

2. Vordiplom: Statistik (2 Stunden)

Benötigte Tabellen befinden sich auf den letzten zwei Seiten der Aufgabenstellungen. Wenn nicht anders vermerkt, führen Sie die Tests auf dem 5%-Niveau durch. Geben Sie jeweils Null- und Alternativhypothese, Verwerfungsbereich und Testentscheidung an.

Berechnungen sind jeweils auf 10% genau verlangt.

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 Glück!

1. (7 Punkte)

Ameisen im Lande Utopia sind entweder rot oder schwarz. Der Anteil der roten Ameisen beträgt nur 20%.

- a) Wie gross ist die Wahrscheinlichkeit, dass sich unter drei (zufällig ausgewählten) Ameisen genau eine rote Ameise befindet?

Jeweils 10'000 (zufällig ausgewählte) Ameisen leben in einem Nest.

- b) Wie gross ist der Erwartungswert und die Varianz der Anzahl roter Ameisen in einem Nest?
- c) Berechnen Sie die Wahrscheinlichkeit, in einem Nest mehr als 2000 rote Ameisen anzutreffen. Benutzen Sie eine geeignete Approximation.
- d) Derselbe Forscher behauptet, dass der Anteil der roten Ameisen in Wahrheit höher als 20% ist. Ein Nest wird zufällig gewählt und es befinden sich 2088 rote Ameisen darin. Kann der Forscher seine These damit statistisch beweisen? Führen Sie einen Test auf dem 5% Niveau durch.

2. (6 Punkte)

Die Wartezeit X eines Kunden an einer Kaufhauskasse sei exponentialverteilt mit Parameter λ (Einheit von X sind Sekunden). Die Dichtefunktion f von X lautet folglich:

$$f(x) = \lambda \exp(-\lambda x) \quad \text{für } x \geq 0$$

- a) Geben Sie die kumulative Verteilungsfunktion an.
- b) Wie gross ist die Wahrscheinlichkeit, dass ein Kunde weniger als 10 Sekunden warten muss, wenn $\lambda = 0.1$?
(**Hinweis:** $\exp(-10) = 4.53 \cdot 10^{-5}$, $\exp(-1) = 0.368$, $\exp(-0.1) = 0.905$, $\exp(0.1) = 1.11$, $\exp(1) = 2.72$, $\exp(10) = 22026$)

- c) Wieviel länger (in Sekunden) steht ein Kunde im Durchschnitt bei einer Kasse mit $\lambda = 0.1$ als bei einer Kasse mit $\lambda = 0.2$ an?
- d) Bei 20 zufällig beobachteten Kunden wurde eine gesamte Wartezeit von 6 Minuten und 40 Sekunden festgestellt. Welchen Schätzwert erhält man damit für λ ?

3. (5 Punkte) Die linke Grafik zeigt den Preis von 100 Büchern (y ; in Pence) in Abhängigkeit von ihrer Dicke (x ; in mm). Die Daten wurden in Zusammenhang mit der Schätzung einer potenziellen Schadenssumme einer Hausrats-Versicherung erhoben.

An diese Daten wurde folgendes Regressionsmodell angepasst:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

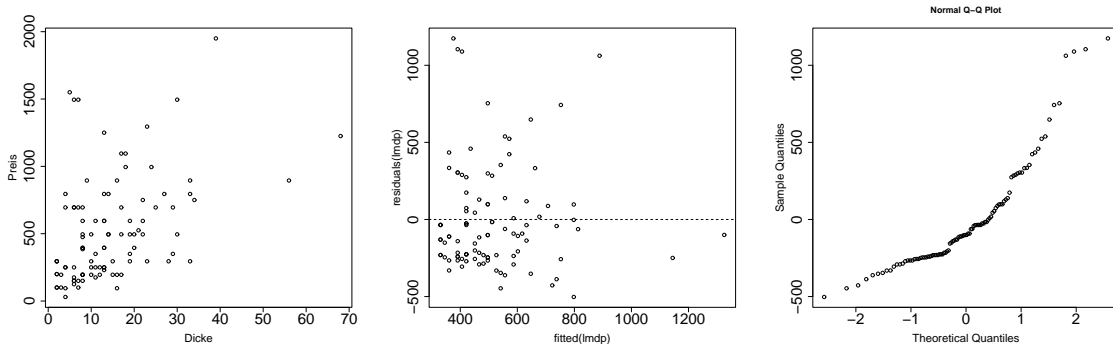
Hier sehen Sie einen Teil des R-Outputs:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	300.485	57.468	5.229	...
Dicke	15.071	3.171	4.752	...

Residual standard error: ? on 98 degrees of freedom

Multiple R-Squared: 0.1873, Adjusted R-squared: 0.179



(a) Scatterplot

(b) Tukey-Anscombe (TA) plot

(c) Normalplot

- Es gibt einen signifikanten Zusammenhang zwischen Buchpreis und Dicke. (β ist signifikant von 0 verschieden.)
 - richtig
 - falsch
- Welches der folgenden Intervalle ist unter Normalverteilungs-Annahme an die Fehler ein exaktes 95%-Vertrauensintervall für β ?
 - $15.071 \pm 1.984 \cdot 3.171$
 - $15.071 \pm 1.984 \cdot 4.752$
 - $15.071 \pm \frac{1.984}{\sqrt{100}} \cdot 3.171$
 - $15.071 \pm \frac{1.984}{\sqrt{100}} \cdot 4.752$
 - Keines der angegebenen Intervalle
- Wie gross ist die Schätzung $\hat{\sigma}$ von σ ("?" im Output) ungefähr?
 - $0 \leq \hat{\sigma} < 10$
 - $10 \leq \hat{\sigma} < 100$
 - $100 \leq \hat{\sigma} < 1000$
 - $1000 \leq \hat{\sigma}$

- 4) Wieviel Pence kostet ein Buch der Dicke 30mm nach Massgabe der Regressionsgerade ungefähr?
- a) 500 b) 750 c) 1000 d) 1250 e) 1500
- 5) Lassen sich Abweichungen von den Modellannahmen feststellen?
- a) Der Zusammenhang zwischen Dicke und Preis ist anscheinend nichtlinear.
- b) Die Normalverteilungs-Annahme ist anscheinend verletzt.
- c) Es lassen sich keine Abweichungen von den Modellannahmen feststellen.

4. (6 Punkte) Die unten angegebenen Daten geben Salzhaltigkeitswerte von zwei getrennten Gewässern der Bimini Lagune, Bahamas, an. In jedem Gewässer wurden mehrere Messungen gemacht. Von Interesse ist die Frage, ob sich die Salzhaltigkeiten der beiden Gewässer unterscheiden, oder ob sie gleich sind.

Daten:

Gewässer 1	39.38	39.70	39.76	39.79	40.17	40.44	40.79	40.80		
Gewässer 2	38.24	38.51	38.53	38.66	38.71	38.89	39.04	39.05	39.21	40.08

Kennzahlen:

Gewässer 1	$\bar{x} = 40.10, s_x^2 = \hat{\sigma}_x^2 = 0.282$
Gewässer 2	$\bar{y} = 38.89, s_y^2 = \hat{\sigma}_y^2 = 0.261$

- 1) Der angemessene t-Test ist...
- a) gepaart b) ungepaart
- 2) Was ist die korrekte Nullhypothese?
- a) Die Salzhaltigkeiten unterscheiden sich nicht.
- b) Die Salzhaltigkeiten unterscheiden sich.
- c) Gewässer 1 hat eine höhere Salzhaltigkeit als Gewässer 2.
- 3) Was ist die korrekte Alternativhypothese?
- a) Die Salzhaltigkeiten unterscheiden sich nicht.
- b) Die Salzhaltigkeiten unterscheiden sich.
- c) Gewässer 1 hat eine höhere Salzhaltigkeit als Gewässer 2.
- 4) Der Wert der geeigneten t-Test-Statistik ist...
- a) 1.64 b) 4.90 c) 9.45 d) 160.26
- 5) Das Ergebnis des geeigneten t-Tests ist...
- a) Der Salzhaltigkeits-Unterschied ist weder signifikant auf dem 5%-, noch auf dem 1%-Niveau.
- b) Die Salzhaltigkeiten unterscheiden sich signifikant auf dem 5%-Niveau, jedoch nicht auf dem 1%-Niveau.
- c) Die Salzhaltigkeiten unterscheiden sich signifikant auf dem 1%-Niveau, jedoch nicht auf dem 5%-Niveau.

- d) Die Salzhaltigkeiten unterscheiden sich sowohl auf dem 1%-Niveau, als auch auf dem 5%-Niveau signifikant.
- 6) Der Wilcoxon-Test führt zu einem p-Wert von 0.0005. Wie ist das zu interpretieren (unabhängig vom Ausgang des t-Tests)?
- Es ist statistisch gesichert, dass sich die Salzhaltigkeiten unterscheiden.
 - Es ist statistisch gesichert, dass sich die Salzhaltigkeiten nicht unterscheiden.
 - Die Salzhaltigkeiten können nicht statistisch sicher unterschieden werden.

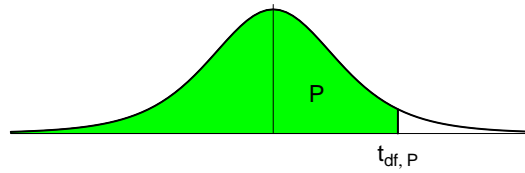
5. (6 Punkte)

Ein Freund will mir einen Brief schicken. Er hat mir versprochen, ihn am Dienstag einzustecken. Die Post stellt Briefe mit einer Wahrscheinlichkeit von 80% am folgenden Tag, mit einer Wahrscheinlichkeit von 15% am übernächsten Tag und mit einer Wahrscheinlichkeit von 5% noch später zu (Zahlen sind erfunden). Mein Freund ist nach meiner Einschätzung nicht ganz zuverlässig und steckt den Brief nur zu 70% tatsächlich am Dienstag ein, zu 15% jedoch erst am Mittwoch und zu 15% noch später.

Man kann die folgenden Ereignisse definieren:

- A_1 : Mein Freund steckt den Brief Dienstag ein.
 - A_2 : Mein Freund steckt den Brief Mittwoch ein.
 - A_3 : Mein Freund steckt den Brief noch später ein.
 - B_1 : Die Post stellt am folgenden Tag zu.
 - B_2 : Die Post stellt am übernächsten Tag zu.
 - B_3 : Die Post stellt noch später zu.
- a) Wie gross ist die Wahrscheinlichkeit, dass der Brief bei mir spätestens am Donnerstag im Briefkasten liegt?
Hinweis: Das ist das Ereignis $[A_1 \cap B_1] \cup [A_1 \cap B_2] \cup [A_2 \cap B_1]$.
- b) Angenommen, der Brief ist auch am Donnerstag noch nicht angekommen. Wie gross ist, gegeben diese Information, die Wahrscheinlichkeit, dass mein Freund sein Versprechen gehalten hat, d.h., dass er den Brief schon Dienstag eingesteckt hat?
- c) Angenommen, am Dienstag abend erfahre ich, dass mein Freund den Brief noch nicht eingesteckt hat. Wie wahrscheinlich ist es dann, dass ich den Brief spätestens Donnerstag bekomme (ohne dass mein Freund noch einmal extra daran erinnert wird)?

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
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