Freiwillige Aufgaben zur Repetition

1. Aufgrund langjähriger Untersuchungen ist bekannt, dass der Bleigehalt X von Kopfsalaten annähernd normalverteilt ist. Ausserdem weiss man, dass der Erwartungswert 32 ppb beträgt und dass 50% der Salate mit zwischen 26 und 38 ppb Blei belastet sind.

a) Mache eine Skizze der Dichte von X und zeichne die Wahrscheinlichkeit, dass ein Kopfsalat zwischen 26 und 38 ppb Blei enthält, in die Skizze ein.

b) Bestimme die Standardabweichung σ_X.
Hinweis: Gehe zur standardisierten Zufallsvariablen Z über.

c) Wie gross ist die Wahrscheinlichkeit, dass ein Salat mehr als 45 ppb Schwermetalle enthält?

d) Wieviele (unabhängige) Bestimmungen des Bleigehaltes müssen geplant werden, wenn der Bleigehalt „auf 5 ppb genau“ bestimmt werden soll, bzw. wenn das Mittel der Messungen mit einer Wahrscheinlichkeit von 95 % um nicht mehr als 5 ppb vom wahren Wert abweichen soll?

2. Ein Photoapparat wird bezüglich Präzision der Verschlusszeiten untersucht. Bei einer Versuchseinstellung von 8 Millisekunden ergaben sich folgende Werte (in Millisekunden):

Versuchsnummer	1	2	3	4	5	6	7	8	9
Verschlusszeit X	8.55	8.17	7.91	8.71	10.89	8.38	8.24	7.99	7.82

Das Mittel der Stichprobe $\bar{x}$ beträgt 8.52, die Standardabweichung $s_x = 0.937$.

Kann man anhand dieser Daten sagen, dass die Verschlusszeit sich signifikant von 8 Millisekunden unterscheidet?

a) Führe einen t-Test auf dem Niveau $\alpha = 0.05$ durch. Formuliere explizit:

- Modellannahmen
- Nullhypothese
- Alternative
- Teststatistik
- Verwerfungsbereich.

b) Interpretiere das Testergebnis in Worten.

c) Liegt $\mu = 8$ im 95%-Vertrauensintervall für den Erwartungswert μ der Verschlusszeit?

d) Der t-Test ist für die vorliegenden Daten nicht besonders geeignet. Weshalb?
Schlage einen besser geeigneten Test vor und begründe Deinen Vorschlag!

3. Berechne die Kleinste-Quadrate-Gerade (d.h. berechne Achsenabschnitt $\widehat{\alpha}$ und Steigung $\widehat{\beta}$) für die folgenden 6 Punkte:

$$(x, y) = (1, 2),\ (2, 3),\ (3, 6),\ (5, 7),\ (6, 8),\ (7, 10).$$

Mache ein Streudiagramm der Daten und zeichne die geschätzte Gerade ein.
Könnte die wahre Steigung 2 sein? (Teste die Nullhypothese $H_0 : \beta = 2$ gegen die Alternative $H_A : \beta \neq 2$.)
Könnte die Gerade auch durch den Nullpunkt gehen? Führe einen geeigneten Test durch, um die Frage zu beantworten.
(Rechenhilfe: Verwende $\widehat{\sigma}(\widehat{\beta}) = 0.14, \widehat{\sigma}(\widehat{\alpha}) = 0.64$)

4.

	08:00		10:00		14:00	
1	121	126	97	131	134	128
2	117	145	145	143	89	133
3	145	114	119	107	108	93
4	108	136	139	86	88	118
5	142	151	143	94	146	126
6	154	105	133	164	153	127
7	115	103	149	139	130	150
8	81	108	107	151	144	138
9	122		154	141	125	119
10	127		102	131	111	142
11	122		108	65	87	
12	141		131	141	162	

Diese Tabelle enthält die Punktzahlen von Studenten dreier Klassen bei einer Statistik-Prüfung. Die Klassen schrieben die Prüfung zu verschiedenen Tageszeiten (8, 10 und 14 Uhr).

Untersuche, ob die Tageszeit einen signifikanten Einfluss auf die Leistung der Studenten hat. Kennzahlen:
$SS_G = 30.92,\ SS_E = 30704.84$

(Quelle: Dixon & Massey, *Introduction to statistical analysis*, 1983, p 188)

5. Unterhalb einer Kläranlage entnehmen wir einen Eimer Wasser aus einem Fluss. Wir wollen nun wissen, ob die Ammoniumkonzentration in diesem Eimer im Rahmen der zufälligen Streuungen mit dem Grenzwert von 200 μgNH_4-N/l übereinstimmen kann. Dazu nehmen wir 16 unabhängige Proben aus diesem Eimer (besser wären 16 Eimer; warum?) und bestimmen deren Ammonium-Konzentration (X_i, in μgNH_4-N/l) mit einem Gerät, von dem wir wissen, dass es einen normalverteilten Fehler der Standardabweichung 10 μgNH_4-N/l hat. Der Mittelwert unserer Proben ergibt $\overline{x} = 203.7$.

a) Wenn die Ammoniumkonzentration im Eimer wirklich 200 μgNH_4-N/l ist (Nullhypothese), wie ist dann $\overline{X}$ verteilt?

b) Wir definieren nun $\overline{x}_{\mathbf{unten}}$ als den Wert, der von $\overline{X}$ nur mit 2.5% Wahrscheinlichkeit unterschritten wird, sofern die Nullhypothese gilt. Wie gross ist $\overline{x}_{\mathbf{unten}}$? (Es soll also gelten: $P[\overline{X} \leq \overline{x}_{\mathbf{unten}}] = 2.5\%$)

c) In welchem Bereich sollte $\overline{X}$ (bei einem zweiseitigen Test) unter der Nullhypothese mit 95% Wahrscheinlichkeit liegen?

d) Kann die Konzentration im Eimer mit dem Sollwert übereinstimmen?

e) Wie sieht der Annahmebereich des z-Tests für unser Beispiel aus?

f) Sinnvollerweise interessiert man sich in einem Beispiel wie diesem nur dafür, ob der Grenzwert *überschritten* wird. Ab welchem Wert von $\overline{X}$ sind wir nicht mehr bereit zu glauben, dass der Grenzwert nicht überschritten wurde? Gib die Null- und die Alternativhypothese sowie den Verwerfungsbereich an.

6. An 224 Ehepaaren einer hessischen Landbevölkerung untersuchte W. Trin, ob eine Beziehung zwischen Körperbau und Gattenwahl besteht. Es wurden bei Ehegatten die Körperbautypen *leptosom*, *athletisch* und *pyknisch* unterschieden. Zu prüfen ist, ob die Gattenwahl unabhängig vom Körperbautyp der beiden Ehepartner erfolgt oder ob sich z.B. häufiger zwei Personen gleichen Körperbautyps heiraten, als dies zufallsmässig zu erwarten wäre.

Körperbautyp des Ehemannes	Körperbautyp der Ehefrau leptosom	athletisch	pyknisch	insges.
leptosom	26	...	17	...
athletisch	...	...	...	136
pyknisch	10	...	22	35
insgesamt	52	116	56	...

a) Ergänze die obige Tabelle.

b) Berechne die Tabelle der erwarteten Anzahlen unter der Nullhypothese der Unabhängigkeit.

c) Prüfe die Nullhypothese der Unabhängigkeit mit dem χ^2-Test.

d) Falls der Test in c) signifikant ist:

i) Wo sind die signifikanten Abweichungen? (Überschlagsmässig mit Faustregel)

ii) Was für Strukturen kannst Du in der Tabelle finden?

7. Täglich wird eine gewisse Substanz in einem Fluss gemessen. Nach einer Sanierung zählt man in den folgenden 6 Monaten 2, 3, 1, 5, 6, und 3 Grenzwertüberschreitungen.

a) Gib ein 95%-Vertrauensintervall für die mittlere monatliche Anzahl Grenzwertüberschreitungen an. (Faustregel oder Tabelle 9.1.c im Buch von Stahel.)
Hinweis: Mache die Annahme, dass die Anzahlen der Grenzwertüberschreitungen pro Monat Poisson-verteilt und unabhängig voneinander sind.

b) Vor der Sanierung hatte man über längere Zeit durchschnittlich 6 Grenzwertüberschreitungen monatlich. Hat eine signifikante Abnahme stattgefunden?
Teste einseitig auf dem 2.5% Niveau. (Faustregel oder exakte Verteilung)

8. Zeitungen durchlaufen bei der Herstellung drei Produktionsstufen: Drucken, Falzen, Bündeln. Die Zeiten, während denen diese Produktionsstufen pannenfrei arbeiten, werden mit T_1, T_2 und T_3 bezeichnet. Die T_i seien unabhängig voneinander und gemäss folgenden Dichten verteilt:

$$f_{T_i} = \begin{cases} c_i \; x \; \exp(-\lambda_i x), & x \geq 0, \\ 0, & \text{sonst} \end{cases} \qquad (i = 1, 2, 3).$$

a) Bestimme c_i für $i = 1, 2, 3$. (*Hinweis:* partielle Integration.)

b) Es sei $\lambda_1 = \frac{1}{600}$, $\lambda_2 = \frac{1}{500}$ und $\lambda_3 = \frac{1}{400}$ (Einheit: $[h^{-1}]$). Berechne die Wahrscheinlichkeit, dass während mindestens 900 Stunden ohne Ausfall Zeitungen produziert werden. (*Hinweis:* Multiplikationssatz für unabhängige Ereignisse.)

9. Auf zwei Fabrikaten A und B ist eine Garantiezeit von einem Jahr gewährleistet. Von 112 Stücken der Sorte A haben 32 einen Defekt im ersten Jahr, von 93 Stücken der Sorte B haben 35 einen Defekt im ersten Jahr.

a) Stelle die Daten in der Form einer Kontingenztafel dar.

b) Führe einen geeigneten Test auf dem 5%-Niveau durch, um zu entscheiden, ob ein signifikanter Unterschied in der Häufigkeit von Defekten im 1. Jahr zwischen dem Fabrikat A und dem Fabrikat B besteht.

SCHÖNE FERIEN