Institut für Stochastik
Universität Karlsruhe
Priv.-Doz. Dr. D. Kadelka

WS 2007/2008
18. April 2008

Nachklausur zur Vorlesung

Statistik für Biologen

Musterlösungen

Aufgabe 1 (15 Punkte)

44 junge Mäuse werden in einem Experiment ständig mit Kraftfutter, von dem angenommen wird, dass es zu einer Gewichtszunahme führt, gefüttert. Nachdem sie ausgewachsen sind, werden sie gewogen. Man erhält die folgenden Werte (in g):

19.7	14.5	18.9	15.4	14.4	13.7	19.4	15.6	16.2
15.4	18.3	19.0	21.1	17.2	15.8	22.9	19.0	16.6
20.8	18.6	17.7	16.7	20.0	13.2	19.0	12.7	13.5
14.1	17.9	20.9	13.0	19.7	17.2	22.5	15.3	19.1
21.1	17.6	16.1	17.4	15.9	21.1	13.8	21.1	

a) Fertigen Sie eine Stamm und Blatt-Darstellung an (Einheit = 1 g).

b) Bestimmen Sie den Quartilsabstand der Stichprobe.

c) Bei normaler Fütterung ist der Median der Grundgesamt bekannt und liegt bei 16.0 g. Testen Sie zum Niveau $\alpha = 0.05$ die Hypothese, dass sich der Median des Gewichts von Mäusen durch die Fütterung mit Kraftfutter nicht ändert.

Lösung:

a) Stamm– und Blatt–Darstellung ($n = 20$, Einheit $= 1dm$):

Stamm	Blätter						
12	7						
13	7	2	5	0	8		
14	5	4	1				
15	4	6	4	8	3	9	
16	2	6	7	1			
17	2	7	9	2	6	4	
18	9	3	6				
19	7	4	0	0	0	7	1
20	8	0	9				
21	1	1	1	1			
22	9	5					

Aus der Stamm- und Blatt-Darstellung erhalten wir die geordnete Stichprobe (wir bezeichnen sie mit x)

$$
\begin{array}{rllllllllll}
x_{()} = & (12.7, & 13.0, & 13.2, & 13.5, & 13.7, & 13.8, & 14.1, & 14.4, & 14.5, \\
& 15.3, & 15.4, & 15.4, & 15.6, & 15.8, & 15.9, & 16.1, & 16.2, & 16.6, \\
& 16.7, & 17.2, & 17.2, & 17.4, & 17.6, & 17.7, & 17.9, & 18.3, & 18.6, \\
& 18.9, & 19.0, & 19.0, & 19.0, & 19.1, & 19.4, & 19.7, & 19.7, & 20.0, \\
& 20.8, & 20.9, & 21.1, & 21.1, & 21.1, & 21.1, & 22.5, & 22.9). &
\end{array}
$$

b) Der Quartilsabstand ist $\tilde{x}_{0.75} - \tilde{x}_{0.25}$. Da $n \cdot 0.25 = 44 \cdot 0.25 = 11$ und $n \cdot 0.75 = 44 \cdot 0.75 = 33$ natürliche Zahlen sind, erhält man

$$
\begin{aligned}
\tilde{x}_{0.25} &= \frac{1}{2}\left(x_{(11)} + x_{(12)}\right) = \frac{1}{2}(15.4 + 15.4) = 15.4 \text{ und} \\
\tilde{x}_{0.75} &= \frac{1}{2}\left(x_{(33)} + x_{(34)}\right) = \frac{1}{2}(19.4 + 19.7) = 19.55
\end{aligned}
$$

und damit den Quartilsabstand $\tilde{x}_{0.75} - \tilde{x}_{0.25} = 19.55 - 15.4 = \underline{4.15}$.

c) Hier bietet sich der Vorzeichentest für den Median an (Skript, S. 141f). Sei M der unbekannte Median des Gewichts der mit Kraftfutter gefütterten Mäuse. Geprüft wird die Hypothese $H_0 \colon M = M_0 = 16.0$ gegen die Alternative $H_1 \colon M \neq M_0$. Die Prüfgröße dieses Tests ist die Anzahl N_+ der x_i, die größer als M_0 sind. Der Stamm- und Blatt-Darstellung entnimmt man (alle Stammwerte 16, 17, … zusammenzählen) $N_+ = 29$. H_0 wird verworfen, wenn $N_+ \leq k$ oder $N_+ \geq n - k$, wobei aus Tabelle A.5 für $n = 44$ und $\alpha = 0.05$ der Wert $k = 15$ folgt. Wegen $N_+ = 29 \geq 44 - 15 = n - k$ ist also die Hypothese abzulehnen.

Aufgabe 2 (15 Punkte)

Von einer bestimmten Erbkrankheit sei bekannt, dass sie bei 10% aller männlichen Tiere einer bestimmten Tierpopulation auftritt.

a) Bei $n = 40$ zufällig ausgewählten männlichen Tieren dieser Population wird überprüft, ob sie diese Erbkrankheit haben. Welche Verteilung besitzt die zufällige Anzahl X der erkrankten Tiere unter den 40 untersuchten?

b) Bestimmen Sie unter den Voraussetzungen von a) die Wahrscheinlichkeit dafür, dass mindestens 2 männliche Tiere an dieser Erbkrankheit leiden.

c) Bestimmen Sie mit einer geeigneten Näherungsformel die Wahrscheinlichkeit dafür, dass unter jetzt $n = 900$ zufällig ausgewählten männlichen Tieren der Anteil der erkrankten Tiere zwischen 8% und 12% liegt.

Lösung:

a) Es liegt hier ein typisches Zufallsexperiment mit den zwei Möglichkeiten „krank“ (*Treffer*) bzw. „nicht krank“ (*Niete*) vor, beschrieben im Skriptum in Abschnitt 5.2.1. Da die Trefferwahrscheinlichkeit hier $p = 0.1$ ist, besitzt X die Verteilung $Bin(40, 0.1)$.

b) Zu berechnen ist $P(X \geq 2)$. Um die Berechnung zu vereinfachen, benützen wir $P(X \geq 2) = 1 - P(X \leq 1)$ und nach Abschnitt 5.1 und Formel (5.1) im Skriptum

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \binom{40}{0} \cdot 0.1^0 \cdot 0.9^{40} + \binom{40}{1} \cdot 0.1^1 \cdot 0.9^{39} \\ &= 0.9^{40} + 40 \cdot 0.1 \cdot 0.9^{39} = 0.0148 + 0.0657 = 0.0805. \end{aligned}$$

und damit $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X \leq 1) = 1 - 0.0805 = \underline{0.9195}$.

c) Wir gehen hier wie in Beispiel 9.2 im Skriptum vor und wenden Formel (9.5) an. Definiert man wie dort die Zufallsvariable X_j durch

$$X_j = \begin{cases} 1\ , & \text{falls das } j\text{-te Tier erkrankt ist,} \\ 0\ , & \text{falls das } j\text{-te Tier nicht erkrankt ist} \end{cases}$$

$(j = 1, \ldots, n)$, so ist $\bar{X}_n$ gerade der Anteil der erkrankten Tiere und mit $\mu = E(X_1) = p = 0.1$ und $\sigma^2 = V(X_1) = p(1-p) = 0.09$ folgt mit der Näherungsformel

$$P(c \leq \bar{X}_n \leq d) \approx \Phi\left(\frac{d - \mu}{\sigma} \cdot \sqrt{n}\right) - \Phi\left(\frac{c - \mu}{\sigma} \cdot \sqrt{n}\right)$$

von S. 102

$$\begin{aligned} P(0.08 \leq \bar{X}_n \leq 0.12) &\approx \Phi\left(\frac{0.12 - 0.10}{\sqrt{0.09}} \cdot \sqrt{900}\right) - \Phi\left(\frac{0.08 - 0.10}{\sqrt{0.09}} \cdot \sqrt{900}\right) \\ &= \Phi(2.00) - \Phi(-2.00) = \Phi(2.00) - (1 - \Phi(2.00)) = 2 \cdot \Phi(2.00) - 1 \\ &= 2 \cdot 0.9772 - 1 = \underline{0.9544}. \end{aligned}$$

Aufgabe 3 (15 Punkte)

Die Körpergröße von Daphnien (in *mm*) eines bestimmten Stammes sei normalverteilt mit Erwartungswert 3.8 und Standardabweichung 0.3. Für ein Experiment werden Daphnien benötigt, die mindestens 3.5 *mm* groß sind.

a) Bestimmen Sie die Wahrscheinlichkeit, dass eine zufällig ausgewählte Daphnie für das Experiment geeignet ist.

b) Sie sind nicht sicher, ob die Standardabweichung σ wirklich wie oben angegeben den Wert von 0.3 hat. Bestimmen Sie für beliebiges $\sigma > 0$ die Wahrscheinlichkeit, dass eine zufällig ausgewählte Daphnie für das Experiment geeignet ist.

c) Wie groß darf die Standardabweichung σ höchstens sein, damit mindestens 90% der Daphnien für das Experiment geeignet sind?

Lösung:

a) Sei X die zufällige Körpergröße einer Daphnie. Nach Voraussetzung gilt $X \sim N(3.8, 0.09)$. Gesucht ist die Wahrscheinlichkeit

$$\begin{aligned} P(X \geq 3.5) &= 1 - P(X < 3.5) \overset{(*)}{=} 1 - P(X \leq 3.5) \\ &= 1 - \Phi\left(\frac{3.5 - 3.8}{0.3}\right) = 1 - \Phi(-1) = 1 - (1 - \Phi(1)) = \Phi(1) = \underline{0.8413}. \end{aligned}$$

Hierbei gilt $(*)$, da X eine stetige Zufallsvariable ist (vergl. Skript, S. 63).

b) Bekannt ist jetzt nur, dass $X \sim N(3.8, \sigma^2)$, wobei $\sigma > 0$ unbekannt ist. Wie in a) kann die Wahrscheinlichkeit $P(X \geq 3.5)$ berechnet werden:

$$\begin{aligned} P(X \geq 3.5) &= 1 - P(X < 3.5) \overset{(*)}{=} 1 - p(X \leq 3.5) \\ &= 1 - \Phi\left(\frac{3.5 - 3.8}{\sigma}\right) = 1 - \Phi\left(\frac{-0.3}{\sigma}\right) = 1 - \left(1 - \Phi\left(\frac{0.3}{\sigma}\right)\right) = \Phi\left(\frac{0.3}{\sigma}\right). \end{aligned}$$

c) Wir wählen σ so, dass $P(X \geq 3.5) = \Phi\left(\frac{0.3}{\sigma}\right) = 0.9$ gilt. Nach Tabelle A.1 gilt $\Phi(1.28) = 0.8997 \approx 0.9$, während $\Phi(1.26) = 0.8962$ und $\Phi(1.30) = 0.9032$. Es muss also

$$\frac{0.3}{\sigma} \approx 1.28$$

gelten. Auflösen dieser Gleichung nach σ ergibt $\sigma = \frac{0.3}{1.28} = 0.234$. σ darf also höchstens $\underline{0.234}$ sein.

Aufgabe 4 (15 Punkte)

Die Höhe von ausgewachsenen (ungedüngten) Sojabohnen werde in der Literatur mit 30 bis 35 cm angegeben. Die Höhe der Sojabohnen kann als annähernd normalverteilt mit unbekanntem Erwartungswert μ und unbekannter Standardabweichung σ angenommen werden. Ein Botaniker hat bei 13 (zufällig und unabhängig voneinander ausgewählten) Sojabohnen die folgende Längen (in cm) gemessen:

31.4, 34.3, 39.4, 30.4, 39.3, 31.7, 44.3, 38.7, 39.0, 37.0, 32.2, 37.4, 36.9.

Hinweis: Sie können folgende Werte verwenden:

$$\bar{x}_{13} = 36.31,\ s_x^2 = 16.54,\ s_x = 4.07.$$

a) Geben Sie ein Konfidenzintervall für μ zur Konfidenzwahrscheinlichkeit 0.9 an.

b) Geben Sie ein Konfidenzintervall für σ^2 zur Konfidenzwahrscheinlichkeit 0.9 an.

c) Testen Sie zum Niveau $\alpha = 0.1$ die Hypothese, dass die mittlere Höhe der Sojabohnen 32.5 cm beträgt.

Lösung:

a) Gesucht wird ein Konfidenzintervall für μ. Dabei wird das $(1-\alpha/2)$–Quantil der t_{n-1}–Verteilung gebraucht:

$$t_{n-1,1-\alpha/2} = t_{12,0.95} \approx 1.78 \ .$$

Untere Schranke für μ:

$$\bar{x}_{13} - \frac{s_x}{\sqrt{n}} \cdot t_{n-1,1-\alpha/2} \approx 36.31 - \frac{4.07}{\sqrt{13}} \cdot 1.78 \approx \underline{34.30} \ .$$

Obere Schranke für μ:

$$\bar{x}_{13} + \frac{s_x}{\sqrt{n}} \cdot t_{n-1,1-\alpha/2} \approx 36.31 + \frac{4.07}{\sqrt{13}} \cdot 1.78 \approx \underline{38.32} \ .$$

Das Konfidenzintervall für μ zum Niveau 0.9 ist also $[34.30, 38.32]$.

b) Gesucht ist nun ein Konfidenzintervall für σ^2. Man braucht dazu die $(1-\alpha/2)$– bzw. $\alpha/2$–Quantile der χ^2_{n-1}–Verteilung:

$$\chi^2_{n-1,1-\alpha/2} = \chi^2_{12,0.95} \approx 21.0 \ , \qquad \chi^2_{n-1,\alpha/2} = \chi^2_{12,0.05} \approx 5.23 \ .$$

Untere Schranke für σ^2:

$$\frac{(n-1)\cdot s_x^2}{\chi^2_{n-1,1-\alpha/2}} \approx \frac{12 \cdot 16.54}{21.0} \approx \underline{9.45} \ .$$

Obere Schranke für σ^2:

$$\frac{(n-1)\cdot s_x^2}{\chi^2_{n-1,\alpha/2}} \approx \frac{12 \cdot 16.54}{5.23} \approx \underline{37.95} \ .$$

Das Konfidenzintervall für σ^2 zum Niveau 0.95 ist also $[9.45, 37.95]$.

c) Da die Varianz der Normalverteilung unbekannt ist, verwenden wir den zweiseitigen Ein-Stichproben-t-Test (Skriptum, S. 140). Die Prüfgröße lautet

$$T(x) = \frac{\sqrt{n} \cdot (\bar{x} - \mu_0)}{s_x}$$

mit $\mu_0 = 32.5$. Die Hypothese $H_0\colon \mu = \mu_0 = 32.5$ wird verworfen, falls $|T(x)| \geq t_{n-1,1-\alpha/2}$. Wegen a) gilt hier $t_{n-1,1-\alpha/2} = t_{12,0.95} \approx 1.78$ und damit

$$T(x) = \frac{\sqrt{13} \cdot (36.31 - 32.5)}{4.07} = 3.38 \geq 1.78 = t_{12,0.95}.$$

Die Hypothese $H_0\colon \mu = \mu_0$ wird also verworfen.

Aufgabe 5 (15 Punkte)

Um festzustellen, ob die Behandlung mit Vitamin B_1 das Wachstum einer bestimmten Pilzsorte beeinflusst, wurde das Gewicht von 8 mit Vitamin B_1 behandelten und 10 unbehandelten Pilzen gemessen. Es ergaben sich die folgenden Werte (in g)

behandelt (x_i)	35.5	26.4	21.6	30.4	27.6	25.3	25.4	28.0		
unbehandelt (y_j)	46.4	42.8	38.4	31.2	34.4	30.3	34.2	37.5	41.3	43.4

a) Nehmen Sie an, dass die x_i aus einer $N(\mu, \sigma^2)$-verteilten Grundgesamtheit und die y_j aus einer $N(\nu, \sigma^2)$-verteilten Grundgesamtheit stammen mit unbekannten μ, ν und σ^2. Testen Sie zum Niveau $\alpha = 0.01$ die Hypothese $H_0 : \mu = \nu$.

Hinweis: $\bar{x} = 27.53$, $\bar{y} = 37.99$, $s_x^2 = 16.85$ und $s_y^2 = 29.67$.

b) Verzichten Sie auf die Normalverteilungsannahme und nehmen Sie statt dessen an, dass die Verteilungsfunktion F des Gewichts der behandelten bzw. G der unbehandelten Pilze durch Verschiebung auseinander hervorgehen. Testen Sie jetzt zum Niveau $\alpha = 0.05$ die Hypothese H_0, dass beide Verteilungsfunktionen F und G übereinstimmen.

Lösung: Es handelt sich hier um ein Zwei-Stichproben-Problem. Die erste Stichprobe sind die Gewichte $x_1, \dots, x_8$ (der $n = 8$ behandelten Pilzen), die zweite Stichprobe sind die Gewichte $y_1, \dots, y_{10}$ (der $m = 10$ unbehandelten Pilze).

a) Die x_i stammen aus einer $N(\mu, \sigma^2)$-, die y_i aus einer $N(\nu, \sigma^2)$-Verteilung. Zu testen ist

$$H_0 : \ \mu = \nu \quad \text{gegen} \quad H_1 : \ \mu \neq \nu.$$

Da die Varianzen gleich sind, ist für diese Situation der zweiseitige Zwei-Stichproben-t-Test angemessen. Er hat die Prüfgröße

$$T = \frac{\sqrt{\frac{m \cdot n}{m+n}} \cdot (\bar{x} - \bar{y})}{\sqrt{\frac{1}{m+n-2} \cdot ((n-1) \cdot s_x^2 + (m-1) \cdot s_y^2)}} \approx \frac{2.108 \cdot (27.53 - 37.99)}{\sqrt{\frac{1}{16} \cdot (7 \cdot 16.85 + 9 \cdot 29.67)}} \approx -4.50.$$

Der kritische Wert ist nach Tabelle A.2

$$t_{m+n-2,1-\alpha/2} = t_{16,0.995} \approx 2.92.$$

Wegen $|T| \geq t_{m+n-2,1-\alpha/2}$ wird die **Hypothese abgelehnt**.

b) Wenn keine Normalverteilungsannahme vorliegt, sollte der t-Test nicht verwendet werden. Unter den getroffenen Voraussetzungen ist der Mann-Whitney-U-Test geeignet. Die Hypothese H_0 lautet: Die Verteilungsfunktion F der ersten Stichprobe und die Verteilungsfunktion G der zweiten Stichprobe sind gleich, also

$$H_0 : \ F = G.$$

Zunächst werden zu **allen** gegebenen 18 Werten die Ränge bestimmt. Nachfolgend sind die Ränge in Klammern nach dem jeweiligen Messwert angegeben:

x_i	21.6(1)	25.3(2)	25.4(3)	26.4(4)	27.6(5)	28.0(6)	30.4(8)	35.5(12)		
y_i	30.3(7)	31.2(9)	34.2(10)	34.4(11)	37.5(13)	38.4(14)	41.3(15)	42.8(16)	43.4(17)	46.4(18)

Nun müssen die Ränge der x_i aufaddiert werden:

$$W = 1 + 2 + 3 + 4 + 5 + 6 + 8 + 12 = 41.$$

Die Prüfgröße des Mann-Whitney-U-Tests ist

$$U = W - \frac{n(n+1)}{2} = 41 - \frac{8 \cdot 9}{2} = 5.$$

erfüllt ist. Aus Tabelle A.7 kann man ablesen:

$$U_{m,n,1-\alpha/2} = U_{10,8,0.975} = 63.$$

Wegen $U = 5 \leq m \cdot n - U_{m,n,1-\alpha/2} = 80 - 63 = 17$ **lehnt der Mann-Whitney-U-Test die Hypothese ab**.

Aufgabe 6 (15 Punkte)

Es soll überprüft werden, ob die Wasserart einen Einfluss auf die Keimungswahrscheinlichkeit von Primula sinensis hat. Jeweils 50 Samen wurden mit Lehmwasser, 50 Samen mit Regenwasser bewässert. Mit Lehmwasser keimten 29 der Samen, mit Regenwasser 13 der Samen.

a) Bestimmen Sie ein (approximatives) 0.95-Konfidenzintervall für den Unterschied $p_1 - p_2$ der Keimungswahrscheinlichkeiten.

b) Testen Sie auf dem 1%-Niveau, ob es einen Unterschied zwischen der Keimwahrscheinlichkeit p_1 bei Lehmwasser und p_2 bei Regenwasser gibt.

Lösung:

a) Konfidenzintervalle für den Unterschied zweier Wahrscheinlichkeiten finden sich im Skriptum auf S. 127. Sei p_1 die Keimwahrscheinlichkeit bei Lehmwasser und p_2 die bei Regenwasser. Dann ist $\hat{p}_1 = \frac{29}{50} = 0.58$ der Trefferanteil unter $n_1 = 50$ mit Lehmwasser bewässerten Samen und $\hat{p}_2 = \frac{13}{50} = 0.26$ der Trefferanteil unter $n_1 = 50$ mit Regenwasser bewässerten Samen.

Da hier in beiden Versuchsreihen mindestens 5 Treffer und 5 Nieten vorkommen, kann man die approximativen unteren und oberen Konfidenzgrenzen für $p_1 - p_2$ von S. 127 verwenden. Daher ist mit $\alpha = 0.05$, also $c_{1-\alpha/2} = c_{0.975} = 1.96$ (das 0.975-Quantil der $N(0,1)$-Verteilung

$$\begin{aligned}
&\hat{p}_1 - \hat{p}_2 - \frac{n_1 + n_2}{2 \cdot n_1 \cdot n_2} - c_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\
&\quad = 0.58 - 0.26 - \frac{100}{5000} - 1.96 \cdot \sqrt{\frac{0.58(1-0.58)}{50} + \frac{0.26(1-0.26)}{50}} = \underline{0.1170}
\end{aligned}$$

eine approximative untere und

$$\begin{aligned}
&\hat{p}_1 - \hat{p}_2 + \frac{n_1 + n_2}{2 \cdot n_1 \cdot n_2} + c_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\
&\quad = 0.58 - 0.26 + \frac{100}{5000} + 1.96 \cdot \sqrt{\frac{0.58(1-0.58)}{50} + \frac{0.26(1-0.26)}{50}} = \underline{0.5230}
\end{aligned}$$

eine approximative obere Konfidenzgrenze für $p_1 - p_2$. Das gesuchte Konfidenzintervall ist daher $[0.1170, 0.5230]$.

b) Wir überprüfen mit dem zweiseitigen Test aus Skriptum, 13.6.1 die Hypothese $H_0\colon p_1 = p_2$ gegen $H_1\colon p_1 \neq p_2$. Wegen $n_1 + n_2 = 100 \geq 60$ können wir die Prüfgröße

$$T = \frac{n_1 \cdot n_2}{n_1 + n_2} \cdot \frac{(\hat{p}_1 - \hat{p}_2)^2}{\hat{p} \cdot (1 - \hat{p})}$$

verwenden. Hierbei ist

$$\hat{p} = \frac{29 + 13}{50 + 50} = 0.42$$

die relative Gesamt-Trefferzahl aus beiden Stichproben. H_0 wird abgelehnt, wenn $T \geq \chi^2_{1,1-\alpha}$, wobei hier $\alpha = 0.01$, also nach Tabelle A.3 $\chi^2_{1,0.99} = 6.63$. Mit den angegebenen Daten folgt

$$T = \frac{50 \cdot 50}{100} \cdot \frac{(0.58 - 0.26)^2}{0.42 \cdot (1 - 0.42)} = 10.51 > 6.63.$$

Die Hypothese wird also abgelehnt.