beta_1$  und  $\beta_2$ .

$H_0$  \_\_\_\_\_  $H_1$  \_\_\_\_\_

Ist dieser Test auf dem 5%-Niveau signifikant? Bestimmen Sie die Prüfgröße und verwenden Sie  $t = 2$  als Näherung für den kritischen Wert.

Wert der Prüfgröße \_\_\_\_\_ Signifikanz ja  nein

### Aufgabe 12 (Allgemeines Lineares Modell) 4 Punkte

Bestimmen Sie die Inverse der folgenden Matrix:

$$\begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

Inverse:

### Aufgabe 13 (Verallgemeinertes Lineares Modell) 2 Punkte

In der logistischen Regression lässt sich die Schätzgleichung für  $\beta$  in die Form

$$X'Y = X'\hat{Y}.$$

bringen. Diese Gleichung kann nicht geschlossen gelöst werden, weil

\_\_\_\_\_ nichtlinear von \_\_\_\_\_ abhängt.

### Aufgabe 14 (Verallgemeinertes Lineares Modell) 2 Punkte

Für die Wahrscheinlichkeiten  $p_1 = 0.2$  und  $p_2 = 0.1$  beträgt die odds ratio: \_\_\_\_\_

### Aufgabe 15 (Verallgemeinertes Lineares Modell) 2 Punkte

In einem logistischen Regressionsmodell ist die odds ratio für das Körpergewicht in Pfund (500 Gramm) etwa 1.015. Wie groß ist die odds ratio angenähert für das Körpergewicht in Kilogramm, wenn Sie die Bernoullische Näherung  $(1+x)^n \approx 1+nx$  verwenden?

\_\_\_\_\_

### Aufgabe 16 (Hauptkomponentenanalyse) 1 Punkt

Bei bivariaten Normalverteilungen mit unabhängigen Komponenten mit ungleichen Varianzen liegen die Punkte gleicher Dichte auf

- Kreisen                       Ellipsen

### Aufgabe 17 (Hauptkomponentenanalyse) 3 Punkte

Bei einer Hauptkomponentenanalyse von 5 Variablen auf Basis der Korrelationsmatrix wurden die 5 Eigenwerte

$$\lambda_1 = \underline{\hspace{2cm}} \quad \lambda_2 = 2.0 \quad \lambda_3 = 0.4 \quad \lambda_4 = 0.2 \quad \lambda_5 = 0.1$$

gefunden. Ersetzen Sie den fehlenden Eigenwert.

Bei einer Hauptkomponentenanalyse von 5 Variablen auf Basis der Korrelationsmatrix wurden der Größe nach geordnet die 5 Eigenwerte

$$\lambda_1 = \underline{\hspace{2cm}} \quad \lambda_2 = 2.0 \quad \lambda_3 = 0.6 \quad \lambda_4 = 0.4 \quad \lambda_5 = \underline{\hspace{2cm}}$$

gefunden. Ersetzen Sie die fehlenden Eigenwerte.

### Aufgabe 18 (Faktoranalyse) 4 Punkte

Betrachten Sie die folgende Korrelationsmatrix von 6 Genen einer Expressionsanalyse:

	Gen 1	Gen 2	Gen 3	Gen 4	Gen 5	Gen 6
Gen 1	1	0.95	0.90	0.10	0.12	0.06
Gen 2		1	0.93	0.08	0.13	0.15
Gen 3			1	0.11	0.08	0.06
Gen 4				1	0.95	0.90
Gen 5					1	0.93
Gen 6						1

A Begründen Sie, warum 2 Hauptkomponenten diesen Datensatz angemessen beschreiben.

---

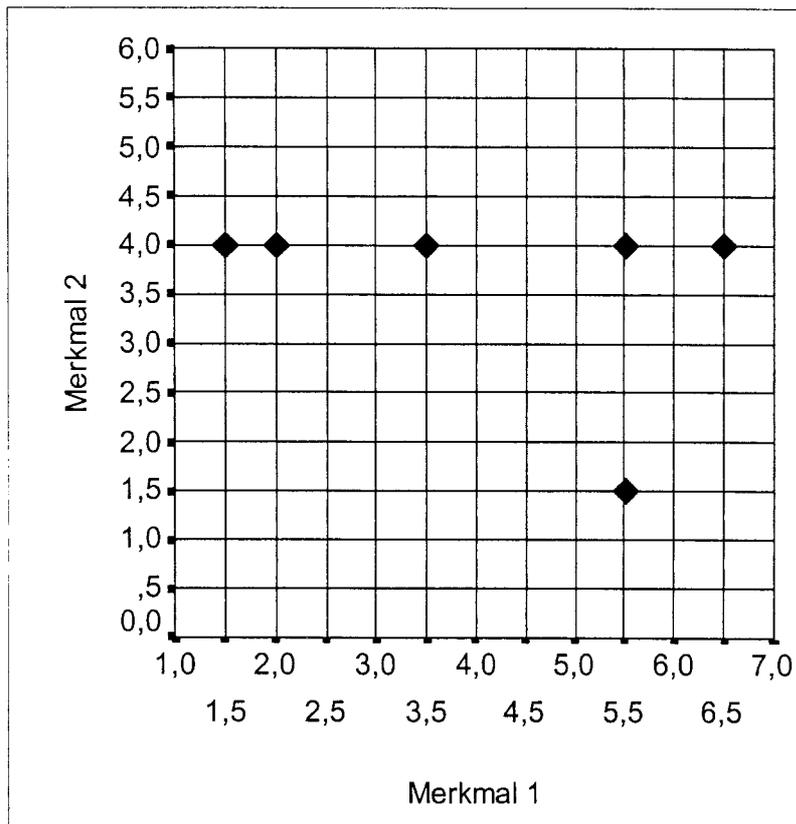
---

B Welche Gene werden für diesen Datensatz nach Faktorrotation jeweils identischen Faktoren zugeordnet?

---

### Aufgabe 19 (Clusteranalyse) 5 Punkte

Betrachten Sie die folgende Graphik. Verwenden Sie den City-Block (=L1) Abstand und führen Sie das single- und das complete linkage Verfahren durch.



### Aufgabe 20 (Multivariate Analyse) 4 Punkte

Entscheiden Sie für jede der genannten Fragestellungen, welches multivariate Verfahren geeignet ist:

- A Für 1000 Notrufe, die zur Einschaltung eines Notarztes führten, sollen typische Muster von Notfällen identifiziert und die einzelnen Notrufe entsprechend gruppiert werden.
- B Ein Fragebogeninstrument zur Lebensqualität enthält 90 Items. Es soll geprüft werden, wie viele unabhängige Achsen die Lebensqualität bestimmen, diese Achsen sollen inhaltlich interpretiert werden.
- C Mit Hilfe von drei verschiedenen Hörtests soll bei Kleinkindern festgestellt werden, ob das Hörvermögen normal oder gestört ist.
- D Aus einer EEG Ableitung mit 20 Kanälen soll diejenige Linearkombination der Einzelergebnisse abgeleitet werden, die am meisten Information über die Daten enthält. Eine inhaltliche Interpretation ist nicht beabsichtigt.

Fragestellung	A	B	C	D
Diskriminanzanalyse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Clusteranalyse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hauptkomponentenanalyse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Faktoranalyse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>