

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 **Plus**punkt und eine falsche Antwort $\frac{1}{2}$ **Minus**punkt. 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. (12 Punkte)

Ein Gerät zur Bestimmung der Schwefel-Konzentration ergibt Werte, die $N(\mu, \sigma^2)$ verteilt sind, wobei μ = "wahre Konzentration" und $\sigma = 10$. Alle Konzentrationen sind in der Einheit $\frac{\mu\text{g}}{\text{l}}$ angegeben.

- a) Angenommen, die wahre Konzentration sei 100, wie gross ist die Wahrscheinlichkeit, dass ein Wert > 104 gemessen wird?
- b) Wenn die Konzentration von 25 unabhängigen Proben einer Lösung mit wahrer Konzentration 100 gemessen und gemittelt wird, welche Verteilung hat dann das arithmetische Mittel \bar{X}_{25} ? Wie gross ist die Wahrscheinlichkeit, dass $\bar{X}_{25} > 104$ ist?

Nehmen Sie im Folgenden an, dass das arithmetische Mittel aus den 25 Proben $N(\mu, 2^2)$ -verteilt ist.

- c) Angenommen, das arithmetische Mittel liefert einen Wert von 104. Testen Sie die Nullhypothese " H_0 : Die wahre Konzentration ist 100" gegen die einseitige Alternative " H_A : Die wahre Konzentration ist grösser als 100" auf dem 5%-Niveau.
- d) Bestimmen Sie den Verwerfungsbereich des einseitigen Tests der Nullhypothese $\mu = 100$ gegen die Alternative $\mu > 100$ auf dem 5% Niveau.
- e) Zeichnen Sie eine grobe Skizze der Wahrscheinlichkeitsdichten von \bar{X}_{25} in den beiden Fällen $\mu = 100$ und $\mu = 105$. Zeichnen Sie in diese Skizze die Fläche ein, welche gleich der Wahrscheinlichkeit eines Fehlers 2.Art für $\mu = 105$ ist, und berechnen Sie diese.

2. (15 Punkte) **Bemerkung:** Geben Sie in dieser Aufgabe jeweils den Buchstaben der richtigen Lösung an! Falls eine Rechnung nötig, schreiben Sie diese auch auf! Die Teilaufgaben geben unterschiedlich viele Punkte.

Wird ein neuer Intelligenztest entwickelt, so muss er zuerst normiert werden. Dazu führen Psychologen den Test mit einer repräsentativen Gruppe aus der Bevölkerung durch. Die Psychologen ermitteln dann, wie viele Testaufgaben durchschnittlich gelöst wurden. Diesen Mittelwert definieren sie als IQ mit dem Wert 100. Als nächstes wird die Verteilung des IQ's so standardisiert, dass genau 34,1 Prozent der Getesteten einen IQ von über 115 haben.

Die so resultierende Verteilung X des IQ's ist in guter Näherung $\mathcal{N}(\mu, \sigma)$ verteilt mit $\mu = 100$.

1) Wie gross ist σ ?

- a) 32.1 b) 36.6 c) 30.2 d) 37.9

Achtung: Für den Rest dieser Aufgabe sollen Sie mit $\sigma = 35$ rechnen, auch wenn Sie den wahren Wert für σ richtig berechnen konnten.

Es wird vermutet, dass der IQ von Schweizer Gymnasiasten über diesen 100 liegt. Um dies zu testen wurde in einer Maturaklasse mit $n = 25$ Schülern ein IQ Test durchgeführt. Die Testresultate der einzelnen Schüler $X_i \sim \mathcal{N}(\mu, 35^2)$ sind unabhängig und identisch verteilt. Das Mittel des IQ's dieser Schüler ist \bar{X}_n .

2) Wie lautet die korrekte Null-Hypothese (H_0), respektive Alternativ-Hypothese (H_A)?

- a) $H_0 : \mu = 100$ b) $H_0 : \mu = 100$
 $H_A : \mu > 100$ $H_A : \mu \neq 100$
- c) $H_0 : \mu > 100$ d) $H_0 : \mu \neq 100$
 $H_A : \mu = 100$ $H_A : \mu > 100$

3) Welche Verteilung hat \bar{X}_n unter der Nullhypothese?

- a) $\mathcal{N}(100, \sigma^2/\sqrt{n})$ b) $\mathcal{N}(100, \sigma^2/n)$
- c) $\mathcal{N}(100, \sigma^2 \cdot n)$ d) $\mathcal{N}(100, \sigma^2 \cdot \sqrt{n})$

4) Nehmen Sie an $\bar{X}_{25} = 115$. Testen Sie H_0 gegen H_A , indem Sie einen z-Test auf dem 5% Niveau durchführen.

- a) Der p -Wert beträgt 0.016, daher wird die Nullhypothese auf dem 5%, sowie auf dem 1% Niveau verworfen.
- b) Der p -Wert beträgt 0.016, daher wird die Nullhypothese auf dem 5%, nicht aber auf dem 1% Niveau verworfen.
- c) Der p -Wert beträgt 0.006, daher wird die Nullhypothese auf dem 5%, sowie auf dem 1% Niveau verworfen.
- d) Der p -Wert beträgt 0.006, daher wird die Nullhypothese auf dem 5%, nicht aber auf dem 1% Niveau verworfen.

5) Geben Sie explizit den Annahmebereich des Testes auf dem 5% Niveau an.

- a) $[88.49, 111.51]$ b) $(-\infty, 111.51]$ c) $(-\infty, 88.49]$ d) $[88.49, \infty)$

6) Wie gross ist die Macht des Testes in Teilaufgabe 4, falls die einzelnen IQ's von Gymnasiasten i.i.d. verteilt sind mit $\mathcal{N}(\mu, 35^2)$, wobei $\mu = 120$?

- a) 0.95 b) 0.89 c) 0.99 d) 0.78

7) Wie verändert sich die Macht des Testes, falls das wahre μ grösser als 120 ist?

- a) Sie wird grösser.
 b) Sie wird kleiner.
 c) Es ist keine Aussage diesbezüglich möglich.

8) Der Fehler 2. Art wird kleiner, falls das Niveau α verkleinert wird.

- a) Stimmt immer.
 b) Stimmt nie.
 c) Dies ist abhängig vom wahren Parameter μ .

9) Welche der folgenden Aussagen ist richtig (nur eine stimmt)?

- a) Verwirft man auf dem 5% Niveau, so kann man automatisch auch auf dem 1% Niveau verwerfen.
 b) Die Wahrscheinlichkeit eines Fehlers 1. Art nimmt zu wenn man das Niveau α des Tests verkleinert.
 c) Verkleinert man das Niveau α , so wird der Verwerfungsbereich grösser.
 d) Liegt μ_0 im Vertrauensbereich, so ist der p -Wert des zugehörigen Tests nicht signifikant.

3. (11 Punkte)

Einige Füchse im Kanton Zürich sind mit Alveolärer Echinokokkose erkrankt (Fuchsbandwurm). Um den Prozentsatz der erkrankten Tiere festzustellen, werden Tiere zufällig eingefangen und auf Befall mit Fuchsbandwurm untersucht.

Jemand vermutet, dass 10% aller Füchse mit Fuchsbandwurm befallen sind. Nimm für die ersten drei Teilaufgaben an, dass diese Vermutung zutrifft.

- a) Es werden 10 Füchse zufällig eingefangen. Wie ist die Verteilung der Anzahl eingefangener, erkrankter Füchse (es kann angenommen werden, dass die gesamte Fuchspopulation im Kanton Zürich sehr gross ist).
 b) Wie gross ist die Wahrscheinlichkeit, dass kein einziger erkrankter Fuchs eingefangen wird? Wie gross ist die Wahrscheinlichkeit, dass mindestens zwei der 10 eingefangenen Füchse erkrankt sind?
 c) Mit welcher Verteilung kann die Anzahl erkrankter Füchse unter 200 eingefangenen Füchsen angenähert werden? Gib mithilfe dieser Näherung die kleinste Anzahl x Füchse an, sodass mit Wahrscheinlichkeit 95% (oder grösserer Wahrscheinlichkeit) weniger als x erkrankte Füchse unter den 200 eingefangenen zu finden sind.
 d) Ist die Annahme von 10% (oder weniger) erkrankter Füchse haltbar, wenn 30 der 200 eingefangenen Füchse an Alveolärer Echinokokkose leiden? *Hinweis:* Benutze hierzu das Resultat aus der vorigen Teilaufgabe.

4. (11 Punkte)

- 1) Welches der folgenden Modelle (in denen y durch x erklärt werden soll) gilt in der Statistik als lineares Modell?

Modell 1 $Y = \beta_0 + \beta_1 e^x + E, E \sim N(\mu, \sigma^2)$

Modell 2 $Y = \frac{\beta_0 + \beta_1 x}{\beta_0 + \beta_2 x} + E, E \sim N(\mu, \sigma^2)$

Modell 3 $Y = \beta_0 e^{\beta_1 x} + E, E \sim N(\mu, \sigma^2)$

- a) Nur Modell 1
 b) Nur Modell 2
 c) Nur Modell 3
 d) Modell 1 und Modell 2
 e) Modell 1 und Modell 3
 f) Modell 2 und Modell 3
 g) alle drei
 h) keines
- 2) In Unteraufgaben 2 und 3 geht es um eine lineare Regression, in der Körpergewicht (in kg) durch Körpergrösse (in m) erklärt werden soll. Welche Einheit hat der Achsenabschnitt?
- a) kg
 b) kg*m
 c) \sqrt{kg}
 d) m
 e) einheitenlos
 f) kg/m
- 3) Es geht um die gleiche Regression wie in Unteraufgabe 2. Welche Einheit hat die Steigung der Geraden?
- a) kg
 b) kg*m
 c) \sqrt{kg}
 d) m
 e) einheitenlos
 f) kg/m

Im Folgenden geht es um den Old Faithful Geyser im Yellowstone Nationalpark der USA. Wie der Name schon vermuten lässt, ist der Geyser bekannt für seine Zuverlässigkeit: Die Ausbrüche treten in regelmässigen Abständen auf und es scheint auch ein Zusammenhang zwischen Dauer (x) der einzelnen Ausbrüche und der Intervalldauer (y) bis zum nächsten Ausbruch zu geben. Um dies zu untersuchen, möchte man den Zusammenhang der Variablen mit einem linearen Modell

$$Y_i = \beta_0 + \beta_1 x_i + E_i, \quad E_1, \dots, E_n, iid \sim N(0, \sigma^2)$$

anpassen. Das wäre z.B. für die Parkwächter nützlich, die den Touristen auf einer Tafel gerne die Uhrzeit des nächsten Ausbruchs angeben würden. Dazu wird das Statistikprogramm R verwendet, welches folgende Ausgabe liefert:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	32.432	4.827	6.719	5.98e-07	***
x	11.211	1.368	8.196	2.05e-08	***

Residual standard error: 7.565 on 24 degrees of freedom
 Multiple R-Squared: 0.7367, Adjusted R-squared: 0.7258
 F-statistic: 67.17 on 1 and 24 DF, p-value: 2.053e-08

- 4) Mit wie vielen Beobachtungen n wurde diese Regression gerechnet?
- | | |
|-------|-------|
| a) 24 | b) 25 |
| c) 26 | d) 27 |
| e) 28 | f) 29 |
- 5) Ist es auf dem 5% Niveau statistisch gesichert, dass in diesem Regressionsmodell ein Zusammenhang zwischen der Ausbruchsdauer y und der Intervalldauer x besteht?
- | | | |
|-------|---------|--------------------------|
| a) Ja | b) Nein | c) Keine Aussage möglich |
|-------|---------|--------------------------|
- 6) Kann man aufgrund des p-Wertes von x entscheiden, ob das zweiseitige 99%-Vertrauensintervall für den Koeffizienten β_1 den Wert 0 enthält?
- | |
|--|
| a) 0 liegt im 99%-Vertrauensintervall |
| b) 0 liegt nicht im 99%-Vertrauensintervall |
| c) Kann mit den vorhandenen Angaben nicht entschieden werden |
- 7) Welches ist ein exaktes 95%-Vertrauensintervall für den Koeffizienten β_1 ?
- | | |
|---|---|
| a) $11.21 \pm 1.64 \cdot 1.368$ | b) $11.21 \pm 1.71 \cdot 1.368$ |
| c) $11.21 \pm 1.96 \cdot 1.368$ | d) $11.21 \pm 2.06 \cdot 1.368$ |
| e) $11.21 \pm 1.96 \cdot \frac{1.368}{\sqrt{24}}$ | f) $11.21 \pm 2.06 \cdot \frac{1.368}{\sqrt{24}}$ |
- 8) Wie lange muss man (in Minuten) nach unserem Modell auf den nächsten Ausbruch warten, wenn die Dauer des letzten Ausbruchs 5.0 Minuten betrug?
- | | |
|---------|---------|
| a) 56.6 | b) 88.5 |
| c) 18.1 | d) 11.7 |
- 9) Wofür wird ein Tukey-Anscombe Plot typischerweise im Zusammenhang mit einer linearen Regression verwendet?
- | |
|--|
| a) Optische Überprüfung, ob Residuen normalverteilt sind. |
| b) Verschafft Überblick über die erklärende Variable. |
| c) Optische Überprüfung, ob die Fehlervarianz konstant ist. |
| d) Optische Überprüfung, ob die Beobachtungen unabhängig sind. |
- 10) Wofür wird ein Quantil-Quantil Plot typischerweise im Zusammenhang mit einer linearen Regression verwendet?
- | |
|--|
| a) Optische Überprüfung, ob Residuen normalverteilt sind. |
| b) Optische Überprüfung, ob das Modell linear ist. |
| c) Optische Überprüfung, ob die Fehlervarianz konstant ist. |
| d) Optische Überprüfung, ob die Beobachtungen unabhängig sind. |
- 11) Wie gross ist die Schätzung von σ^2 ?
- | | | |
|----------|----------|----------|
| a) 0.542 | b) 0.736 | c) 0.858 |
| d) 2.750 | e) 7.565 | f) 57.23 |

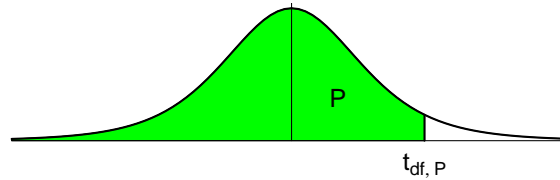
5. (6 Punkte)

Verschiedene Aufgaben aus mehreren Gebieten:

- 1) Mit welcher Verteilung lässt sich am ehesten die Wartezeit an einer Kaufhauskasse beschreiben?
- | | |
|------------------------|---------------------------|
| a) Uniforme Verteilung | b) Poisson-Verteilung |
| c) Binomial-Verteilung | d) Exponential-Verteilung |
- 2) Welche Verteilung eignet sich am ehesten für die Beschreibung der Anzahl Staubteilchen pro Kubikmeter Luft?
- | | |
|------------------------|---------------------------|
| a) Uniforme Verteilung | b) Poisson-Verteilung |
| c) Binomial-Verteilung | d) Exponential-Verteilung |

- 3) Welche Verteilung eignet sich am besten für die Beschreibung der letzten Ziffer einer zufällig aus dem Telefonbuch gewählten Nummer?
- a) Uniforme Verteilung
 - b) Poisson-Verteilung
 - c) Binomial-Verteilung
 - d) Exponential-Verteilung
- 4) Angenommen, 30% der Bevölkerung sind Raucher. Wie gross ist ungefähr die Standardabweichung der Anzahl Raucher unter 100 zufällig ausgewählten Personen?
- a) 5.5
 - b) 30
 - c) 5
 - d) 4.6
- 5) Für die Auswahl der Personen in 4) werden 50 Männer und 50 Frauen ausgewählt. Wenn sich die Raucherraten zwischen Männern und Frauen unterscheiden, so ist jetzt die Standardabweichung der Anzahl Raucher unter den 100 ausgewählten Personen
- a) grösser als in 4).
 - b) kleiner als in 4).
 - c) gleich wie in 4).
- 6) Es soll mit einer Befragung gezeigt werden, dass immer noch 30% oder mehr der Bevölkerung rauchen. Von 100 zufällig ausgewählten Personen sind 25 Raucher. Ein *einseitiger* Test wird durchgeführt. Das zum Test gehörige Vertrauensintervall für den Prozentsatz der rauchenden Bevölkerung ist ungefähr gegeben durch
- a) $[0, 32]$
 - b) $[18, 100]$
 - c) $[32, 100]$
 - d) $[18, 32]$
 - e) $[25, 100]$
 - f) $[0, 25]$

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