

\_\_\_\_\_

Name, Vorname

|  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|

Matrikelnummer



## Modulklausur 31821 – Multivariate Verfahren

Datum

Punkte

Note

**Termin:** 22. September 2017, 11.30 - 13.30 Uhr

**Erstprüfer:** Univ.-Prof. Dr. H. Singer

## Hinweise zur Bearbeitung der Modulklausur 31821

1. Füllen Sie zunächst den **Kopf des Deckblatts** aus!
2. Es können insgesamt 100 Punkte erreicht werden. Bei Erreichen von 50 Punkten ist die Klausur bestanden. **Bitte kontrollieren Sie sofort, ob Sie ein vollständiges Klausurexemplar erhalten haben.**
3. Zugelassen ist Kurseinheit 1 des Moduls 31821 (Kursnr. 00883) mit farblichen Markierungen, kleinen Aufklebern und/oder textbezogenen Anmerkungen. Nicht zugelassen sind selbst ausgedruckte oder kopierte Kursmaterialien.  
Die Verwendung eines Taschenrechners ist dann und nur dann erlaubt, wenn dieser einer der drei folgenden Modellreihen angehört:
  - Casio fx86 oder Casio fx87
  - Texas Instruments TI 30 X II
  - Sharp EL 531

Eventuelle Vorgänger- oder Nachfolgemodelle, die nicht in der oben aufgeführten Liste enthalten sind, sind ebenfalls nicht erlaubt. Die Verwendung anderer Taschenrechnermodelle wird als Täuschungsversuch gewertet und mit der Note „nicht ausreichend“ (5,0) sanktioniert.

Ob ein Taschenrechner einer der drei Modellreihen angehört, können Sie selbst überprüfen, indem Sie die vom Hersteller auf dem Rechner angebrachte Modellbezeichnung mit den oben angegebenen Bezeichnungen vergleichen: Bei vollständiger Übereinstimmung ist das Modell erlaubt. Ist die auf dem Rechner angebrachte Modellbezeichnung umfangreicher, enthält aber eine der oben angegebenen Bezeichnungen vollständig, ist das Modell ebenfalls erlaubt. In allen anderen Fällen ist das Modell nicht erlaubt.

4. Bitte benutzen Sie für Ihre Rechnungen nur die beigelegten Lösungsbögen.
5. Wenn Sie die einzelnen Blätter der Klausur voneinander trennen, **vermerken Sie auf jedem Blatt Ihre Matrikelnummer**. Legen Sie bitte am Ende der Klausur die Blätter wieder zusammen.

Wir wünschen Ihnen viel Erfolg!

## Aufgabe 1

(15 Punkte)

Die Datenmatrix für eine zufällige Stichprobe der Größe  $N = 3$  aus einer bivariat normalverteilten Grundgesamtheit sei  $\mathbf{X} = \begin{bmatrix} 8 & 3 \\ 6 & 9 \\ 10 & 6 \end{bmatrix}$ . Kennzeichnen Sie die folgenden Aussagen mit **R** für richtig oder **F** für falsch.

Der Mittelwertvektor der Variable  $\mathbf{x}$  ist gleich  $\bar{\mathbf{x}} = [8, 6]'$ .

Die empirische Kovarianzmatrix der Variable  $\mathbf{x}$  ist gleich  $\mathbf{S} = \begin{bmatrix} 4 & 3 \\ 3 & 9 \end{bmatrix}$ .

Es soll die Hypothese  $\boldsymbol{\mu} = \boldsymbol{\mu}_0$  für  $\boldsymbol{\mu}_0 = [9, 5]'$  geprüft werden. Nehmen Sie an, dass die Kovarianzmatrix unbekannt und  $\mathbf{S}^{-1} = \begin{bmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{bmatrix}$  ist. Die Prüfgröße Hotelling- $T^2$  ist dann gleich  $7/9$ .

Es soll die Hypothese  $\boldsymbol{\mu} = \boldsymbol{\mu}_0$  für  $\boldsymbol{\mu}_0 = [9, 5]'$  geprüft werden. Zum Signifikanzniveau von  $\alpha = 0.05$  kann die Nullhypothese nicht abgelehnt werden.

Die Korrelationsmatrix ist gleich  $\mathbf{R} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ . Mit dem Likelihood-Quotienten-Test wird geprüft, ob  $\mathbf{R}=\mathbf{I}$  ist. Mit der Statistik  $T = 0.863$  kann die Nullhypothese  $\mathbf{R}=\mathbf{I}$  zum Signifikanzniveau von  $\alpha = 0.05$  abgelehnt werden.

*Hinweis: Für jede korrekte Kennzeichnung werden 3 Punkte vergeben. Jede falsche Kennzeichnung sowie nicht oder unlesbar gekennzeichnete Felder werden mit 0 Punkten bewertet. Die minimale Punktzahl der Aufgabe beträgt 0 Punkte.*

## Aufgabe 2

(35 Punkte)

Mit einer logistischen Regression soll der Zusammenhang zwischen Geschwindigkeit, Geschlecht des Fahrers und der Schwere von Unfällen untersucht werden. Dazu werden von 7 Unfällen die Variablen Geschwindigkeit (1 für Geschwindigkeit über 100 km/h), Geschlecht (1 für Mann) und Schweregrad des Unfalls (1 für einen schweren Unfall) aufgezeichnet:

| Unfall | Geschwindigkeit | Geschlecht | Schweregrad |
|--------|-----------------|------------|-------------|
| A      | 1               | 0          | 0           |
| B      | 1               | 0          | 1           |
| C      | 1               | 1          | 1           |
| D      | 0               | 1          | 0           |
| E      | 1               | 1          | 1           |
| F      | 1               | 0          | 0           |
| G      | 0               | 0          | 0           |

Gehen Sie im Folgenden davon aus, dass **Geschwindigkeit** und **Geschlecht** die unabhängigen Variablen sind, die auf die abhängige Variable **Schweregrad des Unfalls** wirken (Logistisches Modell).

a) Geben Sie die  $4 \times 4$ -Designmatrix für die vier möglichen Fälle sowohl in Dummy- als auch in Effektkodierung an! Berücksichtigen Sie dabei das Intercept und die Interaktionseffekte! (8 P.)

b) Bestimmen Sie nun die Likelihoodfunktion für ein Modell mit Konstante und Interaktionseffekt! (12 P.)

c) Bei der Parameterschätzung hat der Optimierungsalgorithmus

$$\hat{\beta} = [\hat{\beta}_0 \ \hat{\beta}_1 \ \hat{\beta}_2 \ \hat{\beta}_{1,2}] = [-21 \ 20 \ 0 \ 21]'$$

ergeben (Dummy-Kodierung).  $\hat{\beta}_1$  ist der geschätzte Koeffizient der Variable **Geschwindigkeit**,  $\hat{\beta}_2$  ist der geschätzte Koeffizient der Variable **Geschlecht** und  $\hat{\beta}_{1,2}$  ist der Koeffizient für den Interaktionseffekt. Es ist bekannt, dass bei einem Unfall das Auto von einer Frau mit der Geschwindigkeit von 120 km/h gefahren wurde. Berechnen Sie die Wahrscheinlichkeit, dass es sich um einen schweren Unfall handelt. (4 P.)

d) Für eine Clusteranalyse aller Variablen soll die Ähnlichkeitsmatrix  $\mathbf{S}$  verwendet werden. Bis auf einige Werte ist sie schon bekannt<sup>1</sup>:

$$\mathbf{S} = \begin{bmatrix} 1 & \frac{2}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 1 & \frac{2}{3} \\ \frac{2}{3} & 1 & \frac{2}{3} & 0 & \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & 1 & \frac{1}{3} & s_{C,E} & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 1 & \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & s_{E,C} & \frac{1}{3} & 1 & \frac{1}{3} & 0 \\ 1 & \frac{2}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 1 & s_{F,G} \\ \frac{2}{3} & \frac{1}{3} & 0 & \frac{2}{3} & 0 & s_{G,F} & 1 \end{bmatrix}$$

Bestimmen Sie den  $M$ -Koeffizienten sowie den  $S$ -Koeffizienten für die fehlenden Werte  $s_{C,E}$ ,  $s_{E,C}$ ,  $s_{F,G}$  und  $s_{G,F}$ ! (8 P.)

e) Welche(s) Cluster wird/werden als erstes gebildet? Berücksichtigen Sie, dass es sich um eine Ähnlichkeitsmatrix handelt, nicht um eine Distanzmatrix. (3 P.)

---

<sup>1</sup>Hier werden nur die sieben ursprünglich untersuchten Unfälle geclustert!

### Aufgabe 3

(20 Punkte)



Eine Möbelkette hat mehrere Filialen in verschiedenen Städten. Es muss untersucht werden, ob der Faktor Größe der Stadt einen Einfluss auf die Anzahl der Kunden pro Tag hat. Es wird eine Varianzanalyse mit SPSS durchgeführt, wobei drei Gruppen á 6 Läden (in kleinen, mittleren und großen Städten) untersucht werden. Leider sind einige wichtige Informationen unleserlich.

#### Univariat

##### ONEWAY deskriptive Statistiken

| Kunden | N  | Mittelwert | Standardabweichung | Standardfehler | 95%-Konfidenzintervall für den Mittelwert |            | Minimum | Maximum |
|--------|----|------------|--------------------|----------------|---|------------|---------|---------|
|        |    |            |                    |                | Untergrenze                               | Obergrenze |         |         |
| groß   | 6  | 310,8333   | 109,56353          | 44,72912       | 195,8535                                  | 425,8132   | 175,00  | 500,00  |
| mittel | 6  | 188,3333   | 35,44949           | 14,47220       | 151,1314                                  | 225,5353   | 150,00  | 250,00  |
| klein  | 6  | 90,3333    | 50,36533           | 20,56156       | 37,4782                                   | 143,1885   | 12,00   | 150,00  |
| Gesamt | 18 | 196,5000   | 115,15935          | 27,14332       | 139,2326                                  | 253,7674   | 12,00   | 500,00  |

##### Einfaktorielle ANOVA

| Kunden                | Quadratsumme  | df | Mittel der Quadrate | F   | Signifikanz  |
|-----------------------|---|----|---------------------|---|--|
| Zwischen den Gruppen  |  | 2  | 73230,500           |  |  |
| Innerhalb der Gruppen |  | 15 | 5265,833            |  |  |
| Gesamt                |   | 17 |                     |   |  |

- Bestimmen Sie die Schätzer für den Gesamterwartungswert des Umsatzes, die Erwartungswerte in den einzelnen Gruppen und die drei Effekte! (7 P.)
- Bestimmen Sie die Quadratsummen  $SQE$ ,  $SQT$  und  $SQR$ ! Verwenden Sie dazu die in Aufgabenteil a) geschätzten Effekte und  $\sum_{ij} y_{ij}^2 = 920469$ . (6 P.)
- Prüfen Sie  $H_0 : \alpha_{\text{klein}} = \alpha_{\text{mittel}} = \alpha_{\text{groß}} = 0$  mit dem entsprechenden statistischen Test! Kann die Nullhypothese zum Niveau 5% abgelehnt werden? (7 P.)

## Aufgabe 4

(30 Punkte)

Ein Immobilienunternehmen untersucht, welche Faktoren einen Einfluss auf den Mietpreis ausüben. In der zugrundeliegenden Analyse sollen die folgenden Variablen die Höhe der monatlichen Miete (Variable **QMPreis** in Euro/qm) für Mietwohnungen erklären: **Größe** für die Größe der Wohnung in qm und **U-Bahn** für die Nähe zur nächsten U-Bahn-Station (in km).

Auf der nächsten Seite finden Sie die Ergebnisse einer Regressionsanalyse.

- a) Schreiben Sie die Regressionsgleichung mit nichtstandardisierten Koeffizienten auf! (2 P.)
- b) Schätzen Sie den Parameter  $\sigma^2$  des multiplen Regressionsmodells. (4 P.)
- c) Bestimmen Sie den Determinationskoeffizienten  $R^2$  und interpretieren Sie den Wert! (4 P.)
- d) Vervollständigen Sie die ANOVA-Tabelle und prüfen Sie, ob die Regressoren zur Erklärung der abhängigen Variable beitragen ( $\alpha = 0.05$ ). (6 P.)
- e) Prüfen Sie die einzelnen Regressions-Parameter auf Signifikanz ( $\alpha = 0.05$ ). (6 P.)
- f) Eine Wohnung ist 1 km von der U-Bahn entfernt und 50 qm groß. Prognostizieren Sie die monatliche Miete (Variable **QMPreis** in Euro/qm) und berechnen Sie ein 95%-Prognoseintervall. (8 P.)




Hilfsgröße:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.171 & -0.032 & -0.003 \\ -0.032 & 0.037 & -0.0003 \\ -0.003 & -0.0003 & 0.0001 \end{bmatrix}$$

### Deskriptive Statistiken

|         | Mittelwert | Standardabweichung | N  |
|---------|------------|--------------------|----|
| QMPreis | 5,8953     | 1,41128            | 30 |
| Ubahn   | 1,107      | ,9727              | 30 |
| Größe   | 34,30      | 19,223             | 30 |

### ANOVA<sup>a</sup>

| Modell |                                | Quadratsumme | df | Mittel der Quadrate  | F   | Sig.  |
|--------|--------------------------------|--------------|----|--|---|---|
| 1      | Regression                     | 18,436       | 2  |  |  |  |
|        | Nicht standardisierte Residuen | 39,324       | 27 |  |   |   |
|        | Gesamt                         | 57,760       | 29 |  |   |   |

a. Abhängige Variable: QMPreis

### Koeffizienten<sup>a</sup>

| Modell |             | Nicht standardisierte Koeffizienten |                | Standardisierte Koeffizienten |
|--------|-------------|-------------------------------------|----------------|-------------------------------|
|        |             | Regressionskoeffizient B            | Standardfehler | Beta                          |
| 1      | (Konstante) | 7,273                               | ,499           |                               |
|        | Ubahn       | -,693                               | ,233           | -,478                         |
|        | Größe       | -,018                               | ,012           | -,242                         |



# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

---

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte



# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

---

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte



# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte

# LÖSUNGSBOGEN

|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|

Punkte