Prof. P. Bühlmann ETH Zürich Herbst 2006

Statistik: Bachelor
D-BIOL, D-CHAB

Name:		Stud. Nr.:	
Vorname:		Studiengang:	

- **Bitte** ...
  - Legen Sie Ihre Legi auf den Tisch!
  - Tragen Sie Ihre Daten in dieses Deckblatt ein und schreiben Sie auf jedes Blatt Ihren Namen! Schreiben Sie auch Ihren Namen auf das Zusatzblatt zu Aufgabe 3, dass Sie abgeben müssen.
  - Beginnen Sie jede Aufgabe auf einer neuen Seite!
- Lesen Sie alle Aufgaben durch, bevor Sie beginnen.
- Beachten Sie die vorhandenen **Tabellen** im Anhang der Prüfung.
- Die Aufgaben sind verschieden gewichtet. Die erreichbaren Punkte stehen an der Aufgabe.
- Es wird nicht erwartet, dass Sie in der gegebenen Zeit alle Aufgaben vollständig lösen.
- **Viel Erfolg!**

Nicht ausfüllen:

Aufgabe	1	2	3	Σ
mögliche Punkte	6	9	12	27
erreichte Punkte				
Kontrolle				

Punktesumme:	
Vollständigkeit:	

Prof. P. Bühlmann ETH Zürich Herbst 2006

Statistik: Bachelor
D-BIOL, D-CHAB

Aufgabe 1) (6 Punkte) Die Schulbehörde der Stadt vermutet, dass eine Stunde von zusätzlichem Deutschunterricht pro Woche sich positiv auf die Rechtschreibung auswirkt. Dazu hat sie folgendes Experiment durchführen lassen: *Von zwei Schulen, die in ganz unterschiedlichen Quartieren der Stadt liegen, mussten die 11-jährigen Volksschülerinnen und -schüler an einem Diktat im Umfang von 1000 Wörtern teilnehmen. Dabei hatten die SchülerInnen der Schule 2 im vergangenen Jahr eine Stunde mehr Deutschunterricht pro Woche als diejenigen der Schule 1.*
In der folgenden Tabelle sind die Anzahl Fehler im Diktat der einzelnen SchülerInnen eingetragen:

SchülerIn	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Schule 1 : $\mathbf{x_i}$	27	14	18	19	25	22	33	25	23	13	32	30
Schule 2 : $\mathbf{y_i}$	12	13	26	31	14	18	21	23	24	14	18	20

Die Kennzahlen sind:
$\bar{x}_{12} = 23.417,\ \bar{y}_{12} = 19.5,\ s_x^2 = 43.174,\ s_y^2 = 33.909,\ s_{\text{pool}} = 6.208.$

a) Handelt es sich um einen gepaarten oder ungepaarten Vergleich? Begründen Sie kurz.

b) Wie lauten Null- und Alternativhypothese?

c) Ist der Test ein- oder zweiseitig?

Man möchte einen geeigneten t-Test auf dem 5%-Niveau durchführen.

d) Bestimmen Sie den Verwerfungsbereich. Wird die Nullhypothese verworfen?

Aufgabe 2) (9 Punkte) Eine Versicherungsgesellschaft möchte Gebäude in der Schweiz gegen das Risiko von Erdbebenschäden versichern. Es wird angenommen, dass die Anzahl der Erdbeben, die jährlich in der Schweiz auftreten, unabhängig und identisch Poisson $\mathcal{P}(\lambda)$-verteilt sind. Die zur Lösung der Aufgabe benötigten Tabellen der verschiedenen Poissonverteilungen finden Sie im Anhang.

a) Die Versicherungsgesellschaft geht von im Mittel weniger als 14 Erdbeben pro Kalenderjahr in der Schweiz aus. Im Jahr 2005 wurden in der Schweiz 6 Erdbeben beobachtet. Kann aufgrund dieser 6 beobachteten Erdbeben die Annahme der Versicherungsgesellschaft statistisch bestätigt werden? Führen Sie hierzu einen geeigneten statistischen Test auf dem 5%-Niveau durch: Geben Sie die Nullhypothese, Alternative, Teststatistik, den P-Wert und die Testentscheidung an.

b) Ein Erdbebenforscher nimmt an, dass im Mittel 12 Erdbeben pro Kalenderjahr in der Schweiz auftreten. Berechnen Sie unter dieser Annahme die Wahrscheinlichkeit eines Fehlers 2. Art für den Test aus Aufgabenteil a).

c) In der folgenden Tabelle sind die Anzahl (rein fiktiv) der Erdbeben pro Kalenderjahr in der Schweiz von 1996 − 2005 dargestellt.

Jahr	96	97	98	99	00	01	02	03	04	05
Anzahl Erdbeben	5	8	7	11	14	10	6	10	13	6

Das arithmetische Mittel ergibt 9. Berechnen Sie mit den Daten in obiger Tabelle ein genähertes zweiseitiges 95%− Vertrauensintervall für den unbekannten Parameter λ. Hierzu können Sie die Normalverteilungsapproximation verwenden. (Zur Erinnerung: Das 97.5%−Quantil der Standardnormalverteilung ist 1.96.)

Aufgabe 3) (12 Punkte) Alle Teilaufgaben ergeben +1 Punkt, falls richtig beantwortet, $-\frac{1}{2}$ Punkt, falls falsch beantwortet und bei keiner Antwort 0 Punkte. Es ist jeweils genau eine Antwort richtig. Falls die Gesamtpunktzahl dieser ersten Aufgabe negativ sein sollte, so werden keine Punkte dafür angerechnet. Bitte tragen Sie Ihre Antworten für Aufgabe 3 in das dafür vorgesehene Blatt ein.

1. Für den p-Wert eines allgemeinen Tests gilt folgende Aussage:
   (a) Um den p-Wert zu berechnen, muss man das Niveau α des Tests **nicht** kennen.
   (b) Der p-Wert ändert sich je nach vorgegebenem Niveau α.
   (c) Unter der Nullhypothese stimmt der p-Wert mit dem Niveau α überein.

2. Seien $X_1, \ldots, X_n$ unabhängige, identisch normalverteilte Zufallsvariablen. Es wird ein Stichprobentest durchgeführt, um die Nullhypothese $\mu_0 = 0$ zu prüfen, und zwar jeweils ein z- und ein t-Test zu gegebenem Niveau α. Betrachten Sie die Alternativhypothese $\mu_A \neq \mu_0$. Welche Aussage ist korrekt?
   (a) Die Wahrscheinlichkeit eines Fehlers erster Art unterscheidet sich zwischen den beiden Tests.
   (b) Die Wahrscheinlichkeit eines Fehlers zweiter Art ist bei den beiden Tests verschieden.
   (c) Die Wahrscheinlichkeiten für Fehler erster und zweiter Art hängen von den konkreten Daten (Realisierungen der Zufallsvariablen) ab.

3. Nehmen Sie an, es wurde in einer Studie eine lineare Regression durchgeführt, um die Signifikanz einzelner erklärender Variablen zu testen, wobei die Zielvariable Y_i durch zwei erklärende Variablen $(x_{i,1}, x_{i,2})$ beschrieben wird, d.h. $Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + E_i$. Es stellt sich nun die Frage, ob das Paar $x_{i,1}, x_{i,2}$ keinen Einfluss auf Y_i hat, d.h. ob die Nullhypothese $H_0 : \beta_1 = \beta_2 = 0$ angenommen werden kann. Die entsprechende Alternativhypothese ist H_A: „Für mindestens ein $j \in \{1, 2\}$ gilt $\beta_j \neq 0$." Wie gehen Sie vor, um dies zu testen?
   (a) Sie führen je einen t-Test für β_1 und β_2 durch. Falls β_1, β_2 je nicht signifikant von Null abweichen, so bestätigt dies die Nullhypothese.
   (b) Sie führen einen globalen F-Test durch. Falls dieser signifikant ist, hat mindestens eine der Variablen $x_{i,1}, x_{i,2}$ einen relevanten Einfluss auf Y_i.
   (c) Ein anderes Vorgehen ist zu wählen.

4. Für n Datenpunkte $X_1, \ldots, X_n$, die unabhängig, identisch normalverteilt sind mit bekannter Varianz σ_X^2, wird basierend auf dem (zweiseitigen) z-Test ein Vertrauensintervall für den unbekannten Parameter μ berechnet. Nehmen Sie an, das berechnete Vertrauensintervall sei $[90, 110]$. Man möchte den Umfang des Datensatzes erweitern, um ein Vertrauensintervall der Länge 10 anstatt 20 zu erhalten, d.h. von der Form $[\hat{\mu} - 5, \hat{\mu} + 5]$. Wie viele Datenpunkte benötigt man dazu?

   (a) Es werden $\sqrt{2}n$ Datenpunkte benötigt.
   (b) Es werden $2n$ Datenpunkte benötigt.
   (c) Es werden $4n$ Datenpunkte benötigt.

5. Bei einem z-Test wie in der vorherigen Aufgabe wird ein Vertrauensintervall für den unbekannten Parameter μ berechnet. Nehmen Sie an, das Vertrauensintervall sei $[95, 105]$ auf dem 95%-Niveau. Welche Aussage kann gemacht werden?

   (a) Die Nullhypothese $\mu_0 = 96$ wird auf dem 95%-Niveau verworfen.
   (b) Die Nullhypothese $\mu_0 = 94$ wird auf dem 95%-Niveau verworfen.
   (c) Die Nullhypothese $\mu_0 = 100$ wird auf dem 95%-Niveau verworfen.

6. Seien X und Y zwei unabhängige, normalverteilte Zufallsvariablen, wobei $\mathbb{E}[X] = 1$ und $\mathbb{E}[Y] = 3$ gelte. Welche der folgenden Zufallsvariablen ist wiederum normalverteilt mit $\mathbb{E}[Z] = 10$?

   (a) $Z := 2(X - Y - 3)$.
   (b) $Z := 4X + Y + 3$.
   (c) Keine der beiden Zufallsvariablen.

7. Seien X und Y zwei unabhängige, Poisson-verteilte Zufallsvariablen, je mit Parameter μ und λ. Welche Aussage ist korrekt?

   (a) $\mathrm{Var}(X + Y) = \mu + \lambda$.
   (b) $\mathrm{Var}(X + Y) = \mu^2 + \lambda^2$.
   (c) $\mathrm{Var}(X + Y) = \mu^2 + 2\mu\lambda + \lambda^2$.

8. Nehmen Sie an, vier Spielern werden je fünf Karten aus einem Deck von 56 Karten verteilt. Sei X die Anzahl Spieler, die mindestens ein Ass besitzen. Dann gilt

   (a) X ist Bernoulli-verteilt.
   (b) X ist Binomial-verteilt.
   (c) X hat keine der beiden genannten Verteilungen.

9. Für welchen der folgenden Tests muss man die theoretische Standardabweichung (Wurzel der Varianz) der untersuchten Daten kennen?

   (a) t-Test
   (b) z-Test
   (c) Wilcoxon-Test

10. Weshalb wird ein Tukey-Anscombe-Plot durchgeführt?

    (a) Um die Signifikanz von einer oder mehreren erklärenden Variablen abzuschätzen.
    (b) Um die Gültigkeit der gewählten Regressionsgleichung zu überprüfen.
    (c) Um die Normalverteilungs-Annahme der Residuen zu überprüfen.

11. Betrachten Sie einen (1- Stichproben) t-Test zum Niveau α. Die Wahrscheinlichkeit eines Fehlers 1. Art ...

    (a) kann durch geeignete Wahl der Alternativhypothese beliebig reduziert werden.
    (b) ist unabhängig davon, ob einseitig oder zweiseitig getestet wird, falls das Niveau α jeweils dasselbe ist.
    (c) hängt von der gewählten Alternativhypothese ab.

12. Seien X, Y zwei unabhängige, log-normal-verteilte Zuvallsvariablen. Welche Aussage ist korrekt?

    (a) $\log(XY)$ ist normalverteilt.
    (b) $\exp(X+Y)$ ist normalverteilt.
    (c) XY ist normalverteilt.

Zu Aufgabe 3) : Kennzeichnen Sie hier bitte klar und deutlich, welche der Antworten richtig ist. Schreiben Sie Ihren **Namen** und ihre **Leginummer** auf dieses Blatt und geben Sie es ab.

Aufgabe	**a**	**b**	**c**
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			
11.			
12.			

Tabelle der Poissonverteilung mit Parameter λ

$\lambda = 6$:

k	0	1	2	3	4	5	6	7	8	9
$P(X \leqslant k)$	0.0025	0.0174	0.0620	0.1512	0.2851	0.4457	0.6063	0.7440	0.8472	0.9161

k	10	11	12	13	14	15	16	17	18
$P(X \leqslant k)$	0.9574	0.9799	0.9912	0.9964	0.9986	0.9995	0.9998	0.9999	1.000

$\lambda = 12$:

k	0	1	2	3	4	5	6	7	8	9
$P(X \leqslant k)$	0.0000	0.0001	0.0005	0.0023	0.0076	0.0203	0.0458	0.0895	0.1550	0.2424

k	10	11	12	13	14	15	16	17	18	19
$P(X \leqslant k)$	0.3472	0.4616	0.5760	0.6815	0.7720	0.8444	0.8987	0.9370	0.9626	0.9787

k	20	21	22	23	24	25	26	27	28
$P(X \leqslant k)$	0.9884	0.9939	0.9970	0.9985	0.9993	0.9997	0.9999	0.9999	1.000

$\lambda = 13$:

k	0	1	2	3	4	5	6	7	8	9
$P(X \leqslant k)$	0.0000	0.0000	0.0002	0.0011	0.0037	0.0107	0.0259	0.0540	0.0998	0.1658

k	10	11	12	13	14	15	16	17	18	19
$P(X \leqslant k)$	0.2517	0.3532	0.4631	0.5730	0.6751	0.7636	0.8355	0.8905	0.9302	0.9573

k	20	21	22	23	24	25	26	27	28	29
$P(X \leqslant k)$	0.9750	0.9859	0.9924	0.9960	0.9980	0.9990	0.9995	0.9998	0.9999	1.0000

$\lambda = 14$:

k	0	1	2	3	4	5	6	7	8	9
$P(X \leqslant k)$	0.0000	0.0000	0.0001	0.0005	0.0018	0.0055	0.0142	0.0316	0.0621	0.1094

k	10	11	12	13	14	15	16	17	18	19
$P(X \leqslant k)$	0.1757	0.2600	0.3585	0.4644	0.5704	0.6694	0.7559	0.8272	0.8826	0.9235

k	20	21	22	23	24	25	26	27	28	29
$P(X \leqslant k)$	0.9521	0.9712	0.9833	0.9907	0.9950	0.9974	0.9987	0.9994	0.9997	0.9999

k	30	31
$P(X \leqslant k)$	0.9999	1.0000

Tabelle der Poissonverteilung mit Parameter λ

$\lambda = 15$:

k	0	1	2	3	4	5	6	7	8	9
$P(X \leqslant k)$	0.0000	0.0000	0.0000	0.0002	0.0009	0.0028	0.0076	0.0180	0.0374	0.0699

k	10	11	12	13	14	15	16	17	18	19
$P(X \leqslant k)$	0.1185	0.1848	0.2676	0.3632	0.4657	0.5681	0.6641	0.7489	0.8195	0.8752

k	20	21	22	23	24	25	26	27	28	29
$P(X \leqslant k)$	0.9170	0.9469	0.9673	0.9805	0.9888	0.9938	0.9967	0.9983	0.9991	0.9996

k	30	31	32
$P(X \leqslant k)$	0.9998	0.9999	1.0000

$\lambda = 18$:

k	0	1	2	3	4	5	6	7	8	9
$P(X \leqslant k)$	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010	0.0029	0.071	0.0154

k	10	11	12	13	14	15	16	17	18	19
$P(X \leqslant k)$	0.0304	0.0549	0.0917	0.1426	0.2081	0.2867	0.3751	0.4686	0.5622	0.6509

k	20	21	22	23	24	25	26	27	28	29
$P(X \leqslant k)$	0.7307	0.7991	0.8551	0.8989	0.9317	0.9554	0.9718	0.9827	0.9897	0.9941

k	30	31	32	33	34	35	36	37
$P(X \leqslant k)$	0.9967	0.9982	0.9990	0.9995	0.9998	0.9999	0.9999	1.0000

Quantile der t-Verteilung

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