2. Vordiplom: Statistik (2 Stunden)

Bemerkungen:

- Lies zuerst alle Aufgaben durch! Verweile nicht zu lange bei einem Aufgabenteil, der Dir grosse Schwierigkeiten bereitet! Für die Note 6 brauchen nicht alle Aufgaben gelöst zu sein!
- Die Benützung des Taschenrechners ist **nicht** erlaubt. Die Berechnungen sind jeweils auf 10% genau verlangt.
- Die nötigen Tabellen befinden sich auf den hintersten Seiten dieser Prüfung.
- Aufgaben 2, 3 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. Trage die korrekten Antworten der Multiple Choice-Aufgaben mit Kreuzchen in das separate Antwortblatt ein.

Viel Erfolg!

1. (8 Punkte) In einem Wald werden neue Bäume gepflanzt. Es soll zunächst untersucht werden, wieviele der gepflanzten Bäume das erste Jahr überleben. Man nimmt an, dass jeder Baum unabhängig von den anderen eine Überlebenswahrscheinlichkeit von p hat (mit einem Wert von p zwischen 0 und 1).

a) Angenommen, der wahre Wert betrage $p = 0.5$. Wie gross ist dann die Wahrscheinlichkeit, dass von 3 Bäumen genau einer das erste Jahr überlebt? Wie gross ist die Wahrscheinlichkeit, dass alle drei Bäume überleben?

b) Im folgenden sei p unbekannt. In einem Versuch werden 100 Bäume gepflanzt und genau 50 dieser Bäume überleben das erste Jahr. Gib ein zweiseitiges 95%-Vertrauensinterall für die Überlebenswahrscheinlichkeit p an (Hinweis: benutze eine geeignete Approximation).

c) Angenommen, Du kannst bestimmen, wieviele Bäume in einem neuen Versuch gepflanzt werden. Du möchtest sicherstellen, dass das (aus diesem neuen Versuch) resultierende zweiseitige 95%-Vertrauensintervall für p eine gesamte Länge von höchstens 0.02 besitzt. Wie viele Bäume müssten in diesem neuen Versuch mindestens gepflanzt werden, um dies zu garantieren (Hinweis: man darf approximieren $\Phi^{-1}(0.975) \approx 2$)?

2. (9 Punkte) Die Firma SHELL macht zur Zeit eine Werbekampagne für ihr Spar-Benzin. Die Behauptung, dass der Verbrauch mit diesem Treibstoff tatsächlich tiefer ist, wird experimentell überprüft. Auf einer abgesperrten Strecke werden mit $n = 8$ verschiedenen Autos jeweils 1000 Kilometer zurückgelegt. Jedes Fahrzeug wird einmal mit dem Spar-Benzin von Shell und einmal mit gewöhnlichem Benzin betankt. Gemessen wird der Benzinverbrauch in Litern.

Daten:

Shell	80	84	88	96	97	98	103	117
Anderes	83	85	92	99	102	103	105	114
Differenz	-3	-1	-4	-3	-5	-5	-2	+3

Kennzahlen:

Shell	$\bar{x}$	=	95.375	s_x	=	11.69
Anderes	$\bar{y}$	=	97.875	s_y	=	10.53
Differenz	$\bar{x} - \bar{y}$	=	-2.500	s_{x-y}	=	2.62

1) In dieser Situation ist ein gepaarter, nicht ein ungepaarter Test angezeigt. Der Grund ist, weil ...

   a) die Differenz des Benzinverbrauchs interessiert

   b) wir in beiden Gruppen dieselbe Anzahl Beobachtungen haben

   c) der Fahrzeugtyp grösseren Einfluss auf den Verbrauch hat als das Benzin

   d) der gepaarte Test einen grösseren Fehler 2. Art hat

2) Wie lautet die korrekte Nullhypothese?

   a) Mit dem Spar-Benzin ist der Verbrauch geringer

   b) Es gibt einen Unterschied im Benzinverbrauch

   c) Der Benzinverbrauch ist für alle Fahrzeuge identisch

   d) Es gibt keinen Unterschied im Benzinverbrauch

3) Wie lautet die korrekte Alternativhypothese?

   a) Mit Spar-Benzin ist der Verbrauch grösser

   b) Mit Normal-Benzin ist der Verbrauch grösser

   c) Es gibt einen Unterschied im Benzinverbrauch

   d) Es gibt keinen Unterschied im Benzinverbrauch

4) Der Wert der geeigneten t-Test-Statistik beträgt

   a) $\frac{\sqrt{n}(\bar{x}-\bar{y})}{s_x - s_y} = -6.14$ b) $\frac{\sqrt{n}(\bar{x}-\bar{y})}{\frac{1}{2}(s_x+s_y)} = -0.64$

   c) $\frac{(\bar{x}-\bar{y})}{\frac{1}{2}\sqrt{(s_x^2+s_y^2)}} = -0.32$ d) $\frac{\sqrt{n}(\bar{x}-\bar{y})}{s_{x-y}} = -2.70$

5) Der kritische Wert der Teststatistik für das 5%-Signifikanz-Niveau ist

   a) $t_{8,0.95}$ b) $t_{7,0.95}$ c) $t_{7,0.975}$ d) $t_{14,0.95}$ e) $t_{14,0.975}$

6) Nehme an, der p-Wert dieses Test betrage 0.03. Dann wird die Nullhypothese

   a) auf dem 5%-Niveau, aber nicht auf dem 1%-Niveau akzeptiert

   b) auf dem 5%-Niveau, aber nicht auf dem 1%-Niveau verworfen

   c) auf dem 1%-Niveau, aber nicht auf dem 5%-Niveau akzeptiert

   d) auf dem 1%-Niveau, aber nicht auf dem 5%-Niveau verworfen

7) Wie gross ist der p-Wert, wenn wir anstatt dem $t-Test$ einen Vorzeichentest durchführen?
a) $\frac{1}{256}$ b) $\frac{2}{256}$ c) $\frac{8}{256}$ d) $\frac{9}{256}$ e) $\frac{16}{256}$ f) $\frac{18}{256}$

8) Nehme an, beim Eintragen der Messwerte in die Tabelle sei rechts oben ein Übertragungsfehler passiert und der Wert fürs Shell-Benzin betrage 127 statt 117. Der p-Wert für den t-Test wird dann
a) kleiner b) gleich bleiben c) grösser d) keine Aussage möglich

9) Wie verändert sich der p-Wert durch den in 8) geschilderten Übertragungsfehler beim Vorzeichen-Test? Er wird dann
a) kleiner b) gleich bleiben c) grösser d) keine Aussage möglich

3. (7 Punkte) Marco und Isabelle machen folgendes Wettspiel: Marco wirft eine Münze hoch. Die beiden (gleichwahrscheinlichen) Ausgänge sind Kopf und Zahl. Wenn die Münze Kopf anzeigt, gewinnt Isabelle einen Franken; bei Zahl passiert nichts. Es sei X der Gewinn von Isabelle bei einem einzigen Münzwurf.

a) Wie hoch ist der erwartete Gewinn $\mathbf{E}[X]$? Wie hoch ist die Varianz $\mathrm{Var}(X)$?

In einem zweiten Spiel werden *zwei* Münzen gleichzeitig hoch geworfen. Isabelle gewinnt einen Franken, wenn *beide* Münzen Kopf anzeigen; ansonsten passiert wieder nichts. Nun sei Y der Gewinn von Isabelle, wenn dieses Spiel *zweimal* hintereinander gespielt wird.

b) Wie hoch ist nun der erwartete Gewinn $\mathbf{E}[Y]$? Wie hoch ist die Varianz $\mathrm{Var}(Y)$?

In einer allgemeinen Variante des Spiels werden k Münzen gleichzeitig hochgeworfen und Isabelle gewinnt einen Franken, wenn *alle* k Münzen Kopf anzeigen; ansonsten gibt es keinen Gewinn.

c) Wie oft muss dieses Spiel hintereinander gespielt werden, damit Isabelle denselben erwarteten Gewinn hat, wie für das (einmalig durchgeführte) Spiel in Teilaufgabe a)? Wie gross ist die Varianz des Gewinns für diese Anzahl von Spielen? Mit welcher Verteilung lässt sich der Gewinn beschreiben für grosse k?

4. (6 Punkte) In einer Studie wurde die maximale Muskelkraft im Oberschenkel gemessen und durch das Alter in Jahren und das Körpergewicht in kg erklärt. Die untersuchten Personen waren durchwegs Männer, die zwischen 20 und 50 Jahre alt waren.

Das Regressionsmodell lautet:

$$\text{Muskelkraft}_i = \alpha + \beta_1 \cdot \text{Körpergewicht}_i + \beta_2 \cdot \text{Alter}_i + E_i, \quad E_i \sim \mathcal{N}\left(0, \sigma^2\right)$$

```
Coefficients: Estimate Std. Error t value
(Intercept)    115.00       42.78     ...
gewicht          3.00        0.47     ...
alter           -2.65        0.53     ...
Residual standard error: 30.44 on 27 degrees of freedom
F-statistic: 30.02 on 2 and 27 DF,  p-value: 1.793e-08
```

1) Mit wievielen Beobachtungen wurde diese Regression gerechnet?
   a) 30 b) 29 c) 28 d) 27 e) 26 f) 24

2) Ist es statistisch gesichert, dass in diesem Regressionsmodell ein Zusammenhang zwischen dem Körpergewicht und der Muskelkraft besteht?
   a) ja b) nein c) keine Aussage möglich

3) Wie gross ist die prognostizierte Muskelkraft eines 40-jährigen, 80kg schweren Mannes
   a) 134 b) 249 c) 118.35 d) 375 e) 9

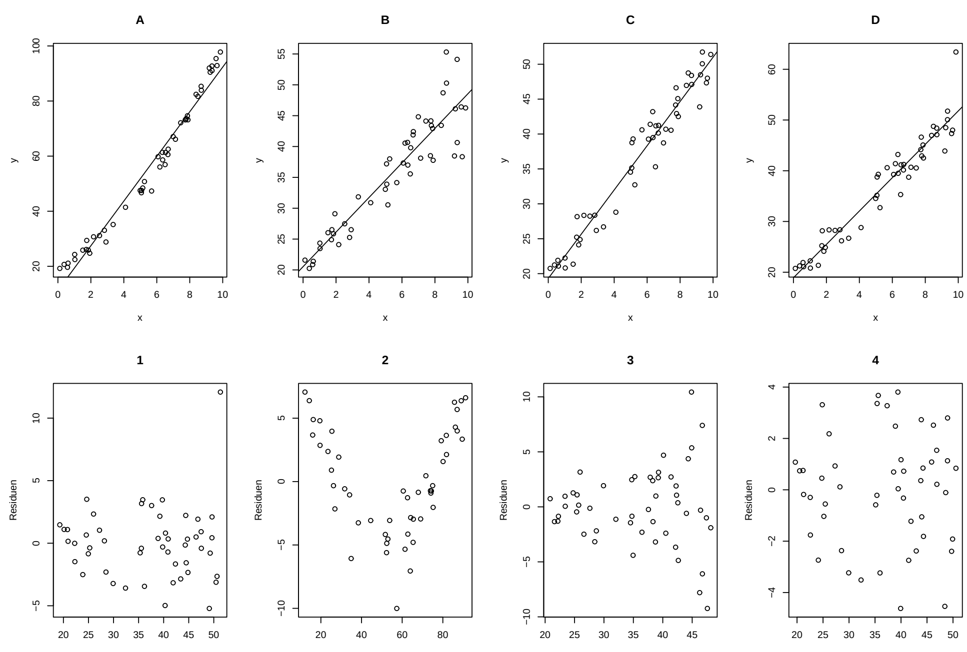
4) Welches Intervall ist ein exaktes 95%-Vertrauensintervalle für β_2?
   a) $-2.65 \pm 2.052 \cdot 0.53$ b) $-2.65 \pm 2.042 \cdot 0.53$
   c) $-2.65 \pm 2.052 \cdot \frac{0.53}{\sqrt{27}}$ d) $-2.65 \pm 2.042 \cdot \frac{0.53}{\sqrt{27}}$

5) Der F-Test in der letzten Zeile des Regressionsoutputs besagt, dass der Achsenabschnitt signifikant von 0 verschieden ist
   a) richtig b) falsch

6) In einer einfachen linearen Regression mit nur noch dem Alter als erklärender Variable, wäre der Einfluss des Alters auf die Muskelkraft signifikant?
   a) auf jeden Fall
   b) auf keinen Fall
   c) kann mit den vorliegenden Angaben nicht entschieden werden

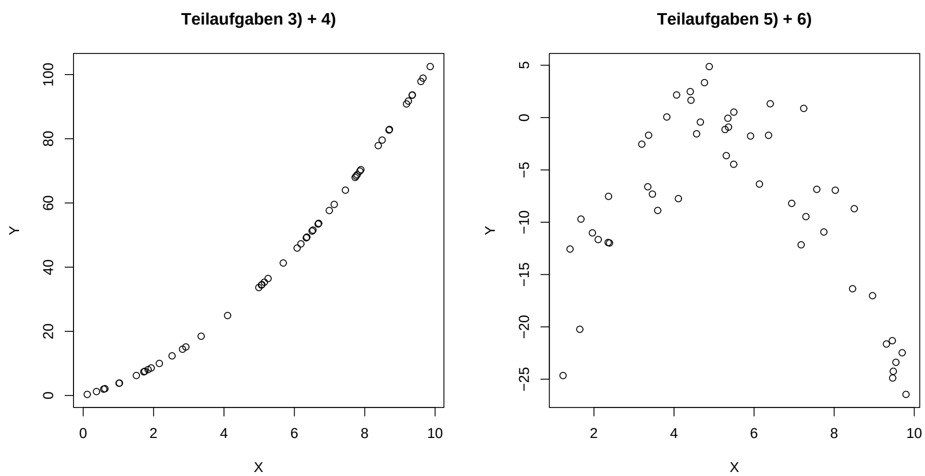
5. (6 Punkte) Verschiedene Aufgaben aus mehreren Gebieten:

1) Die richtige Zuordnung von Streudiagrammen und Tukey-Anscombe-Plots ist
   a) A2 B3 C1 D4 b) A4 B3 C2 D1 c) A2 B3 C4 D1
   d) A3 B2 C4 D1 e) A1 B3 C4 D2 f) A2 B1 D3 C4



2) Wie gross ist in Teilaufgabe 1) die Wahrscheinlichkeit für einen Studenten der keine Ahnung hat, zufällig alle 4 Zuordnungen richtig zu machen?
a) $\frac{1}{4}$ b) $\frac{1}{8}$ c) $\frac{1}{16}$ d) $\frac{1}{24}$ e) $\frac{1}{64}$ f) $\frac{1}{256}$

Betrachte das folgende Streudiagramm



Welche Aussagen sind im linken Plot zutreffend?

3) A: X und Y sind unabhängige Zufallsvariablen
B: X und Y sind unkorrelierte Zufallsvariablen
a) nur A b) nur B c) weder A noch B d) A und B

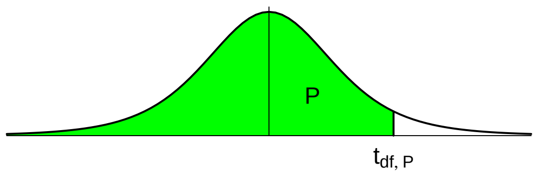
4) C: Die Korrelation von X und Y ist kleiner als 1
D: Es gibt eine exakte Abhängigkeit zwischen X und Y
a) nur C b) nur D c) weder C noch D d) C und D

Welche Aussagen sind im rechten Plot zutreffend?

5) A: Der Wert von Y steigt, falls der Wert von X ansteigt
B: Die Korrelation zwischen X und Y ist kleiner als 0.8
a) nur A b) nur B c) weder A noch B d) A und B

6) C: Es besteht bloss eine lineare Abhängigkeit zwischen X und Y
D: Eine Logarithmus-Transformation für Y macht die Abhängigkeit linear
a) nur C b) nur D c) weder C noch D d) C und D

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