

Ihre Version ist     . Bitte vergessen Sie nicht, die Version in den Lösungsbogen zu übertragen. Lösen Sie die Aufgaben zunächst hier (und ggf. auf dem Schmierpapier) und übertragen Sie die Lösungen zum Schluss in den Lösungsbogen. Geben Sie den Lösungsbogen ab, und behalten Sie dieses Aufgabenblatt. Wir wünschen Ihnen viel Erfolg!

*Es gibt unterschiedliche Versionen — Sie finden die richtigen Antworten jeweils unter verschiedenen Buchstaben (a,b,c,d,e), die Antworten sind aber in allen Versionen die gleichen.*

**Aufgabe:**  $X$  ist gleichverteilt (rechteckverteilt) über dem Intervall  $[a, b]$ . Wir interessieren uns für die Obergrenze dieses Intervalls,  $b$ . Das erste Moment der Gleichverteilung ist  $E(X) = (a + b)/2$ . Ihre Stichprobe enthält 9 unabhängige Beobachtungen:  $x_1, \dots, x_9$ . Welche Schätzfunktionen für  $E(b)$  sind erwartungstreu?

(mehrere Antworten möglich, 15 Punkte)

1a:  $x_9$

1b:  $x_1 + x_7 - a$

1c:  $2x_1 + 2x_7 - 2x_9 - a$

1d:  $x_1 - x_7 - x_9 + 2 \cdot a$

1e:  $a + x_1 + x_7 - x_9$

*Diese und die folgende Aufgabe sind eine Variante von Aufgabe 3 aus Wettbewerbsblatt 2.*

**Aufgabe:** Welche Schätzfunktionen für  $E(b)$  sind effizienter?

(mehrere Antworten möglich, 15 Punkte)

2a:  $x_2 - a + x_3$  ist effizienter als  $\frac{1}{2}(4x_1 - 2a)$

2b:  $\frac{1}{2}(4x_1 - 2a)$  ist effizienter als  $2 \cdot x_1 + x_7 - x_6 - a$

2c:  $2 \cdot x_1 - a$  ist effizienter als  $2 \cdot x_1 + x_7 - x_6 - a$

2d:  $-a + \frac{2}{9} \sum_{i=1}^9 x_i$  ist effizienter als  $2 \cdot x_1 + x_7 - x_6 - a$

2e:  $x_2 - a + x_3$  ist effizienter als  $2 \cdot x_1 + x_7 - x_6 - a$

**Aufgabe:** Eine Zufallsvariable ist gleichverteilt (rechteckverteilt) über dem Intervall  $[a, b]$ . Ihre Stichprobe enthält drei Beobachtungen: 1, 2 und 3. Welche Aussagen über den ML (Maximum-Likelihood) Schätzer treffen zu?

(mehrere Antworten möglich, 5 Punkte)

3a: Der ML Schätzer für  $a$  ist 0

3b: Der ML Schätzer für  $b$  ist 3

3c: Der ML Schätzer für  $b$  ist 4

3d: Der ML Schätzer für  $a$  ist 1

3e: Der ML Schätzer für  $a$  ist 2

*Um eine positive Likelihood zu erhalten muss  $\hat{a} \leq 1$  und  $\hat{b} \geq 3$  sein. Je größer  $\hat{a}$  ist, und je kleiner  $\hat{b}$  ist, um so größer wird die Likelihood. Also ist  $\hat{a} = 1$  und  $\hat{b} = 3$ .*

**Aufgabe:** Ihre Stichprobe enthält wieder drei Beobachtungen: 1, 3 und 5. Außerdem wissen Sie, dass  $b - a = 10$  ist. Was ist der Momentenschätzer für  $a$  auf Basis des ersten Moments?

(5 Punkte)

4:	a	b	c	d	e
	anderer Wert	5	-5	<del>7/2</del>	-3

*Das erste Moment (der Mittelwert) ist 3. Das theoretische Moment ist  $(a + b)/2$  oder, nach Annahme  $(2a + 10)/2$ . Wir lösen also  $(2a + 10)/2 = 3$  nach  $a$  auf und erhalten  $a = -2$ .*

**Aufgabe:** Die Länge von Knipps ist normalverteilt mit Standardabweichung 100 und unbekanntem Mittelwert. Eine Stichprobe von 25 Stück Knipps ergibt eine mittlere Länge von 2000. Wie bestimmen Sie die Obergrenze des 99% Konfidenzintervalls für den Mittelwert?

(3 Punkte)

5a:  $2000 - q_{norm}(0.005)$

5b:  $2000 + 20 \cdot q_{norm}(0.005)$

5c:  $2000 - 20 \cdot q_{norm}(0.005)$

5d:  $2000 + 4 \cdot q_{norm}(0.99)$

5e:  $2000 - 100 \cdot q_{norm}(0.005)$

*Die Standardabweichung von  $\bar{X}$  ist  $100/\sqrt{n} = 100/5 = 20$ . Bei einem 99% Intervall bleiben links und rechts*

*0.005 übrig. Da  $q_{norm}(0.005)$  negativ ist, erhält man mit  $2000 - 20 \cdot q_{norm}(0.005)$  die Obergrenze.*

**Aufgabe:** Wie bestimmen Sie die Untergrenze des 90% Konfidenzintervalls für den Mittelwert?

(3 Punkte)

6a:  $2000 - 20 \cdot q_{norm}(0.05)$

6b:  $2000 - 20 \cdot q_{norm}(0.95)$

6c:  $2000 + 4 \cdot q_{norm}(0.1)$

6d:  $2000 - 100 \cdot q_{norm}(0.05)$

6e:  $2000 - q_{norm}(0.005)$

*Bei einem 90% Intervall bleiben links und rechts 0.05 übrig. Da  $q_{norm}(0.95)$  positiv ist, erhält man mit  $2000 - 20 \cdot q_{norm}(0.95)$  die Untergrenze.*

**Aufgabe:** In der Vorlesung haben wir das Verfahren der Varianzanalyse (AOV) behandelt. Welche der folgenden Aussagen trifft für dieses Verfahren zu?

(mehrere Antworten möglich, 5 Punkte)

7a: Es vergleicht die Mediane mehrerer Stichproben

7b: Es betrachtet den Anteil der erklärten Varianz an der nicht erklärten Varianz

7c: Es lässt sich nur anwenden, wenn die erklärende Variable diskret ist

7d: Es ist ein nichtparametrisches Verfahren.

7e: Es lässt sich nur anwenden, wenn die abhängige Variable F-verteilt ist

**Aufgabe:** Die Länge von Knorz ist normalverteilt. Ihre Nullhypothese ist, die mittlere Länge von Knorz sei kleiner oder gleich 12. Ihre Alternativhypothese ist, die mittlere Länge von Knorz sei größer 12. In Ihrer (eher kleinen) Stichprobe der Größe  $n$  messen Sie einen Mittelwert von  $\bar{x}$ . Für  $\bar{x}$  schätzen sie eine Standardabweichung von  $\hat{\sigma}_{\bar{x}}$ . Ihr Signifikanzniveau ist 10%. Sie berechnen eine Teststatistik

$$g = \frac{12 - \bar{x}}{\hat{\sigma}_{\bar{x}}}$$

Sei  $Q_n^t$  die Quantilsfunktion der t-Verteilung mit  $n$  Freiheitsgraden und  $Q^N$  die Quantilsfunktion der Normalverteilung. Sie lehnen die Nullhypothese ab, falls

(mehrere Antworten möglich, 10 Punkte)

8a:  $g > Q_{n-1}^t(0.9)$

8b:  $g > Q_{n-1}^N(0.9)$

8c:  $g < Q_{n-1}^t(0.95)$

8d:  $g > Q_{n-1}^t(0.05)$

8e:  $g < Q_{n-1}^t(0.1)$

**Aufgabe:** Jetzt sei Ihr Signifikanzniveau 1%. Sie lehnen die Nullhypothese ab, falls

(mehrere Antworten möglich, 10 Punkte)

9a:  $g < -Q_{n-1}^t(0.99)$

9b:  $g < Q_{n-1}^t(0.01)$

9c:  $g < Q^N(0.01)$

9d:  $g < Q^N(0.005)$

9e:  $g < -Q^N(0.005)$

**Aufgabe:** Ihr Signifikanzniveau ist immer noch 1%, allerdings ist Ihre Nullhypothese nun, die mittlere Länge von Knorz sei gleich 12. Ihre Alternativhypothese ist nun, die mittlere Länge von Knorz sei ungleich 12. Sie lehnen die Nullhypothese ab, falls

(mehrere Antworten möglich, 10 Punkte)

10a:  $|g| > Q_{n-1}^t(0.995)$

10b:  $-|g| < Q_{n-1}^t(0.005)$

10c:  $-|g| < Q^N(0.005)$

10d:  $-|g| < Q^N(0.01)$

10e:  $|g| < Q_{n-1}^t(0.995)$

**Aufgabe:** Um den Beratungserfolg zu messen, vergleich eine Unternehmensberatung bei 12 Unternehmen den Gewinn jeweils im Jahr vor und nach der Umstrukturierung der von ihr beratenen Unternehmen. Der Gewinn nach der Beratung steht in der Variablen  $x$ . Der Gewinn vor der Beratung steht in  $y$ . Nun führen Sie einen einseitigen Wilcoxon Test durch. Ihr Signifikanzniveau  $\alpha$  ist 1%. Sie erhalten folgendes Ergebnis:

Wilcoxon signed rank test

data: x and y

V = 72, p-value = 0.003418

alternative hypothesis: true location shift is greater than 0

Was ist Ihre Schlussfolgerung?

(mehrere Antworten möglich, 10 Punkte)

- 11a: Die Gewinne nach Beratung sind signifikant größer.
- 11b: Wenn die Alternativhypothese ist, der Gewinn sei nach Beratung größer als vorher, ist es besser, einen zweiseitigen Test durchzuführen.
- 11c: Bei einem größeren Signifikanzniveau ( $\alpha$  wird also größer, das muss nicht besser sein!) würde man keinen signifikanten Unterschied mehr finden.
- 11d: Bei einer so kleinen Stichprobe ist es besser, einen F-Test durchzuführen.
- 11e: Die gemessenen Unterschiede sind klein und deuten nicht auf einen signifikanten Unterschied hin.

**Aufgabe:** Die Unternehmensberatung misst die Steigerung des Gewinns nach Beratung der 12 Unternehmen als  $S$ . Sie betrachten zwei mögliche Faktoren, die den Erfolg der Beratung vielleicht beeinflussen: Dauer der Beratung  $L$  in Stunden und Sektor des Unternehmens. Der Sektor ist codiert als  $D=1$  für Dienstleistung und  $D=0$  sonst. Sie erhalten das folgende Ergebnis: `lm(formula = S ~ L * D)`

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	10	5	2	0.081 .
L	20	5	4	0.004 ***
D	-10	10	-1	0.347
L:D	-5	1	-5	0.001 ***

Ein Unternehmen aus dem Sektor „Dienstleistung“ wird 100 Stunden beraten. Welche Steigerung des Gewinns erwarten Sie? (3 Punkte)

12: 

a	1000	b	2010	c	-490	d	anderer Wert	e	2000
---	------	---	------	---	------	---	--------------	---	------

$100 \cdot (20 - 5) = 100 \cdot 15 = 1500$

**Aufgabe:** Ein Unternehmen aus einem anderen Sektor (nicht „Dienstleistung“) wird ebenfalls beraten. Wie groß ist die marginale Steigerung des Gewinns pro Stunde? (3 Punkte)

13: 

a	-5	b	30	c	10	d	15	e	20
---	----	---	----	---	----	---	----	---	----

Da hier  $D=0$  ist, zählt nur der Koeffizient von  $L$ .

**Aufgabe:** Ihr Signifikanzniveau ist 5%. Welche der folgenden Aussagen treffen zu? (mehrere Antworten möglich, 5 Punkte)

- 14a: Die durch Beratung erzielte Steigerung des Gewinns hängt nicht signifikant vom Sektor ab. *Der Koeffizient von  $L:D$  ist signifikant von Null verschieden.*
- 14b: Die  $t$ -Statistik folgt immer einer  $t$ -Verteilung, egal welcher Verteilung die Störterme folgen. *Wir müssen normalverteilte Störterme voraussetzen, damit die  $t$ -Statistik auch einer  $t$ -Verteilung folgt.*
- 14c: Die obigen Schätzergebnisse zeigen, dass man den Interaktionsterm zwischen  $L$  und  $D$  besser weglassen sollte. *Warum sollte man einen signikanten Term weglassen?*

14d: Wenn man  $L$  aus der Schätzgleichung weglässt, wird das  $R^2$  voraussichtlich sinken. *Das ist immer der Fall, wenn man einen Koeffizienten  $\neq 0$  weglässt.*

14e: Die Dauer der Beratung hat einen signifikanten Einfluss auf den Gewinn. *Der Koeffizient von  $L$  ist signifikant von Null verschieden.*

**Aufgabe:** Sie erheben Daten zu einer abhängigen Variablen  $Y$  und einer unabhängigen Variablen  $X$ . Die Daten sind durch folgende Tabelle gegeben:

X	0	0	2	2
Y	0	4	0	4

Nun schätzen Sie ein lineares Regressionsmodell ohne Konstante:

$$Y_i = \beta_1 X_i + u_i$$

Wie groß ist der OLS Schätzer  $\hat{\beta}_1$ ? (5 Punkte)

15: 

a	-1	b	2	c	4	d	anderer Wert	e	1
---	----	---	---	---	---	---	--------------	---	---

Zeichnen Sie sich die vier Punkte einmal auf und versuchen Sie, jeweils eine Gerade durch die vier Punkte zu legen. An der Stelle  $X = 0$  haben sie (ohne Konstante) keine Wahl. Die Gerade muss durch  $Y = 0$  gehen. Aber an der Stelle  $X = 2$  bekommen Sie den besten „Fit“, wenn Ihre Gerade gerade zwischen den beiden Punkten  $Y = 0$  und  $Y = 4$  hindurchgeht. Also ist  $\hat{\beta}_1 = 1$ .

**Aufgabe:** Nun erweitern Sie ihr Modell um einen konstanten Term:

$$Y_i = \beta_1 X_i + \beta_0 + u_i$$

Wie groß ist jetzt Ihr OLS Schätzer  $\hat{\beta}_1$ ? (5 Punkte)

16: 

a	-1	b	anderer Wert	c	0	d	1	e	2
---	----	---	--------------	---	---	---	---	---	---

Nun können Sie Ihre Gerade auch an der Stelle  $X = 0$  verschieben. Auch hier bekommen Sie den besten „Fit“, wenn Ihre Gerade gerade zwischen den beiden Punkten  $Y = 0$  und  $Y = 4$  hindurchgeht. Die Gerade ist nun waagrecht. Also ist  $\hat{\beta}_1 = 0$ .

**Aufgabe:** Welche Aussagen treffen zu: (mehrere Antworten möglich, 5 Punkte)

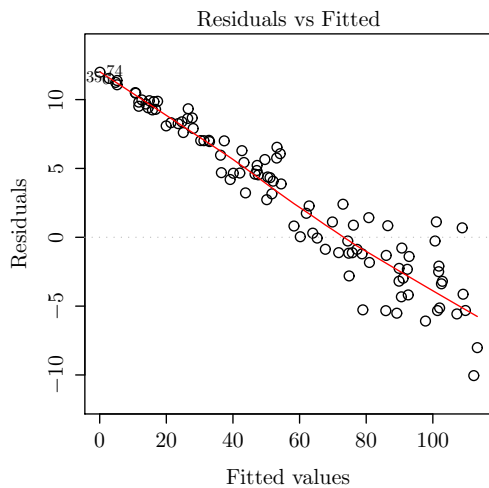
- 17a: Das zweite Modell hat ein kleineres  $R^2$  als das erste.
- 17b: Das zweite Modell hat ein größeres  $R^2$ .
- 17c: Das  $R^2$  in beiden Modellen ist gleich.
- 17d: Die Summe der Fehlerquadrate im zweiten Modell sind größer als im ersten Modell.
- 17e: Die Summe der Fehlerquadrate im zweite Modell sind kleiner als im ersten Modell.

Zeichnen Sie sich die vier Punkte einmal auf und versuchen Sie, jeweils eine Gerade durch die vier Punkte zu legen. Im ersten Modell muss die Gerade irgendwie steigen. Also ist das  $R^2$  positiv. Im zweiten Modell ist die Gerade horizontal, also ist das  $R^2 = 0$ . Da das zweite Modell einen zusätzlichen Parameter hat (und sich die Gerade auch verschiebt) kann die Summe der Fehlerquadrate nur kleiner werden.

**Aufgabe:** Sie schätzen ein lineares Regressionsmodell ohne Konstante

$$Y_i = \beta_1 X_i + u_i$$

und erhalten den folgenden diagnostischen Plot ihrer Regression. Welche Schlussfolgerungen ziehen Sie?



(mehrere Antworten möglich, 10 Punkte)

- 18a:** Zwischen unabhängiger und abhängiger Variablen besteht ein fallender Zusammenhang. *Zwar fallen die Residuen, das hat aber mit der abhängigen Variablen nichts zu tun.*
- 18b:** Die Annahme  $E(u_i|X_i = x) = 0$  ist nicht erfüllt. *Wenn diese Annahme erfüllt wäre, wären die Residuen im Mittel überall Null. Das ist nicht der Fall. Links sind die positiv, rechts negativ.*
- 18c:** Die Annahme,  $\text{var}(u_i|X_i = x)$  ist konstant, ist nicht erfüllt. *Links ist die Varianz der Residuen klein, rechts groß.*
- 18d:** Die Annahme, große Ausreißer in  $X$  und  $Y$  seien selten, ist nicht erfüllt. *Weder hat  $R$  irgendetwas markiert, noch sieht man mit bloßem Auge irgendetwas ungewöhnliches.*
- 18e:** Wenn Sie das Modell um eine Konstante erweitern, wird sich der Fit der Regression verbessern. *An der Graphik sehen wir, dass in den Nähe des Koordinatenursprungs die Residuen etwa 12 sind. Das heißt, die Werte für  $Y$  sind dort auch etwa 12. Also brauchen wir eine Konstante.*

**Aufgabe:** Eine Gruppe von Sportlern bereitet sich auf einen Wettkampf vor. Über die Sportler liegen folgende Informationen vor: Alter ( $A$ ), Geschlecht ( $G$ ; 1 falls weiblich, 0 sonst), tägliches Training ( $T$ , 1 falls ja, 0 sonst), gesunde Ernährung ( $E$ ; 1 falls ja, 0 sonst) und Ranglistenpunkte ( $R$ ). Alter und Geschlecht sind mit den anderen Variablen nicht korreliert. Sie vermuten, dass Sportler nur dann besonders viele Ranglistenpunkte haben, wenn sie täglich trainieren und sich gesund ernähren; ein tägliches Training wirkt nur zusammen mit gesunder Ernährung. Was wären mögliche Spezifikationen des Modells um Ihren Verdacht zu überprüfen. (Hier ist *nicht* nach der „besten“ Spezifikation gefragt)

(mehrere Antworten möglich, 10 Punkte)

- 19a:**  $R = \beta_0 + \beta_1 \cdot T + \beta_2 \cdot E + u$
- 19b:**  $R = \beta_0 + \beta_1 \cdot A + \beta_2 \cdot G + \beta_3 \cdot T + \beta_4 \cdot E + \beta_5 \cdot T \cdot E + u$
- 19c:**  $R = \beta_0 + \beta_1 \cdot T + \beta_2 \cdot E + \beta_3 \cdot T \cdot E + u$
- 19d:**  $T \cdot E = \beta_0 + \beta_1 \cdot R + u$
- 19e:**  $R = \beta_0 + \beta_1 \cdot A + \beta_2 \cdot G + \beta_3 \cdot T + \beta_4 \cdot E + u$

*Erforderlich ist auf jeden Fall die Interaktion  $T \cdot E$ . Dann braucht aber auch  $T$  und  $E$  in der Schätzgleichung (ansonsten würde man mit dem Koeffizienten von  $T \cdot E$  auch den Effekt von  $T$  oder den von  $E$  messen). Man kann  $A$  und  $G$  hinzufügen, um für Alter und Geschlecht zu kontrollieren, muss das aber nicht tun.*

---

maximal erreichbare Punktzahl: 137

davon durch Randomisieren erreichbar: 60.4

hinreichend: 87.9