Aufgabensammlung zur Übung „Statistische Messdatenauswertung für Biotechnologen“

Musterlösungen

Version 2012d

Inhaltsverzeichnis

1 Statistik-Grundlagen

L 1 Lage- und Streuungsparameter

$n = 20$

arithmetisches Mittel: $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1}{20}(4 + 5 + \ldots + 6 + 5) = 5{,}15$

Streuung: $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2} = \sqrt{\frac{1}{19} \cdot 68{,}55} = \sqrt{3{,}61} = 1{,}9$

Varianz: $S^2 = 3{,}61$

x_i	3	3	4	4	4	4	4	5	5	5	5	5	5	5	5	6	6	6	7	12
n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Modalwert: $M = 5$ (8 Elemente)

Median: $\tilde{x} = \frac{x_{10} + x_{11}}{2} = 5$

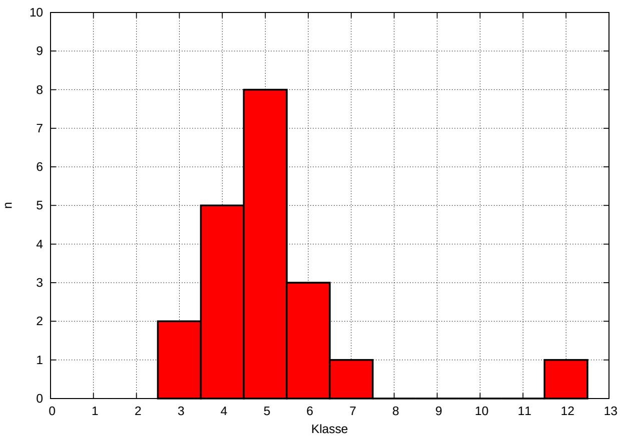
Spannweite: $12 - 3 = 9$ unteres Quartil: 4

oberes Quartil: 5,5

Quartilsabstand: 1,5 (bzw. o. Quartil$= 5$ und Quartilsabstand$= 1$, wenn abgerundet wird (ist so in einigen Literaturstellen zu finden))

Klasse:	3	4	5	6	7	8	9	10	11	12
n:	2	5	8	3	1	0	0	0	0	1

5 Klassen + 1 Ausreißer



L 2 Lage- und Streuungsparameter – Blattlängen

(a) $n = 40$

Mittelwert: $\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n} = 53{,}762\,\mathrm{mm}$

Varianz: $S^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1} = 37{,}13\,\mathrm{mm}^2$

Streuung: $S = 6{,}0932\,\mathrm{mm}$

geordnete Stichprobe:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
39,5	42,5	43,5	45,5	48,5	48,5	49,0	49,0	49,5	49,5	49,5	50,5	50,5	50,5	51,5	51,5	52,5	52,5	53,5	53,5

21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
54,0	54,5	55,0	55,5	56,0	56,0	56,0	56,0	56,5	56,5	57,0	58,5	59,0	59,0	59,0	59,5	62,5	65,0	66,0	68,0

Modalwert: $M = 56{,}0\,\mathrm{mm}$ (4 Elemente)

Medianwert: $\tilde{x} = \frac{x_{20} + x_{21}}{2} = \frac{53{,}5\,\mathrm{mm} + 54{,}0\,\mathrm{mm}}{2} = 53{,}75\,\mathrm{mm}$

Spannweite: $68\,\mathrm{mm} - 39{,}5\,\mathrm{mm} = 28{,}5\,\mathrm{mm}$

unteres Quartil: $x_{10} = 49{,}5\,\mathrm{mm}$ bzw. $\frac{x_{10} + x_{11}}{2} = 49{,}5\,\mathrm{mm}$

oberes Quartil: $x_{30} = 56{,}5\,\mathrm{mm}$ bzw. $\frac{x_{30} + x_{31}}{2} = 56{,}75\,\mathrm{mm}$

Quartilsabstand: $56{,}5\,\mathrm{mm} - 49{,}5\,\mathrm{mm} = 7\,\mathrm{mm}$ (bzw. $7{,}25\,\mathrm{mm}$)

Konfidenzintervall:

Die Messgröße x ist normal verteilt, σ ist unbekannt:

$$\left[\overline{x} - \frac{s}{\sqrt{n}} \cdot t_{n-1\,;\,1-\frac{\alpha}{2}}\,,\, \overline{x} + \frac{s}{\sqrt{n}} \cdot t_{n-1\,,\,1-\frac{\alpha}{2}}\right]$$

$\alpha = 1 - 0{,}95 = 0{,}05$

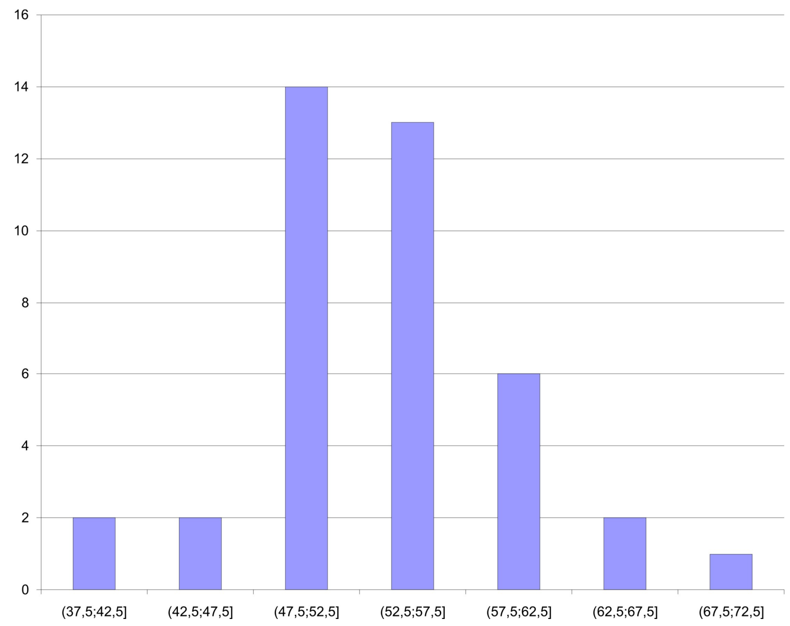
$t_{n-1\,,\,1-\frac{\alpha}{2}} = t_{39\,,\,0{,}975} \approx t_{40\,,\,0{,}975} = 2{,}02$

$$\Rightarrow \left[53{,}762 - \frac{6{,}0932}{\sqrt{40}} \cdot 2{,}02\,;\, 53{,}762 + \frac{6{,}0932}{\sqrt{40}} \cdot 2{,}02\right]$$

$\Rightarrow [51{,}82\,;\,55{,}71]$ Damit können wir 95 % der Messwerte in dem obigen Bereich erwarten und auch der wahre Mittelwert wird hier zu finden sein.

(b) 7 Klassen

Klasse	(37,5;42,5]	(42,5;47,5]	(47,5;52,5]	(52,5;57,5]	(57,5;62,5]	(62,5;67,5]	(67,5;72,5]
n	2	2	14	13	6	2	1



L 3 Gaußsche Normalverteilung – Vertrauensbereich

a) Schätzwerte des Erwartungswertes und Standardabweichung

$\overline{x} = 80{,}019 \qquad S = 0{,}517 \qquad\qquad n = 26 \quad \Rightarrow \quad s = 25$

b) Vertrauensbereich

Parameter $p = 0{,}995 = (1 - \frac{\alpha}{2}) = 1 - \frac{0{,}01}{2}$

$\Rightarrow t = 2{,}79$

$$c = \frac{t}{\sqrt{n}} \cdot S = \frac{2{,}79}{\sqrt{26}} \cdot 0{,}517$$
$$c = 0{,}283$$

c) Wahrscheinlichkeit zwischen 79,8 und 80,2

Wahrscheinlichkeiten aus der Summenfunktion der standardisierten Normalverteilung

$$\Phi(z) = f(z) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} e^{-\frac{t^2}{2}} dt \qquad \text{mit} \quad z = \frac{x-\mu}{\sigma}$$

$\Rightarrow$ Die Werte erhält man z.B. aus der Tabelle im Anhang des Scriptes.

hier: $z = \dfrac{x_i - \overline{x}}{S}$ mit $x_{min} = 79{,}8$ und $x_{max} = 80{,}2$

$$\begin{array}{llll}
z_{min} = \dfrac{79{,}8 - 80{,}019}{0{,}517} = & -0{,}424 \rightarrow Tabelle \rightarrow & \Phi(z_{min}) & = 1 - \Phi(-z_{min}) = \\
 & & & = 1 - 0{,}662757 \\
 & & & = 0{,}3372 \\
z_{max} = \dfrac{80{,}2 - 80{,}019}{0{,}517} = & 0{,}350 & \Phi(z_{max}) & = 0{,}636831
\end{array}$$

$P = \Phi_{max} - \Phi_{min} = 0{,}636831 - 0{,}3372 = 0{,}2996 = 29{,}96\,\% \approx \underline{30\,\%}$

L 4 Fadenlänge von Seidenkokons

$X_1, X_2, \ldots$ sei eine unabhängige Folge von

$N(\mu = 800\,\mathrm{m}, \sigma^2 = 6400\,\mathrm{m}^2)$-verteilten Zufallsvariablen.

$\Rightarrow \sigma = \sqrt{6400\,\mathrm{m}^2} = 80\,\mathrm{m}$

a)

$P(x) = \Phi(\dfrac{x-\mu}{\sigma}), \qquad x \in \mathbb{R}$

$P(X_i \geq 700\,\mathrm{m}) = 1 - \Phi(\dfrac{700\,\mathrm{m} - 800\,\mathrm{m}}{80\,\mathrm{m}}) = 1 - \Phi(-1{,}25)$

$= 1 - (1 - \Phi(1{,}25)) = \Phi(1{,}25) = 0{,}89435 \approx 89{,}4\,\%$

$P(X_i > 1000\,\mathrm{m}) = 1 - \Phi(\dfrac{1000\,\mathrm{m} - 800\,\mathrm{m}}{80\,\mathrm{m}}) = 1 - \Phi(2{,}5) = 0{,}00621 \approx 0{,}6\,\%$

b)

Die Gesamtlänge der von 100 000 Kokons abgewickelten verwertbaren Seidenfäden wird durch die Zufallsvariable $X_{\text{ges}} = X_1 + \ldots + X_{100000}$ beschrieben. X_{ges} ist ebenfalls normal verteilt, und zwar mit Erwartungswert $\mu = 8 \cdot 10^7\,\mathrm{m} \quad (800\,\mathrm{m} \cdot 100000)$ und Varianz $\sigma^2 = 640 \cdot 10^6\,\mathrm{m}^2$ $(6400\,\mathrm{m}^2 \cdot 100000)$.

Die gesuchten Grenzen x_u und x_o sollen die folgenden Bedingungen erfüllen:

$P(x_u \leq S \leq x_o) = P(S \leq x_o) - P(S < x_u) = 0{,}95$

$P(S < x_u) = P(S > x_o) = 1 - P(S \leq x_o)$

$2 \cdot P(S \leq x_o) - 1 = 0{,}95$

Mit der Formel aus (a) erhält man:

$P(S \leq x_o) = \Phi(\frac{x_o - \mu}{\sqrt{\sigma^2}}) = 0{,}975$

$\Phi(z) = 97{,}5\,\%$

$z = \dfrac{x_o - \mu}{\sqrt{\sigma^2}} = 1{,}96$

$x_o = \mu + 1{,}96 \cdot \sqrt{\sigma^2}$

$y = 80049584{,}51\,\mathrm{m}$

Eine analoge Rechnung für x_u liefert:

$x_u = \mu - 1{,}96 \cdot \sqrt{\sigma^2} = 79950415{,}49\,\mathrm{m}$

(c)

Die Zufallsvariable $X_{\text{ges,n}} = X_1 + ... + X_n, n \in \mathbb{N}$, ist $N(n \cdot 800\,\mathrm{m}, n \cdot 6400\,\mathrm{m}^2)$-verteilt. Gesucht sind alle $n \in \mathbb{N}$, für die gilt:

$P(X_{\text{ges,n}} \geq 10^8\,\mathrm{m}) \geq 0{,}99 \qquad (10^8\,\mathrm{m} = 100000\,\mathrm{km} = 100000\,\mathrm{m} \cdot 1000)$

Man erhält:

$P(X_{\text{ges,n}} \geq 10^8\,\mathrm{m}) \geq 0{,}99$

$\Phi(\frac{10^8\,\mathrm{m} - 800\,\mathrm{m} \cdot n}{80\,\mathrm{m} \cdot \sqrt{n}}) \leq 0{,}01 \qquad \Phi(z) = 1\,\%$

$z = -2{,}326$ $\quad\rightarrow$ muss gleich sein: $\dfrac{10^8\,\mathrm{m} - 800\,\mathrm{m} \cdot n}{80\,\mathrm{m} \cdot \sqrt{n}}$

$800\,\mathrm{m} \cdot (\sqrt{n})^2 - 2{,}326 \cdot 80\,\mathrm{m} \cdot \sqrt{n} - 10^8\,\mathrm{m} \geq 0$

(quadratische Gleichung in ξ für $\xi = \sqrt{n}$)

$\underline{n \geq 125083}$

2 Abweichungsfortpflanzung

L 5 Zylindervolumen

Mittelwerte:	Streuungen:	$s = n-1$	P/α	p	t
$\overline{d} = \frac{\sum_{i=1}^{n_d} d_i}{n_d} = 30\,\mathrm{mm}$	$S_d = \sqrt{\frac{\sum_{i=1}^{n_d}(d_i - \overline{d})^2}{n_d - 1}} = 0{,}87\,\mathrm{mm}$	9	95 %/0,05	0,975	2,26
$\overline{l} = \frac{\sum_{i=1}^{n_i} l_i}{n_l} = 400\,\mathrm{mm}$	$S_l = \sqrt{\frac{\sum_{i=1}^{n_l}(l_i - \overline{l})^2}{n_l - 1}} = 7{,}65\,\mathrm{mm}$	4	95 %/0,05	0,975	2,78

Der Erwartungswert des Volumens μ_V liegt im Intervall:

$$\left[V(\overline{d},\overline{l}) - c_V;\ V(\overline{d},\overline{l}) + c_V\right] \qquad \text{mit} \qquad c_V = \sqrt{\left(\frac{\partial V}{\partial d}\bigg|_{\overline{d},\overline{l}} c_d\right)^2 + \left(\frac{\partial V}{\partial l}\bigg|_{\overline{d},\overline{l}} c_l\right)^2}$$

$$\begin{aligned}
V(\overline{d},\overline{l}) &= \frac{1}{4}\overline{d}^2\pi\overline{l} = 282743{,}34\,\mathrm{mm}^3 \\
\frac{\partial V}{\partial d}\bigg|_{\overline{d},\overline{l}} &= \frac{1}{2}\overline{d}\pi\overline{l} \ \ = 18849{,}56\,\mathrm{mm}^2 \\
\frac{\partial V}{\partial l}\bigg|_{\overline{d},\overline{l}} &= \frac{1}{4}\overline{d}^2\pi \ \ = 706{,}86\,\mathrm{mm}^2
\end{aligned}$$

c_d und c_l :

$$\begin{aligned}
c_d &= \frac{t_{n_d-1;1-\alpha/2}}{\sqrt{n_d}} \cdot S_d = \frac{2{,}26}{\sqrt{10}} \cdot 0{,}87 = 0{,}62\,\mathrm{mm} \\
c_l &= \frac{t_{n_l-1;1-\alpha/2}}{\sqrt{n_l}} \cdot S_l = \frac{2{,}78}{\sqrt{5}} \cdot 7{,}65 = 9{,}51\,\mathrm{mm}
\end{aligned}$$

$$\begin{aligned}
c_V &= \sqrt{\left(\frac{\partial V}{\partial d}\bigg|_{\overline{d},\overline{l}} c_d\right)^2 + \left(\frac{\partial V}{\partial l}\bigg|_{\overline{d},\overline{l}} c_l\right)^2} \\
&= \sqrt{(18849{,}56 \cdot 0{,}62)^2 + (706{,}86 \cdot 9{,}51)^2}\,\mathrm{mm}^3 \\
&\underline{= 13530{,}8\,\mathrm{mm}^3}
\end{aligned}$$

Das Vertrauensintervall ist:

$$\underline{\underline{\left[282743{,}34\,\mathrm{mm}^3 - 13530{,}8\,\mathrm{mm}^3\ ;\ 282743{,}34\,\mathrm{mm}^3 + 13530{,}8\,\mathrm{mm}^3\right] \quad P\% = 95\,\%}}$$

L 6 Unbekanntes Prüfteil

$\bar{l} = 1\,\mathrm{dm}$

Streuung: $S_l = 0{,}07071\,\mathrm{dm}$

mit $s = 4$ und $p = 0{,}975 \Longrightarrow t = 2{,}78$ $\qquad (1 - \frac{0{,}05}{2} = 1 - 0{,}025)$

$c_l = 0{,}08791\,\mathrm{dm}$

$\underline{l = (1 \pm 0{,}08791)\,\mathrm{dm} \qquad P = 95\%}$

$\bar{b} = 0{,}28\,\mathrm{dm} \qquad S_B = 0{,}04472\,\mathrm{dm}$

$c_b = 0{,}0556\,\mathrm{dm}$

$\underline{b = (0{,}28 \pm 0{,}0556)\,\mathrm{dm} \qquad P = 95\%}$

$\bar{h} = 0{,}64\,\mathrm{dm} \qquad S_h = 0{,}05477\,\mathrm{dm}$

$c_h = 0{,}06809\,\mathrm{dm}$

$\underline{h = (0{,}64 \pm 0{,}06809)\,\mathrm{dm} \qquad P = 95\%}$

$\rho = f(l,b,h,m) = \frac{m}{lbh}$

$\bar{\rho} = \frac{1{,}402kg}{1 \cdot 0{,}28 \cdot 0{,}64\,\mathrm{dm}^3} = 7{,}824\frac{\mathrm{kg}}{\mathrm{dm}^3}$

$\frac{\partial \rho}{\partial m}\Big|_{\bar{l},\bar{b},\bar{h}} = \frac{1}{lbh} \qquad \frac{\partial \rho}{\partial l}\Big|_{\bar{m},\bar{b},\bar{h}} = -\frac{m}{l^2bh} \qquad \frac{\partial \rho}{\partial b}\Big|_{\bar{m},\bar{l},\bar{h}} = -\frac{m}{lb^2h} \qquad \frac{\partial \rho}{\partial h}\Big|_{\bar{m},\bar{l},\bar{b}} = -\frac{m}{lbh^2}$

$$c_S = \sqrt{\left(\frac{0{,}041}{1 \cdot 0{,}28 \cdot 0{,}64}\right)^2 + \left(\frac{-1{,}402}{1 \cdot 0{,}28 \cdot 0{,}64} \cdot 0{,}08791\right)^2 + \left(\frac{-1{,}402}{1 \cdot 0{,}28^2 \cdot 0{,}64} \cdot 0{,}0556\right)^2 + \left(\frac{-1{,}402}{1 \cdot 0{,}28 \cdot 0{,}64^2} \cdot 0{,}006809\right)^2}\;\frac{\mathrm{kg}}{\mathrm{dm}^3}$$

$$c_S = \sqrt{3{,}632}\frac{\mathrm{kg}}{\mathrm{dm}^3} = 1{,}906\frac{\mathrm{kg}}{\mathrm{dm}^3}$$

Übersicht:

i	$x_1(l)/\,\mathrm{dm}$	$x_2(b)/\,\mathrm{dm}$	$x_3(h)/\,\mathrm{dm}$	$x_4(m)/\,\mathrm{dm}$
1	1,0	0,3	0,7	-
2	1,0	0,3	0,6	-
3	0,9	0,3	0,6	-
4	1,0	0,2	0,7	-
5	1,0	0,3	0,6	-
Mittelwerte $\overline{x}_m$	1,0	0,28	0,64	1,402
Streuung S_{x_m}	0,0707	0,0447	0,0548	
$t_{4\,;\,0{,}975}$	2,78	2,78	2,78	
c_{x_m}	0,0879	0,0556	0,0681	0,041
Abl. ρ bei Mittelwert	-7,824	-27,942	-12,224	5,580
$(c_{x_m} \cdot Ableitung)^2$	0,4731	2,4137	0,6927	0,0523

Erwartungswert und Konfidenzintervall der Dichte für statistische Sicherheit $P\% = 95\,\%$:

$\underline{\rho = 7{,}824\,\mathrm{kg/dm^3} \pm 1{,}906\,\mathrm{kg/\,dm^3}}$

L 7 Torsionsmodul Stahldraht

Vorgehen:

1. Formel für Torsionsmodul G aufstellen
2. sämtliche Vertrauensbereiche für *gleiche Aussagesicherheiten* berechnen

   (evtl. umrechnen)
3. Vertrauensintervall ermitteln

Hier: $G = \dfrac{8\pi L\Theta}{R^4T^2}$

Periodendauer

Mittelwert : $\overline{T} = 11{,}2\,\mathrm{s} \qquad$ für $n = 7$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(T_i - \overline{T})^2} = \sqrt{\frac{1}{6}(0{,}04 + 0 + 0{,}04 + 0{,}04 + 0{,}09 + 0 + 0{,}01)}\,\mathrm{s}$$

$$S = \sqrt{\frac{1}{6}\cdot 0{,}22\,\mathrm{s}^2} = 0{,}191485\,\mathrm{s}$$

$c = \dfrac{t}{\sqrt{n}}\cdot S \qquad$ mit $t = t_{s\,;p}$ $\quad$ ($s =$ Freiheitsgrad und $p =$ Wahrscheinlichkeit)

hier: $s = n - 1 = 6$

$p = 1 - \alpha/2 \qquad P = 95\,\% \Rightarrow \alpha = 0{,}05$

$\Rightarrow p = 0{,}975$

$\Rightarrow$ aus Tabelle: $t_{6\,;0{,}975} = 2{,}45$

$$c_T = \frac{2{,}45}{\sqrt{7}}\cdot S = 0{,}926\cdot 0{,}191485\,\mathrm{s} = 0{,}1773\,\mathrm{s} \qquad (\approx 0{,}18\,\mathrm{s})$$

$\underline{\Rightarrow T = 11{,}2\,\mathrm{s} \pm 0{,}1773\,\mathrm{s} \qquad P = 95\,\%}$

Radius

$R = d/2, \quad \overline{d} = 0{,}4924\,\mathrm{mm} \Rightarrow \overline{R} = 0{,}2462\,\mathrm{mm}$

mit: $n = 5$ $\quad$ Radien: $\quad 0{,}245 \quad 0{,}2465 \quad 0{,}2455 \quad 0{,}2465 \quad 0{,}2475$

$S = \sqrt{\frac{1}{4} \cdot (1{,}14 \cdot 10^{-6} + 9 \cdot 10^{-8} + 4{,}9 \cdot 10^{-7} + 9 \cdot 10^{-8}}\,\mathrm{mm}$

$S = 9{,}75 \cdot 10^{-4}\,\mathrm{mm}$

$s = 4 \qquad p = 0{,}975$

$\Rightarrow t = 2{,}78$

$c_R = \frac{t}{\sqrt{n}} \cdot S = \frac{2{,}78}{\sqrt{5}} \cdot 9{,}75 \cdot 10^{-4}\,\mathrm{mm}$

$\underline{c_R = 1{,}212 \cdot 10^{-3}\,\mathrm{mm}}$

$\underline{\Rightarrow R = 0{,}2462\,\mathrm{mm} \pm 1{,}2 \cdot 10^{-3}\,\mathrm{mm} \qquad P = 95\,\%}$

Länge

$\overline{L} = 121{,}02\,\mathrm{cm} \qquad n = 5$

$S = \sqrt{\frac{1}{4} \cdot (0{,}0484 + 0{,}0004 + 0{,}0324 + 0{,}0144 + 0{,}0324}$

$\underline{S = 0{,}18\,\mathrm{cm}}$

$\Rightarrow t = 2{,}78$

$c_L = \frac{2{,}78}{\sqrt{5}} \cdot 0{,}18\,\mathrm{cm}$

$\underline{c_L = 0{,}223\,\mathrm{cm}}$

$\underline{L = 121{,}02\,\mathrm{cm} \pm 0{,}223\,\mathrm{cm} \qquad P = 95\,\%}$

Wichtig: Einheit ist cm $\Rightarrow$ beim Einsetzen berücksichtigen!

Trägheitsmoment

$\Theta = 1{,}3 \cdot 10^{-3}\,\mathrm{kg} \cdot \mathrm{m}^2$

$n = 8 \qquad c_\Theta(99\,\%) = 1{,}5\,\%$

Problem: Vertrauensbereich bezieht sich auf 99 % Aussagesicherheit $\Rightarrow$ *muss umgerechnet werden!*

$c_\Theta(99\,\%) = 1{,}5\,\% \equiv 1{,}3 \cdot 10^{-3}\,\mathrm{kg\,m^2} \cdot 0{,}015 = 1{,}95 \cdot 10^{-5}\,\mathrm{kg\,m^2}$

$c_\Theta(99\,\%) = \frac{t_{n-1\,;\,1-\alpha_1/2}}{\sqrt{n}} \cdot S$

$$c_\Theta(95\,\%) = \frac{t_{n-1\,;\,1-\alpha_2/2}}{\sqrt{n}} \cdot S$$

S ist bei beiden gleich, da es der selbe Prozess ist.

$\alpha_1 \neq \alpha_2$ somit auch die t-Werte unterschiedlich

$$S = c_\Theta(95\,\%) \cdot \frac{\sqrt{n}}{t_{n-1\,;\,0{,}975}} = c_\Theta(99\,\%)\frac{\sqrt{n}}{t_{n-1\,;\,0{,}995}}$$

$$c_\Theta(95\,\%) = \frac{t_{7\,;\,0{,}975}}{t_{7\,;\,0{,}995}} \cdot c_\Theta(99\,\%)$$

$$c_\Theta = \frac{2{,}36}{3{,}50} \cdot 1{,}95 \cdot 10^{-5}\,\mathrm{kg\,m^2}$$

$\underline{c_\Theta = 1{,}315 \cdot 10^{-5}\,\mathrm{kg\,m^2}}$ (P=95 %)

Mit diesen Einzelmomenten ergibt sich das Torsionsmodul zu:

$$\overline{G} = \frac{8\pi \cdot \overline{L} \cdot \overline{\Theta}}{\overline{R^4} \cdot \overline{T^2}} = \frac{8\pi \cdot 121{,}02 \cdot 10^{-2}\,\mathrm{m} \cdot 1{,}3 \cdot 10^{-3}\,\mathrm{kg\,m^2}}{(0{,}2462 \cdot 10^{-3}\,\mathrm{m})^4 \cdot (11{,}2\,\mathrm{s})^2}$$

$$\overline{G} = 8{,}579 \cdot 10^{10}\frac{\mathrm{kg}}{\mathrm{m \cdot s^2}} \qquad [G] = \frac{\mathrm{N}}{\mathrm{m^2}} = \mathrm{Pa}$$

Dimensionskontrolle: $\dfrac{\mathrm{kg\,m^2 \cdot m}}{\mathrm{m^4 \cdot s^2}} = \dfrac{\mathrm{kg}}{\mathrm{m \cdot s^2}}$

Abweichungsfortpflanzung / Vertrauensbereich

$$\sigma = \sqrt{\sum_{i=1}^{m}\left(\frac{\partial f}{\partial x_i}\sigma_i\right)^2} \qquad \text{bzw.} \qquad c = \sqrt{\sum_{i=1}^{m}\left(\frac{\partial f}{\partial x_i}c_i\right)^2}$$

$$\begin{aligned}
c_G &= \sqrt{\left(\frac{\partial G}{\partial L} \cdot c_L\right)^2 + \left(\frac{\partial G}{\partial R} \cdot c_R\right)^2 + \left(\frac{\partial G}{\partial T} \cdot c_T\right)^2 + \left(\frac{\partial G}{\partial \Theta} \cdot c_\Theta\right)^2} \\
&= \sqrt{\left(\frac{8\pi \cdot \overline{\Theta}}{\overline{R}^4 \cdot \overline{T}^2} \cdot c_L\right)^2 + \left(-4 \cdot \frac{8\pi \cdot \overline{L} \cdot \overline{\Theta}}{\overline{R}^5 \cdot \overline{T}^2} \cdot c_R\right)^2 + \left(-2 \cdot \frac{8\pi \cdot \overline{L} \cdot \overline{\Theta}}{\overline{R}^4 \cdot \overline{T}^3} \cdot c_T\right)^2 + \left(\frac{8\pi \cdot \overline{L}}{\overline{R}^4 \cdot \overline{T}^2} \cdot c_\Theta\right)^2} \\
&= \sqrt{(1{,}5723 \cdot 10^8)^2 + (1{,}6852 \cdot 10^9)^2 + (2{,}7162 \cdot 10^9)^2 + (8{,}698 \cdot 10^8)^2}\,\mathrm{Pa} \\
&= \sqrt{(2{,}472 \cdot 10^{16} + 2{,}8399 \cdot 10^{18} + 7{,}378 \cdot 10^{18} + 7{,}5657 \cdot 10^{17}}\,\mathrm{Pa} \\
&= 3{,}32 \cdot 10^9\,\mathrm{Pa} \qquad \text{(entspricht 3,9 %)}
\end{aligned}$$

Das Produkt aus Ableitung und Vertrauensbereich muss für jedes x_i die gleiche Dimension der Endgröße haben (hier [G]= Pa)! Hier empfiehlt sich Dimensionskontrolle beim Einsetzen!

Gesamtergebnis: $\underline{\underline{G = (8{,}579 \cdot 10^{10} \pm 3{,}32 \cdot 10^9)\ \mathrm{Pa} \qquad P = 95\,\%}}$

3 Student‘scher t-Test

3.1 t-Test für den Erwartungswert

L 8 Gewicht von Kohlköpfen

Einseitiger Test

Der Test erfolgt einseitig, da (aus Sicht der Ladenkette) höhere Kopfgewichte natürlich zugelassen sind.

Nullhypothese: H_0: $\mu = \mu_0 = 1000\,\mathrm{g}$
Alternativhypothese: H_1: $\mu < \mu_0 = 1000\,\mathrm{g}$

Berechnete Statistiken

Mittelwert: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 960{,}7\,\mathrm{g}$

Streuung: $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = 46{,}5\,\mathrm{g}$

Die Testgröße: $t_0 = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{960{,}7\,\mathrm{g} - 1000\,\mathrm{g}}{46{,}5\,\mathrm{g}/\sqrt{7}} = -2{,}24$

Die Freiheitsgrade: $df = n - 1 = 7 - 1 = 6$

Ablehnung von H_0, wenn $t_0 < -t_{df\,;1-\alpha}$

Den kritischen t-Wert ($t_{df\,;1-\alpha}$) aus Tabelle: $t_{6\,;0{,}95} = 1{,}94$

In unserem Beispiel: $t_0 = -2{,}24 < -t_{6\,;0{,}95} = -1{,}94$

Damit wird H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ abgelehnt.

(Wichtig: Die ursprüngliche Aufgabenstellung beantworten!) ⇒ **Die Chinakohlköpfe des Lieferanten sind signifikant zu leicht.**

Zweiseitiger Test

Der Test erfolgt zweiseitig, wenn wir die Frage so formulieren, daß auch die höheren Kopfgewichte nicht zugelassen sind. (z.B. die Verbraucher tolerieren auch die höhere Kopfgewichte nicht, weil die sich nicht gut in den Kühlschrank legen lassen)

Nullhypothese: $\mathrm{H}_0 : \mu = \mu_0 = 1000\,\mathrm{g}$
Alternativhypothese: $\mathrm{H}_1 : \mu \neq \mu_0 = 1000\,\mathrm{g}$

Berechnete Statistiken:

Mittelwert: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 960{,}7\,\mathrm{g}$

Streuung: $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = 46{,}5\,\mathrm{g}$

Die Testgröße: $t_0 = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{960{,}7\,\mathrm{g} - 1000\,\mathrm{g}}{46{,}5\,\mathrm{g}/\sqrt{7}} = -2{,}24$

Die Freiheitsgrade: $df = n - 1 = 7 - 1 = 6$

Ablehnung von H_0, wenn $|t_0| > t_{df\,;\,1-\alpha/2}$

Der kritische t-Wert $(t_{df\,;\,1-\alpha/2})$ aus Tabelle: $t_{6\,;\,0{,}975} = 2{,}45$

In unserem Beispiel:

$|t_0| = 2{,}24 < t_{6\,;\,0{,}975} = 2{,}45$

Es besteht kein Grund H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ abzulehnen.

Das beobachtete „Untergewicht“ liegt also noch im Rahmen der erwarteten statistischen Schwankung.

L 9 Ozongehalt der Luft

Testproblem:

$\mathrm{H}_0 : \mu_x = \mu_0$ gegen $\mathrm{H}_1 : \mu_x > \mu_0$

$\mathrm{H}_0 : \mu_x = 120\,\mu\mathrm{g}/\,\mathrm{m}^3$ gegen $\mathrm{H}_1 : \mu_x > 120\,\mu\mathrm{g}/\,\mathrm{m}^3$

Statistiken:

$$\overline{x} = \sum_{i=1}^{n}\frac{x_i}{n} = \frac{613}{5} = 122{,}6\frac{\mu\mathrm{g}}{\mathrm{m}^3}$$

$$S = +\sqrt{S^2} = \sqrt{\sum_{i=1}^{n}\frac{(x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{179{,}2}{4}} = 6{,}69\frac{\mu\mathrm{g}}{\mathrm{m}^3}$$

$$\mu_0 = 120\,\frac{\mu\mathrm{g}}{\mathrm{m}^3}$$

Die Anzahl der Freiheitsgrade ist:

$df = n - 1 = 5 - 1 = 4$

Die Testgröße:

$$t_0 = \frac{\overline{x} - \mu_0}{S} \cdot \sqrt{n} = \frac{(122{,}6 - 120)\,\mu\mathrm{g/\,m^3}}{6{,}69\,\mu\mathrm{g/\,m^3}}\sqrt{5} = 0{,}8686$$

$t_{n-1\,;\,1-\alpha} = t_{4\,;\,0{,}95} = 2{,}13$

$t_0 = 0{,}8686 < 2{,}13 = t_{4\,;\,0{,}95}$

Die Nullhypothese kann auf einem Signifikanzniveau von 5 % nicht abgelehnt werden.

Man kann also nicht behaupten, dass der Ozongehalt signifikant größer ist als der Grenzwert von 120 µg/ m³.

3.2 t-Test für Vergleich zweier Erwartungswerte unabhängiger Stichproben

L 10 Futtereinfluss auf Gewichtszunahme bei Schweinen

Da interessiert, ob das Futter mit hohen Proteingehalt höhere Gewichtzunahme liefert, wird die **einseitige Alternativhypothese** gewählt.

Nullhypothese: $\mathrm{H}_0 : \mu_H = \mu_N$
Alternativhypothese: $\mathrm{H}_1 : \mu_H > \mu_N$

Statistiken:

Mittelwerte: $\overline{H} = 697{,}3\,\mathrm{g}$ $\quad \overline{N} = 628{,}5\,\mathrm{g}$

Streuungen: $S_H = 36{,}8\,\mathrm{g}$ $\quad S_N = 47{,}4\,\mathrm{g}$

Die Testgröße:

Weil der Umfang n der beiden Stichproben gleich ist, kann man die einfachere Gleichungsform für t_0 verwenden:

$$t_0 = \sqrt{n}\frac{\overline{H} - \overline{N}}{\sqrt{S_H^2 + S_N^2}} = \sqrt{8}\frac{697{,}3\,\mathrm{g} - 628{,}5\,\mathrm{g}}{\sqrt{36{,}8^2\,\mathrm{g}^2 + 47{,}4^2\,\mathrm{g}^2}} = 3{,}24$$

Freiheitsgrade: $df = n_H + n_N - 2 = 8 + 8 - 2 = 14$

Entscheidung: Ablehnung von H_0, wenn $t_0 > t_{df\,;\,1-\alpha}$

Der kritische t-Wert($t_{df\,;\,1-\alpha}$) aus Tabellen: $t_{14\,;\,0{,}99} = 2{,}62$

In unserem Beispiel: $t_0 = 3{,}24 > t_{14\,;\,0{,}99} = 2{,}62$

Also wird H_0 auf dem Signifikanzniveau $\alpha = 0{,}01$ abgelehnt und die Alternativhypothese H_1 wird angenommen.

Das Futtermittel mit dem höheren Proteingehalt liefert höhere Gewichtszunahme.

L 11 Medikamenteneinfluss auf Tierwachstum

Zweiseitiger t-Test für den Vergleich zweier Erwartungswerte.

Die Nullhypothese lautet:

$H_0 : \mu_\mathrm{x} = \mu_\mathrm{y}$ $H_1 : \mu_\mathrm{x} \neq \mu_\mathrm{y}$

Die Mittelwerte lauten:

$\overline{x} = 72\,\mathrm{cm}$ $\qquad$ $\overline{y} = 75\,\mathrm{cm}$

Die Streuungen sind:

$S_x = S_y = 3{,}9279\,\mathrm{cm}$ (nur zufällig gleich)

Die Anzahl der Freiheitsgrade ist:

$df = n_x + n_y - 2 = 16 - 2 = 14$

zweiseitiger t-Test:

$$t_0 = \frac{\overline{x} - \overline{y}}{\sqrt{S_x^2 + S_y^2}} \cdot \sqrt{n} = \frac{(72 - 75)\,\mathrm{cm}}{\sqrt{30{,}8568}\,\mathrm{cm}} \cdot \sqrt{8} = -1{,}5275$$

$t_{df\,;\,1-\frac{\alpha}{2}} = t_{14\,;\,0{,}975} = 2{,}15$

$|t_0| = 1{,}5275 < 2{,}15 = t_{14\,;\,0{,}975}$

Die Hypothese H_0 kann auf dem Signifikanzniveau von $\alpha = 0{,}05$ nicht abgelehnt werden.

Das Medikament hat somit keinen signifikanten Einfluss auf das Wachstum.

L 12 Ertrag verschiedener Pilzarten

Das Testproblem ist:

$\mathrm{H}_0 : \mu_x = \mu_y$ gegen $\mathrm{H}_1 : \mu_x \neq \mu_y$

Die Mittelwerte sind:

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{62{,}1}{10} = 6{,}21\,\mathrm{kg} \qquad \overline{y} = \sum_{i=1}^{n} \frac{y_i}{n} = \frac{62{,}7}{12} = 5{,}225\,\mathrm{kg}$$

Die Streuungen lauten:

$$S_x = +\sqrt{S_x^2} = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{16{,}929}{9}} = \sqrt{1{,}881} = 1{,}37\,\mathrm{kg}$$

$$S_y = +\sqrt{S_y^2} = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \overline{y})^2}{n-1}} = \sqrt{\frac{1{,}7025}{11}} = \sqrt{0{,}155} = 0{,}393\,\mathrm{kg}$$

$$S = \sqrt{\frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}} = \sqrt{\frac{18{,}63}{20}} \approx 0{,}965\,\mathrm{kg}$$

($S \mathrel{\widehat{=}}$ gewichtete Streuung)

Die Anzahl der Freiheitsgrade beträgt:

$$df = n + m - 2 = 10 + 12 - 2 = 20$$

Die Testgröße ist:

$$t_0 = \frac{\overline{x} - \overline{y}}{S} \sqrt{\frac{n \cdot m}{n+m}} = \frac{0{,}985\,\mathrm{kg}}{0{,}965\,\mathrm{kg}} \sqrt{\frac{120}{22}} = 2{,}38$$

$$t_{n+m-2\,;\,1-\frac{\alpha}{2}} = t_{20\,;\,0{,}975} = 2{,}09$$

$$t_0 = 2{,}38 > 2{,}09 = t_{20\,;\,0{,}975}$$

Damit kann die Nullhypothese zum Signifikanzniveau abgelehnt werden und wir können sagen, dass die Erträge der beiden Pilze auf einem Signifikanzniveau von 5 % unterschiedlich sind.

3.3 t-Test für Vergleich zweier Erwartungswerte verbundener Stichproben

L 13 Einfluss der Benzinqualität auf Treibstoffverbrauch

Die Stichproben sind in diesem Fall verbunden, weil der Verbrauch eines Autotyps mit beiden Benzinmarken in der gleichen Größenordnung liegt, die Unterschiede zwischen verschiedenen Autotypen sind jedoch groß. Aus den beiden Stichproben E_i und A_i werden die Differenzen $d_i = E_i - A_i$ gebildet und man testet die Hypothese, ob die Grundgesamtheit, aus der diese Differenz d_i stammt, den Erwartungswert $\mu_d = 0$ besitzt (in diesem Falle sind natürlich die Stichprobenumfänge automatisch gleich groß, $n_E = n_A = n$).

Da interessiert, ob ein Unterschied in beiden Richtungen entsteht, oder nicht, wird ein **zweiseitiger Test** durchgeführt.

Nullhypothese: $\mathrm{H}_0 : \mu = \mu_d = 0$
Alternativhypothese: $\mathrm{H}_1 : \mu = \mu_d \neq 0$

Der Test läuft weiter, wie beim einfachen t-Test für den Erwartungswert.

Berechnete Statistiken:

Mittelwert: $\overline{d} = -0{,}32\,\mathrm{L}/100\,\mathrm{km}$

Streuung: $S_d = 0{,}217\,\mathrm{L}/100\,\mathrm{km}$

Die Testgröße:

$$t_0 = \frac{\overline{d}}{S_d/\sqrt{n}} = \frac{-0{,}32}{0{,}217/\sqrt{5}} = -3{,}3$$

Die Freiheitsgrade:

$df = n - 1 = 5 - 1 = 4$

Die Entscheidung:

Ablehnung von H_0, wenn $|t_0| > t_{df\,;\,1-\alpha/2}$

Der kritische t-Wert ($t_{df\,;\,1-\alpha/2}$) aus Tabellen: $t_{krit} = t_{4\,;\,0{,}975} = 2{,}78$

Hier: $|t_0| = 3{,}3 > t_{4\,;\,0{,}975} = 2{,}78$

Also wird H_0 auf dem Signifikanzniveau $\alpha = 0{,}05$ abgelehnt, und die Alternativhypothese H_1 wird angenommen.

Anders formuliert: **Die Benzinsorten haben einen signifikanten Qualitätsunterschied.**

L 14 Wirksamkeit von Schlafmitteln

Diese Messungen sind verbundene Stichproben, da die Versuchspersonen individuell verschieden auf Medikamente reagieren. In diesem Fall kann man den Einfluss der Variabilität der Individuen, also deren unterschiedliche Reaktion auf Medikamente, eliminieren. Dies ist der Vorteil des Paarvergleichs. Wird die Gruppe jedoch in zwei Hälften geteilt, wobei jede Hälfte ausschließlich das eine Medikament erhält, dann sind die Stichproben unabhängig.

Vermutung: Die Erwartungswerte der Schlafverlängerung durch das Mittel A und das Mittel B sind gleich.

Nullhypothese: $\mathrm{H}_0 : \mu_A = \mu_B$

Alternativhypothese: $\mathrm{H}_1 : \mu_A \neq \mu_B$

Die Anzahl der Freiheitsgrade ist:

$df = 10 - 1 = 9$

Patient i	$\mathrm{A_i}$	$\mathrm{B_i}$	$\mathrm{d_i = A_i - B_i}$	$\mathrm{d_i}^2$
1	1,9	0,7	1,2	1,44
2	0,8	-1,6	2,4	5,76
3	1,1	-0,2	1,3	1,69
4	0,1	-1,2	1,3	1,69
5	-0,1	-0,1	0,0	0,00
6	4,4	3,4	1,0	1,00
7	5,5	3,7	1,8	3,24
8	1,6	0,8	0,8	0,64
9	4,6	0,0	4,6	21,16
10	3,4	2,0	1,4	1,96
$\sum$	23,3	7,5	15,8	38,58
$\sum/10$	$\overline{\mathrm{A}} = 2{,}33$	$\overline{\mathrm{B}} = 0{,}75$	$\overline{\mathrm{d}} = 1{,}58$	

Die Streuung der Differenzen lautet:

$$S_d^2 = \frac{1}{9} \cdot \left(38{,}58 - \frac{1}{10} \cdot 15{,}8^2\right) = 1{,}513$$

$\Rightarrow S_d \approx 1{,}23$

Die Testgröße ist dann:

$$t_0 = \frac{\overline{d}}{S_d} \cdot \sqrt{n} = \frac{1{,}58}{1{,}23} \cdot \sqrt{10} = 4{,}06$$

Mit $\alpha = 0{,}01 = 1\%$ erhält man $t_{9\,;\,0{,}995} = 3{,}25$.

$|t| = 4{,}06 > 3{,}25 = t_{9\,;\,0{,}995}$

Die Nullhypothese über die Gleichheit der beiden Mittel wird auf dem 1%-Signifikanzniveau abgelehnt.

Die Schlafmittel unterscheiden sich signifikant in ihrer Wirksamkeit.

Es soll an dieser Stelle ausdrücklich darauf hingewiesen werden, dass ein Testergebnis kein Beweis für eine Theorie ist. Es kommt sehr stark auf die Formulierung der Hypothesen, die unterstellten Voraussetzungen und die Versuchsdurchführung an. Das zuerst verabreichte Medikament kann im vorliegenden Fall noch in der folgenden Nacht wirksam sein und so eine Schlaf verlängernde Wirkung des zweiten Medikaments vortäuschen. Dieser Effekt ist der sog. **carry-over**. Es wird also eine Wirkung in eine folgende Versuchsperiode mit hinübertragen. Es wäre eventuell sinnvoll, eine Pause zwischen der ersten und zweiten Verabreichung einzulegen, um einen **wash-out** zu erreichen. Dies könnte allerdings wiederum zur Folge haben, dass die Versuchspersonen nach dieser Zeitspanne physisch und psychisch anders reagieren als unmittelbar nach der ersten Nacht.

4 Einfache Varianzanalyse (ANOVA)

L 15 Wirkung von Fungiziden

Die Gesamtvariation wird durch die Summe der quadrierten Abweichungen aller 15 Beobachtungen von Gesamtmittel $\overline{x}$ erfasst, und mit SQ_{total} bezeichnet. Sie wird in zwei Komponenten zerlegt:

1. Die quadrierten Abweichungen **zwischen** den Gruppen (SQZ) (Abweichungen der Gruppenmittelwerte vom Gesamtmittelwert)

2. Die quadrierten Abweichungen **innerhalb** der Gruppen (SQI)

		Behandlung (Gruppe) $j = 1 \ldots k$			
		1	2	3	
Wiederholungen i	1	5,2	3,9	2,2	
	2	5,2	3,4	2,6	
	3	2,5	2,5	3,4	
	4	4,4	2	4,4	
	5	6,2	1,2	3,4	Gesamt:
$\overline{x}_j$		**4,7**	**2,6**	**3,2**	$\overline{x} = 3{,}5$

Die folgende Gleichung ermöglicht es, SQI als Differenz von SQ_{total} und SQZ zu berechnen, was u. U. schneller geht, wenn Statistikfunktionen des Taschenrechners verwendet werden.

$$SQ_{total} = SQZ + SQI$$

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \overline{x})^2 = \sum_{j=1}^{k} n_j(\overline{x}_j - \overline{x})^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \overline{x_j})^2$$

Hilfsgrößen:

$$n = k \cdot n_j \quad \text{(nur hier!)} \qquad \overline{x} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij} \qquad \overline{x_j} = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij}$$

$SQ_{total} = 26{,}92 \quad (= S^2 \cdot (n-1) = 1{,}38667^2 \cdot 14 = 1{,}923 \cdot 14)$

$SQZ = 11{,}70 \quad (= 5 \cdot (4{,}7 - 3{,}5)^2 + 5 \cdot (2{,}6 - 3{,}5)^2 + 5 \cdot (3{,}2 - 3{,}5)^2)$

$SQI = 15{,}22$

Der Schätzwert S_z^2 für die Varianz zwischen den Gruppen ergibt sich:

$$S_z^2 = MQZ = \frac{SQZ}{FGZ} = \frac{SQZ}{k-1} = \frac{11{,}70}{3-1} = 5{,}85$$

Der Schätzwert S_i^2 für die Varianz innerhalb der Gruppen ergibt sich (unter der Annahme, dass die drei Verteilungen die gleiche Varianz besitzen):

$$S_i^2 = MQI = \frac{SQI}{FGI} = \frac{SQI}{n-k} = \frac{15{,}22}{15-3} = 1{,}27$$

Statistischer Test

Hypothese: Alle Stichproben (Behandlungen) haben den gleichen Erwartungswert

Testgröße: $F_0 = \dfrac{MQZ}{MQI} = \dfrac{5{,}85}{1{,}27} = 4{,}61$

Bestimmung des kritischen Werts:

Anzahl der Freiheitsgrade: $FGZ = k - 1 = 2 \qquad FGI = n - k = 12$

Das geforderte Signifikanzniveau: $\alpha = 0{,}05$

$F_{kritisch} = F_{FGZ\,;\,FGI\,;\,1-\alpha} = F_{2\,;\,12\,;\,0{,}95} = 3{,}89$

Die Entscheidung: Hypothese wird angenommen falls $F_0 < F_{kritisch}$!

In unserem Fall $F_0 = 4{,}61 > F_{kritisch} = 3{,}89$, also können wir die Hypothese bei $\alpha = 0{,}05$ nicht annehmen.

D. h., mindestens ein Fungizid beeinflusst das Pilzwachstum.

5 χ^2-Test

L 16 Unkrautzählung

i	X_i	n_i	$X_i n_i$	$(X_i - \overline{x})^2$	$(X_i - \overline{x})^2 n_i$
1	0	3	0	13,84	41,52
2	1	9	9	7,40	66,59
3	2	15	30	2,96	44,38
4	3	21	63	0,52	10,89
5	4	19	76	0,08	1,49
6	5	15	75	1,64	24,58
7	6	10	60	5,20	51,98
8	7	5	35	10,76	53,79
9	8	3	24	18,32	54,96
$K = 9$		$\mathbf{n = \sum n_i = 100}$	**372**	**60,71**	**350,16**

i	Nummer der Klasse
K	Anzahl der Klassen ($K = 9$)
X_i	Anzahl der gefundenen Unkrautpflanzen für Klasse i
n_i	Anzahl der Teilflächen, die der Klasse X_i zugeordnet werden
Δx	Klassenbreite ($\Delta x = X_{i+1} - X_i = 1$)

Nullhypothese H_0: u ist Poisson-verteilt

Berechnete Statistiken:

Mittelwert: $\overline{x} = \dfrac{\sum (X_i n_i)}{\sum n_i} = 3{,}72$

Empirische Varianz: $S^2 = \dfrac{\sum (X_i - \overline{x})^2 n_i}{n-1} = 3{,}54$

Empirische Streuung: $S = +\sqrt{S^2} = 1{,}88$

Die Poisson-Verteilung $h(x)_{Poisson} = \lambda^x \dfrac{e^{-\lambda}}{x!}$ enthält den Parameter λ, der aus den Messwerten geschätzt wird: $\lambda = \overline{x} = 3{,}72$

Mit diesen Werten erhalten wir für die jeweiligen X_i folgende Werte der Poisson-Verteilung:

X_i	$H_e(X_i)$	$h(X_i)_{Poisson}$	$H_t(X_i)$
0	0,03	0,024	0,024
1	0,09	0,090	0,090
2	0,15	0,168	0,168
3	0,21	0,208	0,208
4	0,19	0,193	0,193
5	0,15	0,144	0,144
6	0,10	0,089	0,089
7	0,05	0,047	0,047
8	0,03	0,022	0,022
$\sum$	**1,00**		**0,986**

$H_e(X_i)$ relative *empirische* Häufigkeit für Klasse $i \Rightarrow H_e(X_i) = n_i/n$. Dieser Wert wird im weiteren aber nicht benötigt und dient nur der Gegenüberstellung zur theoretischen Verteilung

$h(X_i)_{Poisson}$ Wert der Poisson-Verteilung mit Parameter $\mu = \overline{x}$ an der Stelle X_i

$H_t(X_i)$ relative *theoretische* Häufigkeit für Klasse $i \Rightarrow H_t(X_i) = h(x)_{Poisson}\Delta x$ mit $\Delta x = 1$

Die Testgröße $\chi^2 = \sum_{i^*=1}^{K^*} \frac{(B_{i^*} - E_{i^*})^2}{E_{i^*}} = \sum_{i^*=1}^{K^*} \frac{(B_{i^*} - nH_t(X_{i^*}))^2}{nH_t(X_{i^*})}$ ist $\chi^2_{(p\,;\,df)}$-verteilt, wenn H_0 zutrifft.

$nH_t(X_i)$	i^*	B_{i^*}	$nH(X_{i^*}) \geq 10$	$B_{i^*} - nH_t(X_{i^*})$	$\frac{(B_{i^*} - nH_t(X_{i^*}))^2}{nH_t(Xi^*)}$
2,42					
9,02	1	12	11,44	0,56	0,03
16,77	2	15	16,77	-1,77	0,19
20,79	3	21	20,79	0,21	0,00
19,34	4	19	19,34	-0,34	0,01
14,39	5	15	14,39	0,61	0,03
8,92	6	18	15,86	2,14	0,29
4,74					
2,20					
	$K^* = 6$				$\chi^2 = \mathbf{0{,}54}$

i^* Nummer der Klassen nach Zusammenlegung besetzungsschwacher Klassen (In diesem Beispiel sei eine Mindestbesetzung von 10 gefordert)

K^* Reduzierte Zahl der Klassen

df Anzahl Freiheitsgrade $df = K^* - p - 1$ (in unserem Beispiel: $df = 6 - 1 - 1 = 4$)

p Anzahl Parameter (in unserem Beispiel: $p = 1$ (λ))

Wenn wir bei einem Signifikanzniveau $\alpha = 0{,}05$ eine Entscheidung treffen wollen:

Aus Tabellen: Kritisches $\chi^2_{(p=0{,}95\,;\,df=4)} = 9{,}49$

Schlussfolgerung: $\chi^2 < \chi^2_{(0{,}95\,;\,4)} = 9{,}49$

$\Rightarrow$ Auf dem Signifikanzniveau $\alpha = 0{,}05$ ist es nicht gerechtfertigt, die Hypothese H_0 zu verwerfen.

D.h., die Anzahl der Unkrautpflanzen pro Fläche ist vermutlich Poisson-verteilt.

L 17 Milchleistung von Kühen

Messergebnisse und Hilfsgrößen

i	X_i	n_i	$X_i n_i$	$(X_i - \overline{x})^2$	$(X_i - \overline{x})^2 n_i$
1	37,5	5	187,5	211,70	1058,51
2	42,5	9	382,5	91,20	820,82
3	47,5	25	1187,5	20,70	517,56
4	52,5	22	1155,0	0,20	4,46
5	57,5	33	1897,5	29,70	980,18
6	62,5	2	125,0	109,20	218,41
7	67,5	4	270,0	238,70	954,81
K=7		$\mathbf{n = \sum n_i = 100}$	**5205**	**701,42**	**4554,75**

i	Nummer der Klasse
K	Anzahl der Klassen ($K = 7$)
X_i	Klassenmitte der Milchleistung von Kühen in $^{\mathrm{dt}}/_{\mathrm{a}}$
n_i	Beobachtete Klassenhäufigkeiten
Δx	Klassenbreite ($\Delta x = X_{i+1} - X_i = 5\,^{\mathrm{dt}}/_{\mathrm{a}}$)

Nullhypothese H_0: Milchleistung ist normalverteilt

Berechnete Statistiken:

Mittelwert: $\overline{x} = \dfrac{\sum (X_i n_i)}{\sum n_i} = 52{,}05\,^{\mathrm{dt}}/_{\mathrm{a}}$

Empirische Varianz: $S^2 = \dfrac{\sum (X_i - \overline{x})^2 n_i}{n-1} = 46{,}01\,^{\mathrm{dt}^2}/_{\mathrm{a}^2}$

Empirische Streuung: $S = +\sqrt{S^2} = 6{,}78\,^{\mathrm{dt}}/_{\mathrm{a}}$

Die Normal-Verteilung $h(x)_{Normal} = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ enthält die Parameter μ und σ, die aus den Messwerten geschätzt werden:

$\mu = \overline{x} = 52{,}05\,{}^{\mathrm{dt}}/_{\mathrm{a}}$ $\qquad\qquad$ $\sigma = S = 6{,}78\,{}^{\mathrm{dt}}/_{\mathrm{a}}$

Variante 1 (bei kontinuierlichen Verteilungen empfohlen):

Die Testgröße $\chi^2 = \sum_{i=1}^{K^*} \frac{(B_i - E_i)^2}{E_i}$ ist $\chi^2_{(p\,;\,df)}$-verteilt, wenn H_0 zutrifft.

i	x_{g_i}	$\Phi(x_{g_i})$	$\Phi(x_{g_i}) - \Phi(x_{g_{i-1}})$	$E_i = n(\Phi(x_{g_i}) - \Phi(x_{g_{i-1}}))$	i^*	$E_{i^*} > 5$	B_{i^*}	$B_{i^*} - E_{i^*}$	$\frac{(B_{i^*} - E_{i^*})^2}{E_{i^*}}$
1	35	0,00587							
2	40	0,03754	0,0317	3,17					
3	45	0,14917	0,1116	11,16	1	14,33	14	-0,33	0,01
4	50	0,38239	0,2332	23,32	2	23,32	25	1,68	0,12
5	55	0,67003	0,2876	28,76	3	28,76	22	-6,76	1,59
6	60	0,87900	0,209	20,9	4	20,9	33	12,1	7,01
7	65	0,97193	0,0929	9,29	5	11,7	6	-5,7	2,78
8	70	0,99598	0,0241	2,41					
					$K^* = 5$				$\boldsymbol{\chi^2 = 11{,}51}$

x_{g_i}	Intervallgrenzen der jeweiligen Klasse, hier z.B. $37{,}5 \pm 2{,}5$ usw.
i^*	Nummer der Klassen nach Zusammenlegung besetzungsschwacher Klassen
K^*	Reduzierte Zahl der Klassen
df	Anzahl Freiheitsgrade $df = K^* - s - 1$ (in unserem Beispiel: $df = 5 - 2 - 1 = 2$)
s	Anzahl Parameter (in unserem Beispiel: $s = 2$ (μ,σ))

Für Signifikanzniveau von $\alpha = 0{,}05$ aus Tabellen:

Kritisches $\chi^2_{(p=0{,}95\,;\,df=2)} = 5{,}99$

Schlussfolgerung: $\chi^2 = 11{,}51 > \chi^2_{(0{,}95\,;\,2)} = 5{,}99$

$\Rightarrow$ Bei einem Signifikanzniveau $\alpha = 0{,}05\%$ können wir die Hypothese H_0 nicht akzeptieren, d.h. die Stichprobe kommt nicht von einer normalverteilten Grundgesamtheit.

Variante 2:

Es wird folgender vereinfachter Ansatz gemacht:

$H_t(X_i) = h(X_i)\Delta x$ mit $\Delta x = 5\,\mathrm{dt}/\mathrm{a}$

Es wird also die Verteilungsdichte in der Intervallmitte mit der Intervallbreite multipliziert, um die Fläche unter der Verteilungsdichtefunktion, und somit die relativen Häufigkeiten H_t in diesem Intervall, anzunähern. **Dieses Verfahren sollte vermieden werden** und die Variante 1 verwendet werden. Es wird aber trotzdem kurz erläutert, da es in einigen Übungsaufgaben und alten Klausuren zu finden ist.

Mit den zuvor berechneten Werten für Mittelwert und Streuung ergeben sich folgende Werte $h(x)$ der Normalverteilung:

X_i	$H_e(X_i)$	$h(X_i)_{Normal}$	$H_t(X_i)$
37,5	0,05	0,0058	0,029
42,5	0,09	0,0218	0,109
47,5	0,25	0,0469	0,235
52,5	0,22	0,0587	0,294
57,5	0,33	0,0426	0,213
62,5	0,02	0,0179	0,090
67,5	0,04	0,0043	0,022
$\sum$	**1,00**		**0,992**

$H_e(X_i)$ relative *empirische* Häufigkeit für Klasse i $\Rightarrow H_e(X_i) = n_i/n$

$h(X_i)_{Normal}$ Wert der Normalverteilung mit Parametern $\mu = \overline{x}, \sigma = s$ an der Stelle X_i

$H_t(X_i)$ relative *theoretische* Häufigkeit für Klasse i
$\Rightarrow H_t(X_i) = h(X_i)_{Normal}\Delta x$
mit $\Delta x = 5\,\mathrm{dt}/\mathrm{a}$

Die Testgröße $\chi^2 = \sum_{i^*=1}^{K^*} \frac{(B_i - E_i)^2}{E_i} = \sum_{i^*=1}^{K^*} \frac{(B_{i^*} - nH_t(X_{i^*}))^2}{nH_t(X_{i^*})}$ ist $\chi^2_{(p\,;\,df)}$-verteilt, wenn H_0 zutrifft.

$nH_t(X_i)$	i^*	B_{i^*}	$nH_t(X_{i^*}) > 5$	$B_{i^*} - nH_t(X_{i^*})$	$\frac{(B_{i^*} - nH_t(X_{i^*}))^2}{nH(X_{i^*})}$
2,94					
10,91	1	14	13,85	0,148	0,002
23,49	2	25	23,49	1,512	0,097
29,36	3	22	29,36	-7,356	1,843
21,30	4	33	21,30	11,702	6,429
8,97	5	6	11,16	-5,163	2,388
2,19					
	$K^* = 5$				$\chi^2 = \mathbf{10{,}76}$

Für Signifikanzniveau von $\alpha = 0{,}05$ aus Tabellen:

Kritisches $\chi^2_{(p=0{,}95\,;\,df=2)} = 5{,}99$

Schlussfolgerung: $\chi^2 = 10{,}76 > \chi^2_{(0{,}95\,;\,2)} = 5{,}99$

$\Rightarrow$ Bei einem Signifikanzniveau $\alpha = 0{,}05$ können wir die Hypothese H_0 nicht akzeptieren, d.h. die Stichprobe kommt nicht von einer normalverteilten Grundgesamtheit.

L 18 Punktzahlen der Statistikklausur

Hier ist die Verteilungsfunktion nicht als Formel, sondern als Tabelle mit der relativen Häufigkeit gegeben.

H_0: Punktzahl ist kryptonormativ-polydiskriminalverteilt :c)

B_i	H_t	$H_t \cdot n = E_i$	$B_i - E_i$	$(B_i - E_i)^2$	$\frac{(B_i - E_i)^2}{E_i}$
6	0,1	10	4	16	1,60
15	0,2	20	5	25	1,25
39	0,3	30	9	81	2,70
40	0,4	40	0	0	0
					$\chi_0^2 = 5{,}55$

Die Verteilungsfunktion hat keine Parameter.

$\Rightarrow s = 4 - 1 = 3$

$p = 0{,}95$

$\chi^2_{krit} = 7{,}81$

$\chi^2_0 < \chi^2_{krit}$

$\Rightarrow$ H_0 ist nicht abzulehnen.

$\Rightarrow$ K. hat recht, Punktzahl ist k.-p.-verteilt.

6 Lineare Regression

L 19 Ertrag von Sommerweizen

i	x_i	y_i	${x_i}^2$	${y_i}^2$	$x_i \cdot y_i$
1	80	56,2	6400	3158,44	4496
2	90	55,7	8100	3102,49	5013
3	110	75,5	12100	5700,25	8305
4	90	55,7	8100	3102,49	5013
5	110	68,4	12100	4678,56	7524
6	130	67,7	16900	4583,29	8801
7	90	63,3	8100	4006,89	5697
8	110	58,3	12100	3398,89	6413
9	130	80,8	16900	6528,64	10504
10	70	52,1	4900	2714,41	3647
11	150	87,3	22500	7621,29	13095
$\sum$	1160	721,0	128200	48595,64	78508

Die Mittelwerte sind:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{11} \cdot 1160 = 105{,}5\frac{\mathrm{kg}}{\mathrm{ha}}$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{11} \cdot 721 = 65{,}5\frac{d\mathrm{t}}{\mathrm{ha}}$$

Die Streuungen sind:

$$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sqrt{\frac{1}{10}5872{,}75} = 24{,}233\frac{\mathrm{kg}}{\mathrm{ha}}$$

$$S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2} = \sqrt{\frac{1}{10}1337{,}39} = 11{,}565\frac{d\mathrm{t}}{\mathrm{ha}}$$

Nach der Methode der kleinsten quadratischen Abweichungen geht die Gerade durch den Schwerpunkt der Punkte $(x_i; y_i)$

$$(y - \overline{y}) = b(x - \overline{x})$$

$$\Rightarrow Y_i = b(x_i - \overline{x}) + \overline{y}$$

Geschätzter Regressionskoeffizient b (Gleichung aus Formelsammlung umgeformt):

$$\frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} = \frac{78508 - 11 \cdot 105{,}5 \cdot 65{,}5}{128200 - 11 \cdot 105{,}5^2} = 0{,}421\frac{dt}{\mathrm{kg}}$$

Der Ordinatenabschnitt der Regressionsgeraden ist:

$$y_0 = \overline{y} - b_1\overline{x} = 65{,}5 - 0{,}421 \cdot 105{,}5\frac{dt}{\mathrm{ha}} = 21{,}1\frac{dt}{\mathrm{ha}}$$

Die Regressionsgerade lautet also:

$$Y_i = 21{,}1\frac{dt}{\mathrm{ha}} + 0{,}421\frac{dt}{\mathrm{kg}} \cdot x_i$$

Die Restvarianz $\hat{\sigma}^2$ (ein Schätzwert für σ^2):

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - Y_i)^2 = \frac{1}{9} \cdot 294{,}1022 = 32{,}678\left(\frac{dt}{\mathrm{ha}}\right)^2$$

$$\Rightarrow \hat{\sigma} = +\sqrt{\hat{\sigma}^2} = 5{,}72\frac{dt}{\mathrm{ha}}$$

Bestimmung der Vertrauensgrenze für diese Schätzung des Steigungsmaßes b: (die geforderte statistische Sicherheit $P\% = (1-\alpha) = 0{,}99$)

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

In unserem Beispiel:

$$\left[0{,}421\frac{dt}{\mathrm{kg}} - \frac{5{,}72 \cdot t_{9\,;\,0{,}995}}{\sqrt{11} \cdot 24{,}233}\frac{dt}{\mathrm{kg}}\,;\,0{,}421\frac{dt}{\mathrm{kg}} + \frac{5{,}72 \cdot t_{9\,;\,0{,}995}}{\sqrt{11} \cdot 24{,}233}\frac{dt}{\mathrm{kg}}\right]$$

Aus Tabellen: $t_{9\,;\,0{,}995} = 3{,}25$

$$\left[(0{,}421 - 0{,}231)\frac{dt}{\mathrm{kg}}; (0{,}421 + 0{,}231)\frac{dt}{\mathrm{kg}}\right]$$

$$\left[0{,}19\frac{dt}{\mathrm{kg}}; 0{,}652\frac{dt}{\mathrm{kg}}\right]$$

Der Erwartungswert β liegt mit der statistischen Sicherheit von P % =0,99 in diesem Intervall.

Der Vertrauensbereich für Y:

Durch die errechnete Gerade können wir einem beliebig gewählten x^*-Wert den $Y^* = b(x^* - \overline{x}) + \overline{y}$ zuordnen.

$$Y^* - \frac{\hat{\sigma} \cdot t_{n-2\,;\,1-\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{1 + \frac{(x^* - \overline{x})^2}{S_x^2}}; \; Y^* + \frac{\hat{\sigma} \cdot t_{n-2\,;\,1-\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{1 + \frac{(x^* - \overline{x})^2}{S_x^2}}$$

In unserem Beispiel:

P % =0,99 $\qquad t_{9\,;\,0{,}995} = 3{,}25$

$$x_1^* = 70 \frac{\mathrm{kg}}{\mathrm{ha}} \qquad x_2^* = 150 \frac{\mathrm{kg}}{\mathrm{ha}}$$

$$Y_1^* = 50{,}57 \qquad Y_2^* = 84{,}25$$

Vertrauensbereich 1:

$$[(50{,}57 - 9{,}942)\,\mathrm{kg}\,;\; (50{,}57 + 9{,}942)\,\mathrm{kg}]$$

$$[40{,}628\,\mathrm{kg}\,;\; 60{,}512\,\mathrm{kg}]$$

Vertrauensbereich 2:

$$[(84{,}25 - 11{,}72)\,\mathrm{kg}\,;\; (84{,}25 + 11{,}72)\,\mathrm{kg}]$$

$$[72{,}53\,\mathrm{kg}\,;\; 95{,}97\,\mathrm{kg}]$$

L 20 Zusammenhang von Körpergröße und Gewicht

Nach der Methode der kleinsten quadratischen Abweichungen geht diese Gerade durch den Schwerpunkt $(\overline{x}; \overline{y})$ der Punkte $(x_i; y_i)$.

$$(Y - \overline{y}) = b(x - \overline{x})$$

$$Y_i = b(x_i - \overline{x}) + \overline{y}$$

b: Regressionskoeffizient (Steigung der Geraden)

Y_i: der berechnete Gleichungswert an der Stelle x_i

Regressionskoeffizient:

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Restvarianz:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - Y_i)^2}{n-2}$$

Berechnete Statistiken:

Mittelwerte:

$$\overline{x} = \frac{\sum x_i}{n} = \frac{3460}{20} = 173\,\mathrm{cm}$$

$$\overline{y} = \frac{\sum y_i}{n} = \frac{1359}{20} = 67{,}95\,\mathrm{kg}$$

Streuungen:

$$S_x = \frac{442}{20-1} = 4{,}82\,\mathrm{cm}$$

$$S_y = \frac{1036{,}95}{20-1} = 7{,}39\,\mathrm{kg}$$

Zunächst weitere Hilfsgrößen berechnen...

Die Urliste und die berechenten Hilfsgrößen tabellarisch dargestellt:

	Größe	**Gewicht**	**Hilfsgrößen**						
i	x_i	y_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$y_i - \overline{y}$	$(y_i - \overline{y})^2$	$(x_i - \overline{x})(y_i - \overline{y})$	Y_i	$(y_i - Y_i)^2$
1	165	56	-8,00	64,00	-11,95	142,80	95,60	55,95	0,0025
2	176	75	3,00	9,00	7,05	49,70	21,15	72,45	6,5025
3	175	70	2,00	4,00	2,05	4,20	4,10	70,95	0,9025
4	168	61	-5,00	25,00	-6,95	48,30	34,75	60,45	0,3025
5	167	61	-6,00	36,00	-6,95	48,30	41,70	58,95	4,2025
6	172	63	-1,00	1,00	-4,95	24,50	4,95	66,45	11,9025
7	175	72	2,00	4,00	4,05	16,40	8,10	70,95	1,1025
8	180	80	7,00	49,00	12,05	145,20	84,35	78,45	2,4025
9	179	76	6,00	36,00	8,05	64,80	48,30	76,95	0,9025
10	173	68	0,00	0,00	0,05	0,00	0,00	67,95	0,0025
11	166	57	-7,00	49,00	-10,95	119,90	76,65	57,45	0,2025
12	178	76	5,00	25,00	8,05	64,80	40,25	75,45	0,3025
13	169	60	-4,00	16,00	-7,95	63,20	31,80	61,95	3,8025
14	169	64	-4,00	16,00	-3,95	15,60	15,80	61,95	4,2025
15	170	63	-3,00	9,00	-4,95	24,50	14,85	63,45	0,2025
16	176	71	3,00	9,00	3,05	9,30	9,15	72,45	2,1025
17	180	78	7,00	49,00	10,05	101,00	70,35	78,45	0,2025
18	169	62	-4,00	16,00	-5,95	35,40	23,80	61,95	0,0025
19	177	75	4,00	16,00	7,05	49,70	28,20	73,95	1,1025
20	176	71	3,00	9,00	3,05	9,30	9,15	72,45	2,1025
$\sum$	**3460**	**1359**	**0,00**	**442,00**	**0,00**	**1036,95**	**663,00**		**42,45**

Geschätzter Regressionskoeffizienz b:

$$b = \frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2} = \frac{663}{442} = 1{,}5\frac{\mathrm{kg}}{\mathrm{cm}}$$

Die Restvarianz $\hat{\sigma}^2$ (ein Schätzwert für σ^2)

$$\hat{\sigma}^2 = \frac{\sum\limits_{i=1}^{n}(y_i - Y_i)^2}{n-2} = \frac{42{,}45\,\mathrm{kg}^2}{20-2} = 2{,}36\,\mathrm{kg}^2 \quad \Rightarrow \quad \hat{\sigma} = \sqrt{\hat{\sigma}^2} = 1{,}54\,\mathrm{kg}$$

Bestimmung der Vertrauensgrenze für diese Schätzung des Steigungsmaßes b: (die geforderte statistische Sicherheit $P\% = 1 - \alpha = 0{,}95$)

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

In unserem Beispiel:

$$\left[1{,}5 - \frac{1{,}54 \cdot t_{18\,;\,0{,}975}}{\sqrt{20} \cdot 4{,}82}; 1{,}5 + \frac{1{,}54 \cdot t_{18\,;\,0{,}975}}{\sqrt{20} \cdot 4{,}82}\right]$$

Aus Tabellen: $t_{18\,;\,0{,}975} = 2{,}10$

$$\left[(1{,}5 - 0{,}15)\frac{\mathrm{kg}}{\mathrm{cm}}; (1{,}5 + 0{,}15)\frac{\mathrm{kg}}{\mathrm{cm}}\right]$$

Der Erwartungswert β für den Regressionskoeffizienten b liegt also mit der statistischen Sicherheit $P\% = 0{,}95$ in diesem Intervall.

b)

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

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

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

In unserem Beispiel: $P\% = 0{,}95$ $\quad t_{18\,;\,0{,}975} = 2{,}1$ $x_1^* = 170{,}00\,\mathrm{cm}$ $\quad Y_1^* = 63{,}45\,\mathrm{kg}$ Vertrauensbereich: $[(63{,}45 - 0{,}85)\,\mathrm{kg}; (63{,}45 + 0{,}85)\,\mathrm{kg}]$

$x_2^* = 172{,}00\,\mathrm{cm}$ $\quad Y_2^* = 66{,}45\,\mathrm{kg}$ Vertrauensbereich: $[(66{,}45 - 0{,}74)\,\mathrm{kg}; (66{,}45 + 0{,}74)\,\mathrm{kg}]$

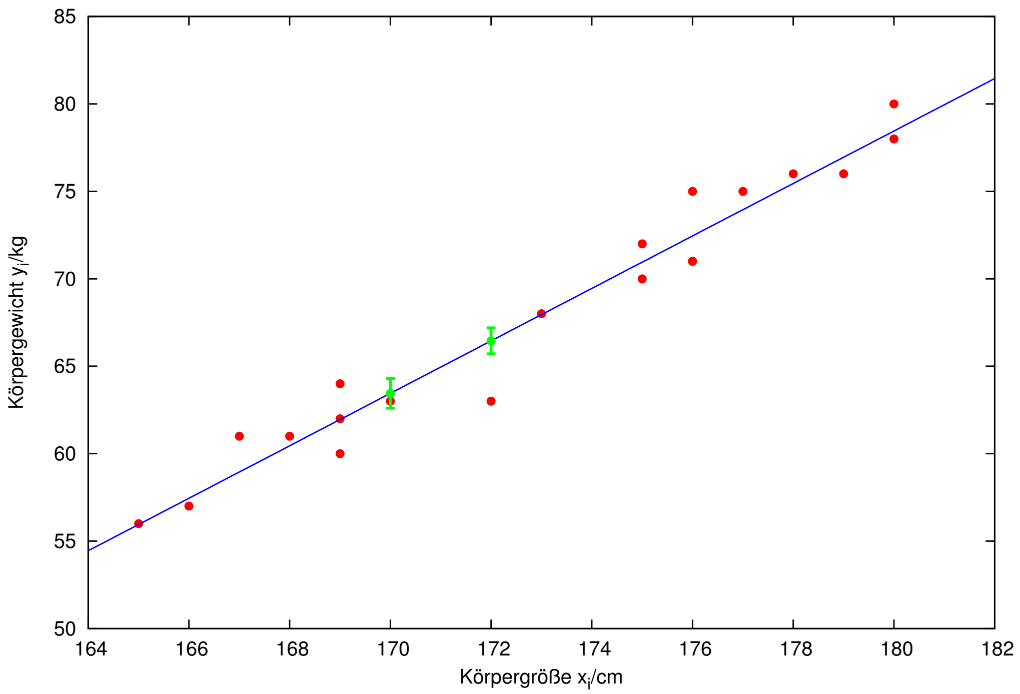


Bild 1: Vertrauensbereiche

7 Aufgabenvarianten zur statistischen Sicherheit

Aufgabenvarianten zur statistischen Sicherheit

Gegeben: μ, σ (Erwartungswert und Standardabweichung)

Mit welcher Wahrscheinlichkeit liegt *ein* nächster Messwert in einem Intervall $[a; b]$?
$\Phi(z(b)) - \Phi(z(a))$ mit $z = \frac{x-\mu}{\sigma}$ (z.B. $z = \frac{a-\mu}{\sigma}$)

Mit welcher Wahrscheinlichkeit ist *ein* nächster Messwert größer als a? $1 - \Phi(a)$

In welchem Intervall $[\mu \pm c]$ liegt *ein* nächster Messwert mit stat. Sicherheit $p = 1 - \alpha$?
$c = k \cdot \sigma$ k ist abhängig von p (Tab. 2.2), c entspricht dem $x - \mu \Rightarrow k = z$

Aufgabenvarianten zur statistischen Sicherheit

Gegeben: x, σ (ein Messwert und Standardabweichung)

In welchem Intervall $[x \pm c]$ liegt μ mit stat. Sicherheit $p = 1 - \alpha$?

$c = k \cdot \sigma$ k ist abhängig von p (Tab. 2.2), c entspricht dem $x - \mu \Rightarrow k = z$

Aufgabenvarianten zur statistischen Sicherheit

Gegeben: $x_1 \ldots x_n \rightarrow \overline{x}, \sigma$ (Mittelwert aus Stichprobe mit n Elementen und Standardabweichung)

In welchem Intervall $[\overline{x} \pm c]$ liegt μ mit stat. Sicherheit $p = 1 - \alpha$? $c = k \cdot \sigma_{\overline{x}} = k \cdot \frac{\sigma}{\sqrt{n}}$ k ist abhängig von p (Tab. 2.2)

Wie viele Messungen sind notwendig, damit das Vertrauensintervall $c < c_k$ ist ($p = 0{,}95$)?
$c = 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} < c_k \quad \rightarrow \quad n > \left(\frac{1{,}96 \cdot \sigma}{c_k}\right)^2$

Wie viele Messungen sind notwendig, damit μ mit einer statistischen Sicherheit von mindestens $p = 0{,}95$ im Intervall $[\overline{x} \pm c]$ liegt? $c = k \cdot \frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad n > \left(\frac{1{,}96 \cdot \sigma}{c}\right)^2$

Aufgabenvarianten zur statistischen Sicherheit

Gegeben: $x_1 \ldots x_n \rightarrow \overline{x}, S$ (Mittelwert und Streuung aus Stichprobe mit n Elementen)

In welchem Intervall $[\overline{x} \pm c]$ **liegt** μ **mit stat. Sicherheit** $p = 1 - \alpha$**?** $c = \dfrac{t_{n-1\,;\,\frac{\alpha}{2}} \cdot S}{\sqrt{n}}$ t ist abhängig von p (bzw. α) und n **Warum** $\dfrac{\alpha}{2}$**?** Aus Symmetrie folgt Wahrscheinlichkeit p, dass ein Wert im Intervall $\mu \pm c$ ist:

$$z_c = \frac{c - \mu}{\sigma}$$

$$p = \Phi(z) - \Phi(-z) = 1 - \Phi(-z) - \Phi(-z) = 1 - 2\Phi(-z)$$

mit $p = 1 - \alpha$ folgt: $\alpha = 2\Phi(-z)$ und damit $\Phi(-z) = \alpha/2$

$\Phi(-z)$ ist jetzt genau die Fläche an einer Seite außerhalb des Vertrauensbereiches $\mu \pm z$.