



3. Übungsblatt

Aufgabe 9	Aufgabe 10	Aufgabe 11	Aufgabe 12	Summe:

Übungsgruppe:

Tutor(in):

Namen:

Aufgabe 9 (Lineare Regression und Prognosen, 5 = 1 + 2 + 2 Punkte).

Ein Forschungsinstitut vermutet einen linearen Zusammenhang zwischen dem Alter (A in Monaten) und dem Durchmesser (D in cm) des Schneckenhauses bei Weinbergschnecken,

$$A = \beta_1 + \beta_2 \cdot D.$$

Das Institut möchte diesen Zusammenhang nutzen, um das Alter von eingefangenen Weinbergschnecken zu bestimmen. Zur Bestimmung von β_1, β_2 werden zunächst einige Daten anhand von gezüchteten Schnecken gesammelt:

Durchmesser in cm	0.78	1.16	1.61	2.23	2.77	3.22	3.73
Alter in Monaten	6.6	4.1	14.4	12.0	15.5	20.9	27.2

Helfen Sie nun bei der Bestimmung des Alters von frisch gefangenen Schnecken mit Hausdurchmessern (in cm):

1.80	3.02	4.52
------	------	------

Gehen Sie wie folgt vor:

- (a) Bestimmen Sie anhand der Daten den KQ-Schätzer $\hat{\beta}$ von $\beta := (\beta_1, \beta_2)'$. Gehen Sie dafür formal vom Modell

$$A_i = \beta_1 + \beta_2 \cdot D_i + \varepsilon_i$$

mit normalverteilten und unabhängigen Messfehlern $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, $\sigma^2 > 0$, aus.

- (b) Bestimmen Sie anhand der Daten ein 0.95-Konfidenzellipsoid für β und fertigen Sie eine maßstabsgetreuen Plot an. *Hinweis:* $F_{2,5,0.95}=5.79$

- (c) Nehmen Sie nun an, dass die Beobachtungen in der zweiten Tabelle neuen Designpunkten $C = (1, \xi)$ mit ξ als Hausdurchmesser entsprechen. Geben Sie zunächst allgemein ein $(1 - \alpha)$ -Konfidenzintervall für die Prognose des Alters $\psi_\xi = C\beta = \beta_1 + \beta_2\xi$ an. Berechnen Sie dann für jede der drei obigen Hausdurchmesser ein 0.95-Konfidenzintervall für das zugehörige Alter.

Hinweis: Verwenden Sie eine direkte Parametrisierung der linearen Regression. Das Beispiel aus der Vorlesung ist nicht direkt anwendbar. Es gilt $F_{1,5,0.95} = 6.61$.

Aufgabe 10 (Vergleichen von Mittelwerten - multiples Testen, 5 = 1 + 1 + 1 + 2 Punkte).

Im Rahmen einer landwirtschaftlichen Versuchsreihe sind wir an den durchschnittlichen Kohlkopferträgen μ_i bei Verwendung $k = 3$ verschiedener Düngemittel ($i = 1, \dots, k$) interessiert. Dazu seien $Y_{ij}, i = 1, \dots, k, j = 1, \dots, N$, Zufallsvariablen, welche die Ernteerträge (in Gramm) des j -ten Kohlkopfs in der zu Dünger i gehörigen Stichprobe beschreiben. Wir nehmen dabei an, dass

$$Y_{ij} = \mu_i + \varepsilon_{ij},$$

wobei $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ unabhängig und normalverteilt mit homogener Varianz $\sigma^2 > 0$ seien. Wir beobachten folgende Kohlkopfgewichte Y_{ij} (in Gramm):

j/i	Dünger 1	Dünger 2	Dünger 3
1	728	974	548
2	955	748	682
3	823	937	763
4	1161	904	617
5	972	869	632
$\bar{Y}_{i\cdot}$	921.2	886.6	648.4

Bearbeiten Sie folgende Aufgaben:

- (a) Ermitteln Sie mittels eines F -Test, ob die Nullhypothese $H_0 : \mu_1 = \mu_2 = \mu_3$ zum Niveau $\alpha = 0.01$ abzulehnen ist oder nicht. Sie brauchen dafür nicht alle theoretischen Größen neu zu berechnen. *Orientieren Sie sich an der Lösung von Blatt 2, Aufgabe 6.*

So genannte Post-Hoc-Tests (oder multipler Vergleichstests) werden verwendet, um die verschiedenen Klassen $i = 1, \dots, k$ in Gruppen von gleichen Mittelwerten einzuteilen. Konkret wollen wir die $\frac{k(k-1)}{2} = 3$ verschiedenen Hypothesen $H_{i,i'} : \mu_i = \mu_{i'}$, testen, wobei $i \neq i'$ für $i, i' \in \{1, 2, 3\}$.

- (b) Zeigen Sie, dass die drei F -Tests $\phi_{1,2}, \phi_{2,3}$ und $\phi_{1,3}$ für das Überprüfen der Hypothesen $H_{1,2}, H_{2,3}$ und $H_{1,3}$ zum Niveau $\alpha' = \alpha/3$ durchgeführt werden müssen, damit die Wahrscheinlichkeit, bei irgendeinem der drei Tests einen Fehler 1. Art zu begehen, durch α beschränkt ist.

Der Test von Tukey bietet eine Alternative zum Vorgehen in (b), bei welchem das Niveau in den einzelnen Teiltests nicht so stark reduziert werden muss. Für ein fest vorgegebenes Niveau α besteht der Test aus der Überprüfung aller $\frac{k(k-1)}{2} = 3$ verschiedenen Hypothesen $H_{i,i'}, i \neq i', i, i' \in \{1, 2, 3\}$. Die Hypothese $H_{i,i'}$ wird abgelehnt, wenn $\frac{|\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}|}{\hat{\sigma} \sqrt{2/N}} > q_{k, (N-1)k, 1-\alpha}$, wobei $\bar{Y}_{i\cdot} := \frac{1}{5} \sum_{j=1}^5 Y_{ij}$ und $q_{k, (N-1)k, 1-\alpha}$ das $(1 - \alpha)$ -Quantil der Verteilung von

$$Z := \frac{\bar{Y}_{\max, \cdot} - \bar{Y}_{\min, \cdot}}{\hat{\sigma} \sqrt{2/N}}$$

bezeichnet, $\bar{Y}_{\max, \cdot} := \max\{\bar{Y}_{i\cdot} : i = 1, \dots, k\}$, $\bar{Y}_{\min, \cdot} := \min\{\bar{Y}_{i\cdot} : i = 1, \dots, k\}$ und $\hat{\sigma}^2 = \frac{1}{(N-1)k} \sum_{i=1}^k \sum_{j=1}^N (Y_{ij} - \bar{Y}_{i\cdot})^2$.

- (c) Zeigen Sie, dass die Verteilung von Z unter H_0 unabhängig von σ und $\mu := \mu_1$ ist.
- (d) Zeigen Sie, dass der Test von Tukey tatsächlich ein Test zum Niveau α ist. Führen Sie den Test mit den obigen Daten zum Niveau $\alpha = 0.01$ durch.

Hinweis: Es ist $q_{3, 12, 0.99} = 3.57$.

Aufgabe 11 (Konfidenzellipsode und -intervalle im Amplitudenmodell, 6 = 1.5 + 1 + 1.5 + 2 Punkte).

Wir beobachten die Amplitude A eines periodischen Signals der Form

$$A_i = \beta_1 \sin(2\pi \cdot t_i) + \beta_2 \cos(2\pi \cdot t_i) + \varepsilon_i, \quad i = 0, \dots, n,$$

wobei $\beta = (\beta_1, \beta_2)$ Amplituden und $t_i := \frac{i}{n}$ die Zeit (in Sekunden) bezeichnet. Wir nehmen an, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ seien unabhängig und identisch verteilte Zufallsvariablen mit $\sigma^2 > 0$.

- (a) Berechnen Sie den KQ-Schätzer $\hat{\beta}$ für β und vereinfachen Sie soweit wie möglich.
Hinweis: $\sum_{i=0}^n \sin^2(2\pi \frac{i}{n}) = \frac{n}{2}$, $\sum_{i=0}^n \cos^2(2\pi \frac{i}{n}) = \frac{n}{2} + 1$, $\sum_{i=0}^n \sin(2\pi \frac{i}{n}) \cos(2\pi \frac{i}{n}) = 0$, $\sum_{i=0}^n \sin(2\pi \frac{i}{n}) = 0$, $\sum_{i=0}^n \cos(2\pi \frac{i}{n}) = 1$.
- (b) Geben Sie ein $(1 - \alpha)$ -Konfidenzellipsoid für β an und skizzieren Sie die Form der Ellipse in einem Koordinatensystem. Nehmen Sie an, dass n groß ist.
- (c) Sei t^* ein weiterer Zeitpunkt. Geben Sie ein $(1 - \alpha)$ -Konfidenzintervall für die Prognose $\psi := \beta_1 \sin(2\pi \cdot t^*) + \beta_2 \cos(2\pi \cdot t^*)$ der Amplitude an.
- (d) Wir beobachten nun für $i = 0, \dots, n = 10$ folgende Amplituden:

i	0	1	2	3	4	5	6	7	8	9	10
A_i	-1.34	-0.44	-0.14	0.92	1.44	0.91	0.46	0.06	-1.05	-1.42	-0.95

Berechnen Sie konkret die Objekte aus (b) und (c) mit $t^* = 2.45$ zum Niveau $\alpha = 0.05$.
Hinweis: $F_{2,9,0.95} = 4.26$ und $F_{1,9,0.95} = 5.12$.

Aufgabe 12 (Anhang 2.1: Ein Daten-Beispiel zur Vorhersage und zum F-Test, 2 = 0.5 + 0.5 + 0.5 + 0.5 Bonuspunkte).

Lesen Sie den Anhang 2.1 (Beispiel 2.13) im Skript auf Seite 26 aufmerksam durch und beantworten Sie folgende vier Fragen:

- (a) Wie groß ist die Anzahl der Stichprobe an intakt gebliebenen männlichen Schädeln, die bei der Farringdon-Street-Reihe verwendet wird?
- (b) Wie lautet das aufgestellte lineare Modell für das Schädelvolumen?
- (c) Zu welchem Schluss kommen die Forscher bei dem Test der Hypothese $H : \beta_1 = \beta_2 = \beta_3$?
- (d) Wie lautet das 0.95-Prognoseintervall für \tilde{V}_0 ?

Abgabe:

In Zweiergruppen, bis spätestens Donnerstag, den **26.11.2020, 11:00 Uhr**.

Bitte geben Sie Ihre Lösungsvorschläge im PDF-Format ab. Es reicht, wenn einer der beiden Personen das Dokument hochlädt.

Homepage der Vorlesung:

<https://ssp.math.uni-heidelberg.de/stat-ws2020/index.html>