

Bachelorprüfung: Mathematik 4 - Statistik (2 Stunden)

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 4 und 5 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. (9 Punkte)

Markus fährt manchmal mit dem Fahrrad und manchmal mit dem Tram von seinem Arbeitsort nach Hause. Jetzt will er endlich herausfinden, ob er mit dem Fahrrad schneller zu Hause ist als mit dem Tram. Hierzu hat er von seinen letzten 8 Fahrrad- und von seinen letzten 8 Tramfahrten die Zeit in Minuten notiert.

Fahrrad (x_i)	16.9	15.9	15.4	13.6	16.9	10.4	13.0	13.2
Tram (y_i)	16.0	18.2	13.7	17.2	16.9	18.8	21.1	16.6

Kennzahlen:

$$\begin{aligned} \bar{x} &= 14.4 & \widehat{\sigma}_x^2 &= 5.1 \\ \bar{y} &= 17.3 & \widehat{\sigma}_y^2 &= 4.7 \\ S_{pool} &= 2.2 \end{aligned}$$

Als Modell für die Fahrzeiten nehmen wir eine Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ an.

- a) Begründen Sie kurz, wieso es sich hier um ungepaarte Stichproben handelt.
- b) Geben Sie die Null- und die Alternativhypothese an. Muss hier ein- oder zweiseitig getestet werden?
- c) Führen Sie den geeigneten t-Test durch: Bestimmen Sie den Wert der Teststatistik, den Verwerfungsbereich und den Testentscheid.
- d) Bestimmen Sie ein einseitiges 95%-Vertrauensintervall für $\mu_x - \mu_y$.
- e) Angenommen es sei $\mu_x = 15$ und $\mu_y = 18$. Zudem seien die Varianzen bekannt: $\sigma_x^2 = 3$, $\sigma_y^2 = 4$. Wie gross ist dann die Wahrscheinlichkeit, dass Markus mit dem Fahrrad mindestens 7 Minuten früher zu Hause ist? Es kann davon ausgegangen werden, dass die beiden Fahrzeiten unabhängig voneinander sind.

2. (10 Punkte) Es ist bekannt, dass nur 3% aller perlbildenden Muscheln perfekt runde Perlen produzieren. Wir nehmen an, dass die Perlenbildung von Muschel zu Muschel unabhängig ist und dass pro Muschel höchstens eine Perle auftritt. Ein Taucher holt nun 100 Muscheln einer perlbildenden Sorte aus dem Meer. Die Anzahl X der perfekt runden Perlen, die sich in diesen Muscheln befinden, folgt also der Verteilung $Binomial(100, 0.03)$.

- a) Man kann die Verteilung von X auch durch geeignete Poisson- oder Normalverteilungen approximieren. Wie lauten in diesen Fällen die Parameter der beiden approximierenden Verteilungen?
- b) Der Taucher öffnet seine 100 Muscheln und findet insgesamt lediglich 2 perfekt runde Perlen. Wie gross ist die Wahrscheinlichkeit, höchstens so viele Perlen vorzufinden, d.h. wie gross ist $P[X \leq 2]$? Welchen Wert hat $P[X \leq 2]$ für die beiden in a) ermittelten Approximationen?
- c) Der Taucher will testen, ob bei einer anderen Muschelart weniger häufig perfekt runde Perlen auftreten. Dazu holt er 200 Muscheln der anderen Art aus dem Meer. Sei Y die Anzahl der darin enthaltenen perfekt runden Perlen. Formulieren Sie die Null- und die Alternativhypothese. Bestimmen Sie den Verwerfungsbereich des Tests und zwar sowohl für die exakte Binomial-Verteilung wie auch für die Normal-Approximation mit vorgegebener Varianz 5.82.

3. (11 Punkte)

Es war einmal vor langer Zeit in einem fernen Land, in dem alle Goldtaler das gleiche Gewicht haben, ein Märchenkönig, der hatte eine wunderschöne Tochter in heiratsfähigem Alter. Er machte im ganzen Land bekannt, dass er für die Hand seiner Tochter 1000 Goldtaler verlange. So geschah es denn, dass sich fünf Jünglinge mit je einem Sack voller Goldtaler im Thronsaal des Königs einfanden. Der König hatte jedoch in Erfahrung gebracht, dass *genau* einer der fünf Bewerber ein Betrüger war, und ihm heimlich nur 999 Goldtaler für seine Tochter bezahlen wollte. So liess der König die fünf Bewerber zufällig in einer Reihe aufstellen und wollte herausfinden, wer der Betrüger ist. Dazu bestimmte er das Sollgewicht von 1000 Goldtalern mit einer Waage und versuchte anschliessend durch Wägen der Goldsäcke der Jünglinge den Betrüger zu ermitteln. Wir wollen annehmen, dass jeder der fünf Jünglinge mit derselben Wahrscheinlichkeit $1/5$ der Betrüger sein könnte.

- a) Die erste Idee des Königs ist die folgende: Er schreitet die Reihe der Jünglinge von vorne nach hinten ab und lässt die Goldstücke in den jeweiligen Goldsäcken von einem Gehilfen wägen. Er geht also zunächst zum ersten Jüngling und lässt überprüfen, ob sein Sack nur 999 Goldtaler enthält, wenn ja, hat er den Betrüger gefunden, wenn nein, geht er zum zweiten Jüngling usw. Für jeden Sack, den der Gehilfe wägen muss, schuldet ihm der König einen Lohn von 20 Goldtalern. Welches ist der erwartete Lohn, den der König dem Gehilfen zu bezahlen hat, bis der Schuldige gefunden ist? Berechnen Sie auch die Standardabweichung dieses Lohnes.

Hinweis: Finden Sie zunächst die Verteilung der benötigten Anzahl Wägungen X . Überlegen Sie sich dazu, wie oft der Gehilfe je nach Falle des Täters wägen muss.

Täter ist Jüngling	Wert von X
1	?
2	?
3	?
4	?
5	?

- b) Der schlaue Hofberater des Königs schlägt ihm aber eine andere Strategie vor. Er soll zunächst die Säcke der ersten drei Jünglinge *gemeinsam* auf die Waage stellen. Wenn das Gewicht der drei Goldsäcke zusammen zu klein ist, dann weiss er, dass er den Betrüger unter den ersten drei Jünglingen zu suchen hat. Ist dies der Fall, so soll er als nächstes die Säcke der ersten beiden Jünglinge *gemeinsam* auf die Waage stellen und schliesslich denjenigen des ersten Jünglings *alleine*. Ist dies aber nicht der Fall, so weiss der König, dass der Schuldige entweder der vierte oder fünfte Jüngling sein muss, und braucht dann nur noch den Goldsack des vierten Jünglings *alleine* zu wägen, um den Betrüger zu finden. Falls der König sich für die Strategie des Hofberaters entscheidet, muss er allerdings dem Hofberater 50 Goldtaler Beratungshonorar bezahlen und bei jeder einzelnen Wägung eines Sackes durch den Gehilfen werden zusätzlich wieder 20 Goldtaler Lohn fällig. Berechnen Sie auch für diese Strategie die erwarteten Ausgaben des Königs und deren Standardabweichung.
- c) Welche der beiden Strategien würden Sie verfolgen, wenn Sie KönigIn wären, und warum?

4. (11 Punkte)

Ein Meteorologe untersucht, wie die Jahresdurchschnittstemperatur (in °C) von der Höhe (Meter über Meer) abhängt. Dazu verwendet er die Daten von den Messstationen in der Schweiz und rechnet eine lineare Regression mit der Höhe als erklärende Variable und der Temperatur als Zielvariable. Das Modell ist

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

Hier sind der R-Output und Plots:

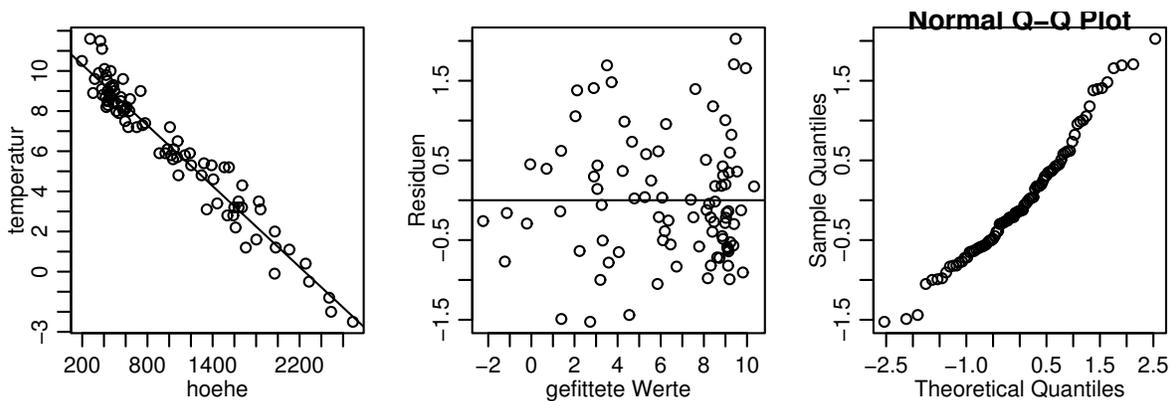
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	11.3186097	0.1536831	???	???
hoehe	-0.0050398	0.0001311	???	???

Residual standard error: ??? on 87 degrees of freedom

Multiple R-Squared: 0.9444, Adjusted R-squared: 0.9437

F-statistic: 1477 on 1 and 87 DF, p-value: < 2.2e-16



- Mit wievielen Beobachtungen wurde die Regression berechnet?
 - 85
 - 87
 - 89
 - 1475
 - 1477
 - 1479
- Wie gross ist die t-Teststatistik für den Test der Nullhypothese $H_0 : \beta_1 = 0$?
 - 0.9444
 - 1477
 - 0.0050398
 - 11.3186097
 - 73.65
 - 38.44
- Wird $H_0 : \beta_1 = 0$ auf dem 5% Niveau verworfen (die Alternative ist $H_A : \beta_1 \neq 0$)?
 - Ja
 - Nein
 - Keine Aussage möglich
- Welches der folgenden Intervalle ist ein exaktes zweiseitiges 95% Vertrauensintervall für β_1 ?
 - $-0.0050398 \pm 1.96 \cdot 0.0001311$
 - $-0.0050398 \pm 1.99 \cdot 0.0001311$
 - $-0.0050398 \pm 1.96 \cdot \frac{0.0001311}{\sqrt{87}}$
 - $-0.0050398 \pm 1.99 \cdot \frac{0.0001311}{\sqrt{87}}$
- Welche durchschnittliche Jahrestemperatur sagt das Regressionsmodell für die Polyterrasse der ETH (Höhe: 450 Meter) voraus?
 - 9.05
 - 13.59
 - 8.47
 - 11.32
 - 10.89

- 6) Betrachten Sie die gezeigten Plots. Welche der nachfolgenden Aussagen ist zutreffend?
- Die Modellannahmen über die Fehler scheinen plausibel.
 - Die Normalitätsannahme der Fehler ϵ_i scheint grob verletzt zu sein.
 - Die Annahme der konstanten Varianz der Fehler ϵ_i scheint grob verletzt zu sein.
 - Die Normalitätsannahme der Fehler ϵ_i und die Annahme der konstanten Varianz der Fehler ϵ_i scheinen grob verletzt zu sein.
- 7) Was passiert mit dem Absolutbetrag der t-Teststatistik für β_1 , wenn an der Stelle (2600, 10) noch eine Beobachtung hinzugefügt wird?
- Wird kleiner
 - Bleibt gleich
 - Wird grösser
 - Keine Aussage möglich
- 8) Wie gross ist die empirische Korrelation $\hat{\rho}$ zwischen der Temperatur und der Höhe?
- $-1 \leq \hat{\rho} < -0.8$
 - $-0.8 \leq \hat{\rho} < -0.5$
 - $-0.5 \leq \hat{\rho} < 0.5$
 - $0.5 \leq \hat{\rho} < 0.8$
 - $0.8 \leq \hat{\rho} < 1$
 - $\hat{\rho} = 1$
- 9) Wie gross ist die Schätzung $\hat{\sigma}$ von σ ?
- 1.5012
 - 2.2e-16
 - 0.2277
 - 0.7722
- 10) Bei der Berechnung der Regression wurden auch die Daten vom Pilatus (Höhe: 2106m, Temperatur: 1.1°C) verwendet. Wie gross ist das Residuum für diese Beobachtung?
- 1.286
 - 0.395
 - 0.705
 - 1.286
 - 0.395
 - 0.705
- 11) Es wird nun noch die Sonnenscheindauer als weitere erklärende Variable ins Modell aufgenommen (multiple Regression). Welche Aussage ist richtig?
- Die Schätzung des Koeffizienten für die Höhe bleibt unverändert.
 - Die Schätzung $\hat{\sigma}$ für σ wird auf jeden Fall kleiner.
 - Multiple Regression liefert ganz allgemein bis auf kleine numerische Abweichungen jeweils gleiche Parameterschätzungen wie die individuellen einfachen linearen Regressionen.
 - Bei der Schätzung des Koeffizienten für die Höhe im multiplen linearen Modell wird der lineare Einfluss von der Sonnenscheindauer auf die Zielgrösse berücksichtigt.
 - Es stimmt überhaupt keine der obigen Aussagen.

5. (8 Punkte)

Grundlagen der Wahrscheinlichkeitstheorie

Betrachten Sie den Grundraum Ω mit den vier Elementarereignissen ω_1 , ω_2 , ω_3 , und ω_4 . Man weiss, dass

$$P[\{\omega_1, \omega_2\}] = \frac{1}{3}, \quad P[\{\omega_2, \omega_4\}] = \frac{1}{2} \quad \text{und} \quad P[\{\omega_1\}] = p$$

für $p \in [0, \frac{1}{3}]$.

- 1) Berechnen Sie $P[\{\omega_2\}]$ in Abhängigkeit von p .
 - a) $\frac{1}{6} - p$
 - b) $\frac{1}{3} - p$
 - c) $1 - p$
 - d) $\frac{2}{3} - p$
 - e) $\frac{1}{4} - p$
 - f) $2p$
- 2) Für welche Werte von p sind die Ereignisse $A = \{\omega_1, \omega_2\}$ und $B = \{\omega_2, \omega_4\}$ unabhängig?
 - a) $0 \leq p \leq \frac{1}{3}$
 - b) $p = \frac{1}{4}$
 - c) $p = 0$
 - d) $p = \frac{1}{6}$
 - e) $p \in \{0, \frac{1}{6}\}$
 - f) $p = \frac{1}{3}$

Kumulative Verteilungsfunktion

Die Zufallsvariable X hat die kumulative Verteilungsfunktion

$$F(x) = \begin{cases} x^3 & x \in [0, 1] \\ 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases}.$$

- 3) Für den Erwartungswert von X gilt:
 - a) $E[X] = 0$
 - b) $E[X] = 1$
 - c) $E[X] = \frac{3}{5}$
 - d) $E[X] = \frac{1}{4}$
 - e) $E[X] = 3$
 - f) Aussagen a) bis e) sind falsch
- 4) Berechnen Sie $P[F(X) \leq y]$ für $y \in [0, 1]$:
 - a) y
 - b) $y^{\frac{1}{3}}$
 - c) y^3
 - d) $1 - y$
 - e) 0
 - f) 1

Summen von i.i.d Zufallsvariablen

X_1, X_2, \dots, X_{75} sind i.i.d. Zufallsvariablen mit $P[X_i = 0] = \frac{5}{12}$, $P[X_i = 3] = \frac{1}{3}$ und $P[X_i = 4] = \frac{1}{4}$.

- 5) Wie ist $\sum_{i=1}^{75} X_i$ approximativ verteilt?
 - a) $\mathcal{N}(100, 225)$
 - b) $\mathcal{N}(150, 225)$
 - c) $\mathcal{N}(150, 200)$
 - d) $\mathcal{N}(175, 225)$
 - e) Poisson(225)
 - f) Poisson(150)
- 6) Berechnen Sie approximativ $P[\sum_{i=1}^{75} X_i \geq 150 + 1.96 \cdot 15]$.
 - a) 0.05
 - b) 0.25
 - c) 0.1
 - d) 0.01
 - e) 0.005
 - f) 0.025

Konfidenzintervalle

Um die Konzentration μ einer Substanz zu bestimmen, gibt es zwei bekannte Verfahren, die sich deutlich im Aufwand und Güte unterscheiden. Aus finanziellen Gründen sind für das erste Verfahren nur $n_1 = 100$ unabhängige Wiederholungen möglich, beim zweiten deren 1000 ($= n_2$). Die Güte eines Verfahren wird durch die Varianz der Messfehler charakterisiert. Diese sind aus früheren Versuchen bekannt. Es gilt $\sigma_1^2 = 1$ für das erste bzw. $\sigma_2^2 = 11$ für das zweite Verfahren. Ausserdem kann man annehmen, dass die Messfehler etwa normalverteilt sind.

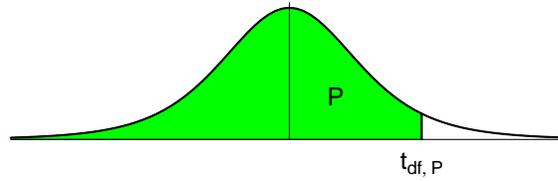
Sie haben also nun die folgenden zwei Varianten, um die wahre Konzentration μ zu schätzen:

A: $\hat{\mu} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$, wobei X_i i.i.d. $\sim \mathcal{N}(\mu, \sigma_1^2)$, $i = 1, 2, \dots, n_1$.

B: $\hat{\mu} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$, wobei Y_i i.i.d. $\sim \mathcal{N}(\mu, \sigma_2^2)$, $i = 1, 2, \dots, n_2$.

- 7) Welche der beiden Varianten ist aus statistischer Sicht besser, d.h. welche liefert das schmalere Konfidenzintervall für μ zum Niveau α .
- Variante A.
 - Variante B.
 - Es sind beide gleich gut.
- 8) Wieviele Wiederholungen w braucht man für das 1. Verfahren, damit die Variante A mit $n_1 = w$ ein halb so breites Konfidenzintervall liefert wie Variante B (mit $n_2 = 1000$)?
- | | | |
|--------|---------|--------|
| a) 400 | b) 360 | c) 180 |
| d) 364 | e) 1000 | f) 182 |

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