

Bachelorprüfung: Statistik (1 Stunde)

Bemerkungen:

- Es sind alle mitgebrachten schriftlichen Hilfsmittel und der Taschenrechner erlaubt.
- Natels sind auszuschalten!
- Lesen Sie zuerst alle Aufgaben durch! Verweilen Sie nicht zu lange bei einem Aufgabenteil, der Ihnen grosse Schwierigkeiten bereitet! Für die Note 6 brauchen nicht alle Aufgaben gelöst zu sein!
- Wenn nicht anders vermerkt, sind die Tests auf dem 5%-Niveau durchzuführen.
- Die nötigen Tabellen befinden sich auf den hintersten Seiten dieser Prüfung.
- Aufgaben 3 und 4 sind Multiple-Choice-Aufgaben. Es ist jeweils genau eine Antwort korrekt. Eine korrekte Antwort gibt 1 **Pluspunkt** und eine falsche Antwort $\frac{1}{2}$ **Minuspunkt**. Minimal erhält man für eine ganze Multiple-Choice Aufgabe 0 Punkte. Tragen Sie die korrekten Antworten der Multiple Choice Aufgaben mit Kreuzchen in das separate Antwortblatt ein.

Viel Erfolg!

1. (6 Punkte)

Im berühmten Wintersportort Weissbergen erwägt man angesichts des Klimawandels, das Schmelzen des Schnees auf den Pisten durch ein künstliches Schneehärtungsmittel (das den Schnee erst bei höheren Temperaturen schmelzen lässt) einzudämmen. Um die Effektivität dieses Mittels zu testen, werden 26 Pisten um Weissbergen herum zufällig ausgewählt, die nach jedem grösseren Schneefall mit dem Härtemittel behandelt werden. Weitere 36 zufällig gewählte Pisten dienen als Kontrollgruppe. Auf all diesen Pisten gibt es bereits eine automatisierte Schneehöhen-Messstation mit integriertem Schneefallsmesser, so dass insgesamt die reine *Abschmelzrate* an jeder Station ermittelt werden kann.

Die Schneeschmelzraten (in mm/Tag) an den Messstationen auf den behandelten Pisten sind X_1, \dots, X_{26} , mit Durchschnitt \bar{X} und Varianz $\hat{\sigma}_X^2 = \frac{1}{25} \sum_{i=1}^{26} (X_i - \bar{X})^2$. Entsprechend werden die Messwerte für die unbehandelten Pisten mit Y_1, \dots, Y_{36} bezeichnet, mit Durchschnitt \bar{Y} und Varianz $\hat{\sigma}_Y^2$. Die gepoolte Varianz über beide Stichproben ist S_{pool}^2 . Ferner nehmen wir an, die Messdaten seien jeweils unabhängig und normalverteilt: die X_i als $\mathcal{N}(\mu_X, \sigma^2)$, die Y_i als $\mathcal{N}(\mu_Y, \sigma^2)$. Ermittelt wurde:

\bar{X}	15.4
$\hat{\sigma}_X^2$	13.8
\bar{Y}	19.7
$\hat{\sigma}_Y^2$	19.7
S_{pool}^2	17.2

- a) Muss ein gepaarter oder ein ungepaarter Test durchgeführt werden? (Begründen Sie!)
- b) Geben Sie die Null- und Alternativhypothese eines t-Tests an, der die Effektivität des Härtemittels beim Bekämpfen der Schneeschmelze testet.
- c) Berechnen Sie die Teststatistik für den t -Test aus (b). Zu welchem Schluss kommt dieser Test?
- d) Würden Sie anhand dieser Ergebnisse zusätzlich einen Vorzeichentest empfehlen? Weshalb, respektive weshalb nicht?

2. (6 Punkte)

An einer Universität studieren sehr viele Studenten Medizin. Ziel dieser Aufgabe ist es, anhand einer Stichprobe von 22 Medizinstudenten den Männeranteil abzuschätzen. Für die ersten beiden Teilaufgaben gehen wir davon aus, dass an dieser Universität 50% der Medizinstudenten Männer sind.

- a) Welche Verteilung besitzt die Anzahl Männer in dieser Stichprobe? Berechnen Sie Erwartungswert und Varianz dieser Verteilung.
- b) Berechnen Sie die Wahrscheinlichkeit, dass in dieser Stichprobe 10 Männer sind.

Eine Auszählung ergab, dass von den 22 Medizinstudenten 6 Frauen sind. Ist diese Häufigkeit vereinbar mit der bisher angenommenen Hypothese, dass die Hälfte der Medizinstudenten an dieser Universität Männer sind?

- c) Führen Sie dazu einen Test auf dem 5% Niveau durch. Benützen Sie dazu die Normal-Approximation. Formulieren Sie dabei Null- und Alternativhypothese, Teststatistik, Verwerfungsbereich und Testentscheid.

3. (10 Punkte)

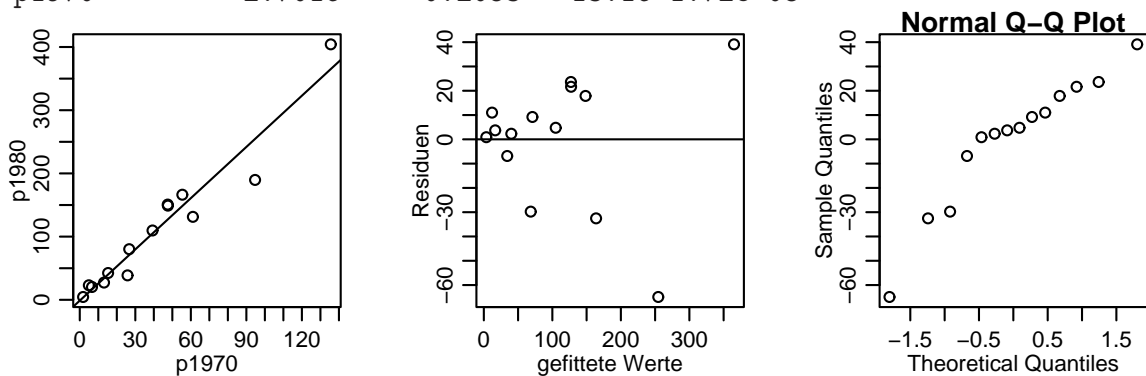
Die Preise für Meeresfrüchte und Fische sind in den USA zwischen 1970 und 1980 stark gestiegen. In einer Studie werden die Preise (in Cent), die ein Fischer im Jahr 1970 und im Jahr 1980 für ein Pfund Ware erhielt, verglichen. Der Datensatz enthält die Angaben zu 14 Fischarten. Es wird eine lineare Regression durchgeführt. Die erklärende Variable ist der Preis im Jahr 1970, und die Zielvariable ist der Preis im Jahr 1980. Das Modell ist

$$p1980_i = \beta_0 + \beta_1 \cdot p1970_i + E_i, \quad E_i \sim \mathcal{N}(0, \sigma^2) \quad \text{i.i.d.}$$

Hier sind der R-Output und einige Plots:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.2338	11.2579	-0.11	0.915
p1970	2.7016	0.2053	13.16	1.72e-08 ***



- 1) Welche der folgenden Aussagen ist richtig?
 - a) Die Modellannahmen über die Fehler scheinen plausibel.
 - b) Der Normalplot deutet auf eine Zunahme der Fehlervarianz mit \hat{y}_i hin.
 - c) Der Tukey-Anscombe Plot deutet auf eine Zunahme der Fehlervarianz mit \hat{y}_i hin.
 - d) Die Plots für die Residuenanalyse sehen nicht so gut aus. Das macht jedoch nichts, da die Modellanpassung ziemlich gut aussieht.
- 2) Was sollte man machen, wenn man bei der Residuenanalyse feststellt, dass die Varianz des Fehlers linear mit \hat{y}_i zunimmt?
 - a) Wenn die Modellanpassung trotzdem gut aussieht, kann man das Modell ohne weiteres verwenden.
 - b) Man muss das Modell abändern. Man sollte die Zielvariable logarithmieren und ausprobieren, ob die Fehler nun besser verteilt sind.
 - c) Wenn der Normalplot gut aussieht, gibt es keinen Grund das Modell zu ändern.
 - d) Man muss das Modell ändern. Variablentransformationen genügen nicht. Man braucht einen ganz anderen Ansatz (keine lineare Regression).

Sowohl die Zielvariable wie auch die erklärende Variable werden logarithmiert (natürlicher Logarithmus). Das neue Modell sieht so aus:

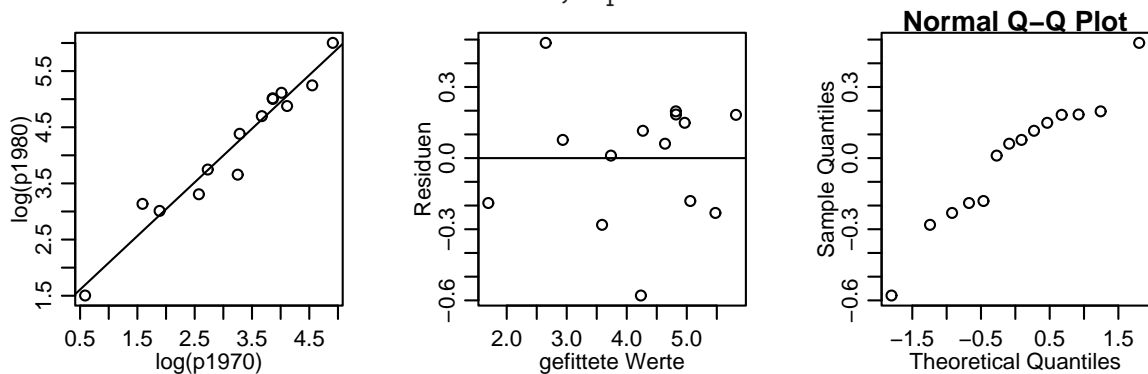
$$\log(p1980_i) = \beta_0 + \beta_1 \cdot \log(p1970_i) + E_i, \quad E_i \sim \mathcal{N}(0, \sigma^2) \quad \text{i.i.d.}$$

Hier sind der R-Output und einige Plots:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.13231	0.21664	5.227	0.000212	***
log(p1970)	0.95465	0.06349	???	???	???

Residual standard error: 0.2776 on ?? degrees of freedom
 Multiple R-Squared: 0.9496, Adjusted R-squared: 0.9454
 F-statistic: 226.1 on 1 and ?? DF, p-value: 3.782e-09



Die Fragen 3 bis 9 beziehen sich auf das neue Modell.

- 3) Wieviele Freiheitsgrade hat die geschätzte Standardabweichung für den Fehler $\hat{\sigma}$?
 a) 12 b) 13 c) 2 d) 14 e) 15 f) 16
- 4) Wie gross ist die t-Teststatistik für den Test der Nullhypothese $H_0 : \beta_1 = 0$?
 a) 0.067 b) 15.036 c) 3.789
 d) 0.063 e) 0.891 f) 23.682
- 5) Wird $H_0 : \beta_1 = 0$ auf dem 5% Niveau verworfen (die Alternative ist $H_A : \beta_1 \neq 0$)?
 a) Ja b) Nein c) Keine Aussage möglich
- 6) Welches der folgenden Intervalle ist ein exaktes zweiseitiges 95% Vertrauensintervall für β_1 ?
 a) $0.95465 \pm 1.96 \cdot \frac{0.06349}{\sqrt{12}}$ b) $0.95465 \pm 2.179 \cdot \frac{0.06349}{\sqrt{12}}$
 c) $0.95465 \pm 1.96 \cdot 0.06349$ d) $0.95465 \pm 2.179 \cdot 0.06349$
- 7) Im R-Output enthält die Spalte Pr(>|t|) für die Zeile log(p1970) den p-Wert für $H_0 : \beta_1 = 0$ und ...
 a) $H_A : \beta_1 < 0$ b) $H_A : \beta_1 > 0$
 c) $H_A : \beta_1 \neq 0$ d) Keine Aussage möglich
- 8) Die Fischer erhielten im Jahr 1970 42.30 Cent für den Lachs. Welchen Preis sagt das Regressionsmodell für den Lachs im Jahr 1980 voraus?
 a) 4.70 b) 41.50 c) 3.75 d) 119.20 e) 110.75

9) Welche der folgenden Aussagen über $R^2 = 0.9496$ stimmt?

- a) Das Regressionsmodell ist mit einer Wahrscheinlichkeit von 5.46% falsch.
- b) Das Regressionsmodell erklärt die Variabilität der Daten mit einer Wahrscheinlichkeit von 94.96%.
- c) Das Regressionsmodell erklärt wenig der totalen Variabilität in den Daten.
- d) Das Regressionsmodell erklärt 94.96% der totalen Variabilität in den Daten.

10) Für die letzte Aufgabe passen wir das Modell nochmals an. Der Fischpreis von 1980 soll anhand der Preise im Jahr 1978 und 1979 erklärt werden. Das Modell sieht so aus:

$$p1980_i = \beta_0 + \beta_1 \cdot p1978_i + \beta_2 \cdot p1979_i + E_i, \quad E_i \sim \mathcal{N}(0, \sigma^2) \quad \text{i.i.d.}$$

Hier ist der R-Output:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.721	6.128	1.260	0.234
p1978	-3.826	6.191	-0.618	0.549
p1979	4.014	6.192	0.648	0.530

Residual standard error: 15.49 on ?? degrees of freedom

Multiple R-Squared: 0.9816, Adjusted R-squared: 0.9783

F-statistic: 294.1 on 2 and ?? DF, p-value: 2.827e-10

Wird die Zielgrösse mindestens von einer der beiden erklärenden Variablen (p1978 und p1979) signifikant beeinflusst?

- a) Ja.
- b) Nein.
- c) Das kann man nur anhand von zwei einfachen Regressionen entscheiden.
- d) Keine Aussage möglich.

4. (6 Punkte)

Das radioaktive Teilchen U-239 zerfällt spontan nach einer Zeit T , gemessen in Sekunden. Man weiss, dass $T \sim \text{Exp}(\lambda)$ mit $\lambda = 0.0004916$.

- 1) Wie gross ist die Wahrscheinlichkeit, dass ein Teilchen erst nach 1000 oder mehr Sekunden zerfällt?

a) 0.6106	b) 0.6116	c) 0.7106
d) 0.5106	e) 0.5	f) Aussagen a) bis e) sind falsch

- 2) Die Zeit t_H , die vergeht, bis die Wahrscheinlichkeit genau 50 Prozent beträgt, dass das Teilchen noch nicht zerfallen ist, heisst (stochastische) Halbwertszeit. Welche der untenstehenden Aussagen gilt also für t_H ?

a) $P[T \leq 0.5] = t_H$	b) $P[T = 0.5] = t_H$
c) $P[T = t_H] = 0.5$	d) $P[T \leq t_H] = 0.5$
e) $P[T = t_H] \leq 0.5$	f) $P[T = 0.5] \leq t_H$

- 3) Berechnen Sie nun t_H , gerundet auf Sekunden.

a) 0.5 Sekunden	b) 1350 Sekunden
c) 1410 Sekunden	d) 120 Sekunden
e) 190 Sekunden	f) Aussagen a) bis e) sind falsch

Sei $X \sim \mathcal{N}(3, 9)$ und $Y = 2 + \frac{1}{3} \cdot X$.

- 4) Berechnen Sie $\text{Var}(Y)$.

a) 0	b) $\frac{1}{3}$	c) 0.5
d) 0.75	e) 1	f) Aussagen a) bis e) sind falsch

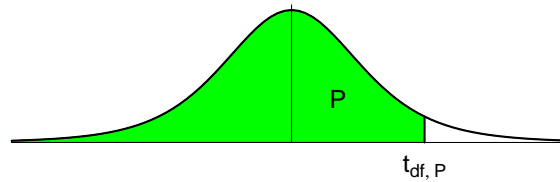
- 5) Berechnen Sie $P[Y \geq 3]$.

a) 0	b) $\frac{1}{3}$	c) 0.5
d) 0.75	e) 1	f) Aussagen a) bis e) sind falsch

- 6) Wie gross ist die Korrelation zwischen X und Y ?

a) 0	b) $\frac{1}{3}$	c) 0.5
d) 0.75	e) 1	f) Aussagen a) bis e) sind falsch

Perzentile der t-Verteilung



Bsp.: $t_{9; 0.975} = 2.262$

df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
31	0.255	0.530	0.853	1.309	1.696	2.040	2.452	2.744
32	0.255	0.530	0.853	1.309	1.694	2.037	2.449	2.738
33	0.255	0.530	0.853	1.308	1.693	2.035	2.445	2.733
34	0.255	0.529	0.852	1.307	1.691	2.032	2.441	2.728
35	0.255	0.529	0.852	1.306	1.690	2.030	2.438	2.724
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576