

Bachelorprüfung: Mathematik 4 - Statistik (2 Stunden)

Bemerkungen:

- Es sind alle 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!
- 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. (8 Punkte)

Vor wenigen Jahren wurde ein neues Verfahren zur Bestimmung der Masse von Sternen entwickelt, das auf der Allgemeinen Relativitätstheorie beruht. Das klassische Verfahren stützt sich hingegen auf die klassische Mechanik und ist weithin akzeptiert, hat aber den Nachteil einer sehr grossen Ungenauigkeit.

Einige Forscher vermuten, dass das neue Verfahren einen systematischen Fehler hat und wollen es nun mit einem umfangreichen Experiment überprüfen. Dazu bestimmen sie die Masse von zehn Sternen gleichen Typs (also sehr ähnlichen Sternen) im Kugelsternhaufen M80 je einmal mit dem traditionellen und einmal mit dem neuen Verfahren. Die über alle zehn Sterne gemittelten Werte sind (jeweils in Sonnenmassen) $\bar{X} = 1.35$ für das neue und $\bar{Y} = 1.56$ für das traditionelle Verfahren. Die geschätzten Standardabweichungen betragen $\hat{\sigma}_X = 0.18$ sowie $\hat{\sigma}_Y = 0.45$. Die geschätzte Standardabweichung für die Differenz zwischen den beiden Verfahren beträgt $\hat{\sigma}_{X-Y} = 0.46$. Wir nehmen im Weiteren an, dass die einzelnen Werte als unabhängige, normalverteilte Zufallsvariablen ($\mathcal{N}(\mu_X, \sigma_X^2)$ für die neue Methode, $\mathcal{N}(\mu_Y, \sigma_Y^2)$ für die traditionelle Methode) modelliert werden können.

Nun soll geprüft werden, ob die beiden Verfahren zu systematisch unterschiedlichen Resultaten führen.

- a) Handelt es sich um einen gepaarten oder einen ungepaarten Test? Begründen Sie!
- b) Geben Sie die Null- und die Alternativhypothese an.
- c) Geben Sie den Verwerfungsbereich und die Testentscheidung (mit Begründung) für einen t-Test auf dem 5% Niveau an.
- d) Berechnen Sie ein zweiseitiges 95%-Vertrauensintervall für die Differenz $\mu_X - \mu_Y$.
- e) Könnte an Stelle des t-Tests auch ein z-Test durchgeführt werden? Begründen Sie!

2. (9 Punkte)

Eine Restaurantkette in den U.S.A. möchte sich gegen sehr grosse Klagen versichern. Um das Problem zu vereinfachen nehmen wir an, dass die Kosten für jede Klage 500'000\$ betragen und dass die Klagen voneinander unabhängig sind. Die folgende Tabelle fasst die Verteilung der Klagen gegen 60 Restaurants im Jahr 2009 zusammen.

Anzahl Klagen in 2009	0	1	2	3	4
Anzahl Restaurants	54	5	1	0	0

(Zum Beispiel: 5 Restaurants haben im Jahr 2009 eine Klage gehabt, aber nur eines hat eine 2. Klage bekommen.)

- Erklären Sie kurz, wieso für die Anzahl Klagen pro Jahr eine Poissonverteilung ein vernünftiges Modell darstellt.
- Passen Sie eine Poissonverteilung an die Daten an: was ist der Schätzwert vom Parameter λ ?
- Wie gross ist die Wahrscheinlichkeit unter der in **b)** gewählten Poissonverteilung, dass eine Firma mindestens einmal in 2009 angeklagt wurde?
- Wie hoch sind die zu erwartenden Kosten durch Klagen pro Jahr und Restaurant?

Eine Versicherung schlägt jedem Restaurant folgenden Vertrag vor: Für 61'500\$ übernimmt die Versicherung die ganzen Kosten der Klagen (falls es solche gibt).

Die Versicherung hat jetzt 24'000 Restaurants versichert.

- Wie viele Klagen insgesamt erwartet die Versicherung? Die Verteilung dieser Klagen kann durch die Normalverteilung $Anzahl\ Klagen \sim \mathcal{N}(\mu, \mu)$ sehr gut approximiert werden. Was ist eine sinnvolle Wahl für μ ? Warum?
- Wie gross ist die Wahrscheinlichkeit, dass die Versicherung einen Verlust am Ende des Jahres schreiben muss? (Benutze die in **e)** gefundene Approximation)

3. (12 Punkte)

Die low-cost Fluggesellschaft 'Air-Patatrack' verkauft (wie auch viele andere Fluggesellschaften) mehr Flugtickets pro Flug als Sitze im betreffenden Flugzeug vorhanden sind (so genanntes 'overbooking'). Grund für dieses Vorgehen ist, dass Kunden oft kurzfristig auf die Reise verzichten.

Air-Patatrack schätzt, dass 90% der gebuchten Tickets benutzt werden.

- a) Wie gross ist die Wahrscheinlichkeit, dass von vier Personen, genau eine die Reise nicht antritt?

Für den Flug Zürich-Agno benutzt die Gesellschaft ein 'Beechcraft C12-J' mit 26 Passagierplätzen. Für den nächsten Flug sind 28 Plätze gebucht.

- b) Welche Verteilung besitzt die Anzahl der Personen, die den Flug antreten möchten? Berechnen Sie den Erwartungswert und die Varianz dieser Verteilung.
- c) Wie gross ist die Wahrscheinlichkeit, dass nicht alle Passagiere Platz im Flugzeug finden?

Für die Flüge Zürich-Johannesburg benutzt die Air-Patatrack einen Airbus A380 mit 853 Sitzplätze. Für solche Flüge werden 890 Tickets verkauft.

- d) Mit welcher stetigen Approximation kann man die Anzahl Personen, die den Flug nehmen möchten nun einfacher modellieren? (Verteilungsfamilie *und* Parameter angeben).
- e) Man möchte, dass höchstens in 5% der Fälle zu viele Leute am Flughafen sind. Wie viele Tickets sollte man in diesem Fall verkaufen?

Von 890 Personen, die den Flug am 05.08 reserviert haben, sind nur 875 am Flughafen erschienen. Ist die Annahme, dass 90% der gebuchten Tickets benutzt werden angesichts dieser Daten plausibel?

- f) Führen Sie einen zweiseitigen Test auf dem 5% Niveau durch. Geben Sie die Null- und die Alternativhypothese, die Teststatistik sowie den Testentscheid an. (Benutzen Sie die in **d**) bestimmte Approximation).
- g) Ist es (im allgemein) sinnvoll einen Test durchzuführen um eine Nullhypothese zu verifizieren? Warum?

4. (8 Punkte)

In einer klinischen Studie wird untersucht, wie bei Patienten mit Prostata-Krebs der Level eines Prostata Antigens vom Volumen, dem Prostata-Gewicht und dem Alter des Patienten abhängt. Dazu wurde das folgende lineare Modell angepasst:

$$\log(psa_i) = \beta_0 + \beta_1 \cdot \log(cavol_i) + \beta_2 \cdot \log(weight_i) + \beta_3 \cdot age_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \text{ i.i.d.}$$

Der Regressionsoutput sieht wie folgt aus:

Residuals:

Min	1Q	Median	3Q	Max
-2.805	-0.411	-0.006	0.448	2.703

Coefficients:

	Estimate	Std.Error	t value	Pr(> t)
(Intercept)	0.518	0.849	0.611	0.543
logcavol	0.642	???	8.551	2.95e-13
logweight	0.468	0.181	2.579	0.012
age	-0.009	???	???	???

Residual standard error: 0.802 on 90 degrees of freedom

Multiple R-squared: 0.491, Adjusted R-squared: 0.474

F-statistic: 28.9 on 3 and 90 DF, p-value: 3.539e-13

- 1) Wieviele Beobachtungen umfasst der verwendete Datensatz?
 - a) 86
 - b) 90
 - c) 93
 - d) 94
 - e) 97
 - f) 92
- 2) Wie gross ist der Standardfehler der Schätzung von β_1 ?
 - a) 0.075
 - b) 0.152
 - c) 0.181
 - d) 0.087
- 3) Wie gross ist die geschätzte Varianz der zufälligen Fehler $\text{Var } \varepsilon_i$?
 - a) 0.643
 - b) 0.585
 - c) 0.802
 - d) keine Aussage möglich
- 4) Wird die Nullhypothese $H_0 : \beta_2 = 0$ auf dem 1%-Niveau verworfen?
 - a) Ja
 - b) Nein
 - c) keine Aussage möglich
- 5) Welches der folgenden Intervalle ist ein exaktes zweiseitiges 95% Vertrauensintervall für β_2 ?
 - a) $0.468 \pm 1.987 \cdot \frac{0.181}{\sqrt{90}}$
 - b) $0.468 \pm 1.96 \cdot \frac{0.181}{\sqrt{90}}$
 - c) $0.468 \pm 1.987 \cdot 0.181$
 - d) $0.468 \pm 1.96 \cdot 0.181$
- 6) Wird die Nullhypothese $H_0 : \beta_3 = 0$ auf dem 5%-Niveau verworfen?
 - a) Ja
 - b) Nein
 - c) keine Aussage möglich
- 7) Ein Patient ist 58 Jahre alt und seine Prostata wiegt 27 Gramm. Welchen Antigen Level (psa) besitzt er?

a) 12.94

b) 19.12

c) Keine Aussage möglich

8) Betrachten Sie die Plots. Welche der folgenden Aussagen ist zutreffend?

- a) Die Modellannahmen über die Fehler sind exakt erfüllt.
- b) Die Modellannahmen über die Fehler scheinen plausibel. Es gibt aber Ausreisserpunkte.
- c) Die Normalitätsannahme der Fehler scheint grob verletzt zu sein, aber die Annahme der konstanten Varianz ist plausibel.
- d) Die Annahme der konstanten Varianz scheint grob verletzt zu sein, aber die Normalitätsannahme der Fehler ist plausibel.
- e) Die Fehler scheinen nicht unabhängig zu sein.

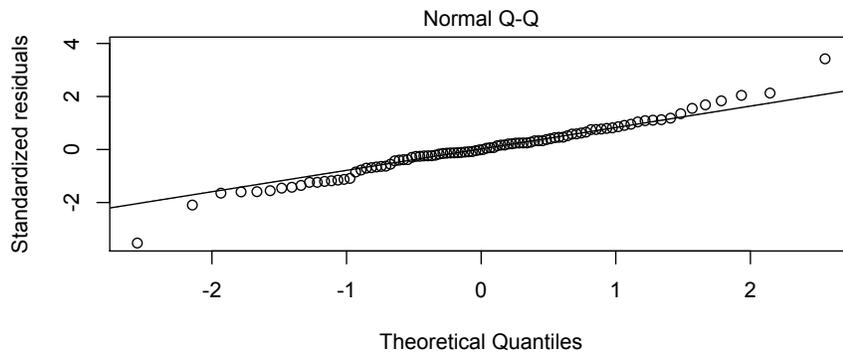
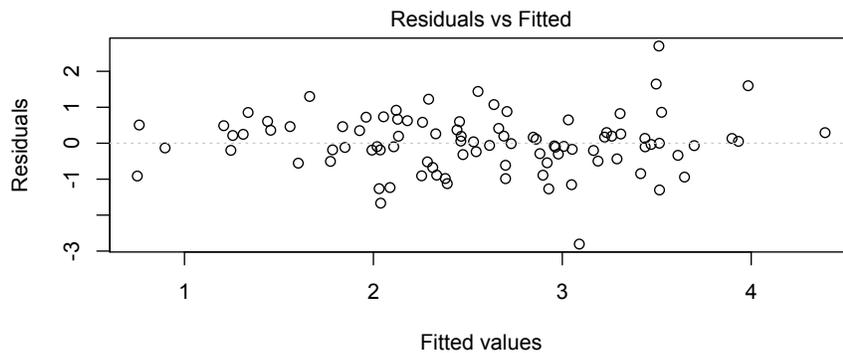
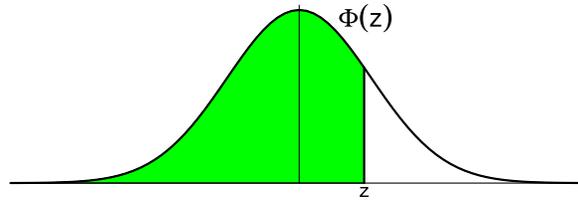


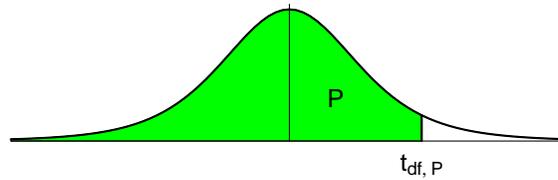
Tabelle der Kumulativen Normalverteilung $\Phi(z) = P[Z \leq z]$, $Z \sim \mathcal{N}(0, 1)$



Bsp.: $P[Z \leq 1.96] = 0.975$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

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
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
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