Musterlösung

zur Klausur „Einführung in die statistische Messdatenauswertung für Biotechnologen“

28.02.2014

Lösung Kurzfragen

1. Fehler 2. Art
   Fehler 1. Art: H_0 ist richtig, wird aber abgelehnt. Hat eher Auswirkungen auf die Ausschussrate des Herstellers, für die Kletterer eher unproblematisch, Fehler 2. Art: H_0 ist falsch, wird aber nicht abgelehnt; die Tragkraft der Seile ist geringer, das wird im Test aber nicht erkannt.

2. Binomialverteilung: $\mathrm{Bin}(n,p) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{60}{10} (\frac{1}{6})^{10} (\frac{5}{6})^{60-10} = 0{,}1370 = 13,7\%$

3. Die Lösung ergibt sich aus den Eigenschaften der Standardabweichung des Mittelwertes. Die Breite des Konfidenzintervalls soll im vorliegenden Fall von $33,95 - 26,45 = 7,5$ auf $31,45 - 28,95 = 2,5$ reduziert werden, also um den Faktor 3. Da die Standardabweichung des Mittelwerts mit $1/\sqrt{n}$ skaliert, muss die Zahl der Messungen um den Faktor $3^2 = 9$ erhöht werden. Bei zuvor 8 Messungen sind dies also $8 \cdot 9 =$ 72 Messungen.

4. Nicht zwangsläufig wg. Fehler 1. Art (Ablehnung bei richtiger H_0).

5. a) Dreiecksignale sind nicht bandbegrenzt, das Shannon-Theorem ist damit nicht erfüllbar

   b) 46 kHz

6. Richtig

7. $h = \frac{H}{n \cdot \Delta x}$ oder $\frac{h}{H} = \frac{1}{n \cdot \Delta x}$ oder „Die relative Häufigkeitsdichte ist die absolute Häufigkeit geteilt durch die Gesamtanzahl der Elemente und die Intervallbreite“.

8. Zeit, el. Stromstärke, Masse, Lichtstärke. (0,5P je Größe)

9. Nein, für das arithmetische Mittel müssen der Verteilung zugrundeliegenden Merkmale mind. intervallskaliert sein. Häufigkeitsverteilung geht aber auch bei Nominal- und Ordinalskala.

Lösung Aufgabe 1

Mittelwerte:
$\bar{x}_{\text{vor}} = 78,25\,g$, $\bar{x}_{\text{nach}} = 71,75\,g$

Streuungen:
$S_{\text{vor}} = 3,28416\,g$, $S_{\text{vor}} = 3,49489\,g$

Lösungsansatz: t-Test zweier Erwartungswerte

Hypothesen:
einseitige Hypothese (Brötchen verlieren an Gewicht)
$H_0 : \mu_{\text{vor}} = \mu_{\text{nach}}$, $H_1 : \mu_{\text{vor}} > \mu_{\text{nach}}$

Berechnung der Testgröße:

$$t_0 = \sqrt{n}\frac{\bar{x}_{\text{vor}} - \bar{x}_{\text{nach}}}{\sqrt{S_{\text{vor}}^2 + S_{\text{nach}}^2}} = \sqrt{8}\frac{78,25 - 71,75}{\sqrt{(3,28416)^2 + (3,49489)^2}} = 3,83(3,83349)$$

Anm.: Zum Teil wurde die Differenz andersherum gebildet ($\bar{x}_{\text{nach}} - \bar{x}_{\text{vor}}$), so dass sich dann $t_0 = -3,83$ ergibt. In dem Fall muss dann allerdings auch die Hypothese $H_1 : \mu_{\text{nach}} = \mu_{\text{x}} < \mu_{\text{y}} = \mu_{\text{vor}}$ mit der anderen Ungleichung gewählt werden, so dass sich dann wieder die Ablehnung der Nullhypothese ergibt ($t_0 < -t_{\text{df};1-\alpha}$).

Ablehnung von H_0, wenn $t_0 > t_{\text{df};1-\alpha}$
Freiheitsgrade: $df = n_{\text{vor}} + n_{\text{nach}} - 2 = 14$
Aus der Vorlesung/Übung: Wenn kein Signifikanzniveau angegeben, gilt $\alpha = 0,05$ $\Rightarrow$ aus Tabelle:
$t_{\text{df};1-\alpha} = t_{14;0,95} = 1,76$

Ergebnis: Die Nullhypothese wird abgelehnt.

Antwortsatz: Mit einer Irrtumswahrscheinlichkeit von 5% ist anzunehmen, dass die Brötchen während des Backens an Gewicht verlieren. (oder sinngemäß)

Lösung Aufgabe 2

a) Regressionskoeffizient

$\bar{x} = 179\,\mathrm{cm}$, $\bar{y} = 42,25$
$S_x = 7,3485\ (7,34847)\,\mathrm{cm}$ ($,\ S_y = 2,1899\ (2,18985)$)
Anm.: S_y wird nur benötigt, wenn man die Restvarianz über die Korrelation r_{xy} berechnet. Bei einigen liefert der Taschenrechner r_{xy} und dann wird die entsprechende Gleichung verwendet. Bei der Berechnung der Restvarianz per Hand wird jedoch nur S_x gebraucht. Die Verwendung von S_y ist daher für die Lösung der Aufgabe nicht nötig $\rightarrow$ keine Punkte, dafür Zeitgewinn.

Rechentabelle nach $b = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ (Anhang zur Klausur):

Körpergröße	Schuhgröße	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
188	45	9	2,75	24,75	81
189	45	10	2,75	27,5	100
182	43	3	0,75	2,25	9
166	38,5	-13	-3,75	48,75	169
180	42,5	1	0,25	0,25	1
182	42,5	3	0,25	0,75	9
172	41	-7	-1,25	8,75	49
176	41,5	-3	-0,75	2,25	9
174	40	-5	-2,25	11,25	25
170	39,5	-9	-2,75	24,75	81
184	44,5	5	2,25	11,25	25
185	44	6	1,75	10,5	36
			Summen:	**173**	**594**

Rechentabelle nach $b = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$ (Anhang zur Klausur):

Körpergröße	Schuhgröße	$x_i y_i$	$n\bar{x}\bar{y}$	x_i^2	$n\bar{x}^2$
188	45	8460	90753	35344	384492
189	45	8505		35721	
182	43	7826		33124	
166	38,5	6391		27556	
180	42,5	7650		32400	
182	42,5	7735		33124	
172	41	7052		29584	
176	41,5	7304		30976	
174	40	6960		30276	
170	39,5	6715		28900	
184	44,5	8188		33856	
185	44	8140		34225	
	Summen:	**90926**		**385086**	

In beiden Fällen ergibt sich $b = 0,2912\,\mathrm{cm}^{-1}$.

Restvarianz: $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \bar{y} + b(\bar{x} - x_i))^2 = 0,2364\ (0,23645)$

Daraus folgt: $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = 0,4863\ (0,48626)$

Vertrauensbereich für $P = 98\%$:

$$\left[b - \frac{\hat{\sigma} t_{n-2;1-\alpha/2}}{\sqrt{n} S_x} ; b + \frac{\hat{\sigma} t_{n-2;1-\alpha/2}}{\sqrt{n} S_x}\right]$$

$t_{n-2;1-\alpha/2} = t_{10;0,99}$ aus Tabelle: 2,76

$$= \left[0,2912\,\mathrm{cm}^{-1} - \frac{0,4863 \cdot t_{10;0,99}}{\sqrt{12} \cdot S_x}\ ;\ 0,2912\,\mathrm{cm}^{-1} + \frac{0,4863 \cdot t_{10;0,99}}{\sqrt{12} \cdot S_x}\right]$$

$$= \left[(0,2912 - 0,0527)\,\mathrm{cm}^{-1}\ ;\ (0,2912 + 0,0527)\,\mathrm{cm}^{-1}\right]$$

b) prognostizierte Schuhgröße inklusive des dazugehörigen Vertrauensbereichs für eine statistische Sicherheit von $P = 95\%$ an für einen 173 cm großen Mann

Durch die berechnete Gerade können wir einem beliebigen gewählten x^*-Wert den Wert

$$y^* = \bar{y} + b(x^* - \bar{x})$$

zuordnen.

Hier:
$P\% = 0,95 \rightarrow t_{n-2;1-\alpha/2} = t_{10;0,975} = 2,23$

$x_1^* = 173\,\mathrm{cm} \rightarrow y^* = 40,50\ (40,5028)$

Der Vertrauensbereich für y^* zur statischen Sicherheit $P\% = 1 - \alpha$:

$$\left[y^* - \frac{\hat{\sigma} t_{n-2;1-\alpha/2}}{\sqrt{n}}\sqrt{1 + \frac{(x^* - \bar{x})^2}{S_x^2}}\ ;\ y^* + \frac{\hat{\sigma} t_{n-2;1-\alpha/2}}{\sqrt{n}}\sqrt{1 + \frac{(x^* - \bar{x})^2}{S_x^2}}\right]$$

Hier:

$$\left[40,50 \pm \frac{0,4863 \cdot 2,23}{\sqrt{12}}\sqrt{1 + \frac{(173cm - 179cm)^2}{(7,3485\,cm)^2}}\right]$$

bzw. bei mehr verwendeten Stellen:

$$\left[40,5028 \pm \frac{0,48626 \cdot 2,23}{\sqrt{12}}\sqrt{1 + \frac{(173cm - 179cm)^2}{(7,34847\,cm)^2}}\right]$$

In beiden Fällen ergibt sich das Vertrauensintervall:

$$[40,50 - 0,4041\ ;\ 40,50 + 0,4041]$$

c) Eine Antwort aus

   a) Jugendliche sind nicht Teil der betrachteten Grundgesamtheit.

   b) 135 cm weit vom durch Messwerte abgedeckten Bereich entfernt, keine Aussage über weiteren linearen Verlauf möglich

   c) und sinngemäß...