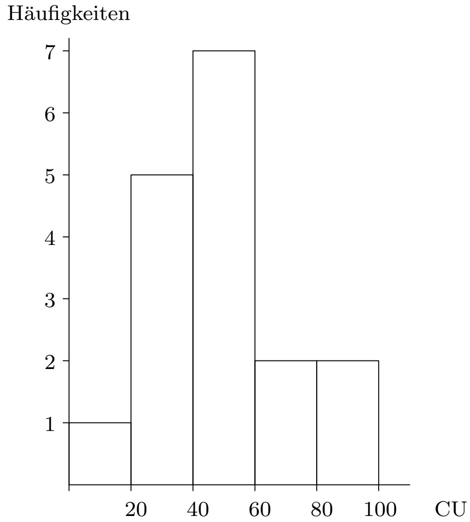
Mathematik für Naturwissenschaftler, Pruscha & Rost
Kap 10 Lösungen

1. **a)**

Für das Histogramm verwenden wir die Intervalle

$$[0,20)\ ,\ [20,40)\ ,\ [40,60)\ ,\ [60,80)\ ,\ [80,100].$$



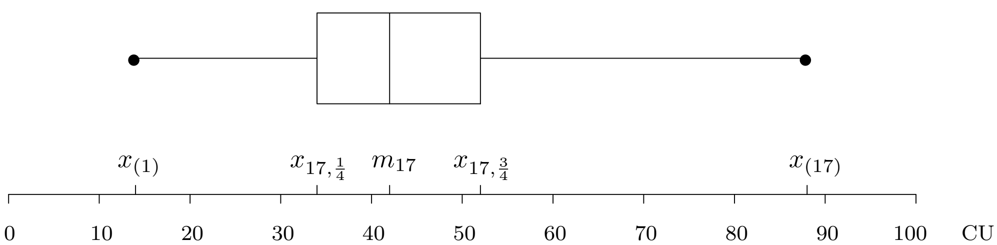
b) Geordnete Stichprobe:

14 25 29 31 34 37 40 41 42 44 49 51 52 62 71 88 88

Mit $n = 17$ ist

$$\begin{array}{llll} k_1 = [17 \cdot \frac{1}{4}] + 1 = 5 & \Longrightarrow & x_{17,\frac{1}{4}} = x_{(5)} = 34 & \text{(1. Quartil)} \\ k_2 = [17 \cdot \frac{3}{4}] + 1 = 13 & \Longrightarrow & x_{17,\frac{3}{4}} = x_{(13)} = 52 & \text{(2. Quartil)} \\ k_3 = \frac{17+1}{2} = 9 & \Longrightarrow & m_{17} = x_{(9)} = 42 & \text{(Median)} \end{array}$$

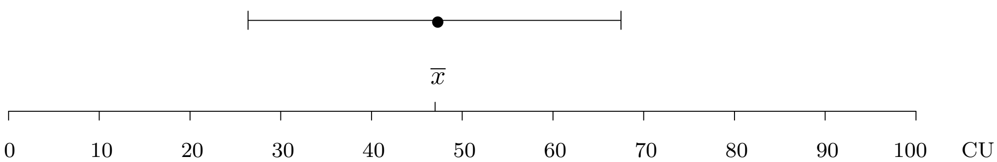
Boxplot:



c) Es ist

$$\overline{x} = \frac{1}{17} \cdot 798 = 46.941 \qquad s = \sqrt{421.809} = 20.538$$

Fehlerbalken:



$\boxed{\textbf{2.}}$

Es ist

$$\overline{x^*} = \frac{1}{n}\sum_{i=1}^{n} x_i^* \;=\; \frac{1}{s\,n}\sum_{i=1}^{n}(x_i - \overline{x}) \;=\; \frac{1}{s}\Big(\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\cdot n\,\overline{x}\Big) \;=\; \frac{1}{s}(\overline{x} - \overline{x}) = 0 \quad \text{und}$$

$$s_{x^*}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^* - \overline{x^*})^2 \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(x_i^*)^2 \;=\; \frac{1}{s^2\cdot(n-1)}\sum_{i=1}^{n}(x_i - \overline{x})^2 \;=\; \frac{1}{s^2}\,s^2 \;=\; 0.$$

$\boxed{\textbf{3.}}$

Die bivariate Stichprobe (x_i, y_i) vom Umfang $n = 8$ lautet

$$(x_i) = (112, 122, 82, 137, 224, 55, 90, 126) \quad RB$$
$$(y_i) = (44, 31, 52, 42, 49, 14, 40, 88) \quad CU$$

a) Die Mittelwerte der beiden Stichproben sind

$$\bar{x} \;=\; \sum_{i=1}^{8} x_i/8 \;=\; 948.0/8 \;=\; 118.5 \quad , \quad \bar{y} \;=\; \sum_{i=1}^{8} y_i/8 \;=\; 360.0/8 \;=\; 45.0\,.$$

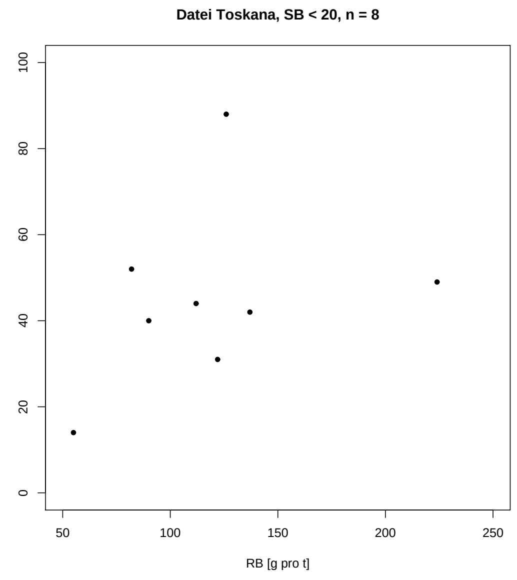
b) Die Summen der Quadrate und der Kreuzprodukte sind

$$\sum_{i=1}^{8} x_i^2 \;=\; 130098.0 \quad , \quad \sum_{i=1}^{8} y_i^2 \;=\; 19306.0 \quad , \quad \sum_{i=1}^{8} x_i\cdot y_i \;=\; 45162.0\,.$$

c) Der Korrelationskoeffizient berechnet sich nach der Rechenformel zu

$$r_{x,y} \;=\; \frac{45162 - 8\cdot 118.5\cdot 45.0}{\sqrt{(130098 - 8\cdot 118.5^2)\cdot(19306 - 8\cdot 45^2)}} \;=\; \frac{2502}{\sqrt{17760\cdot 3106}} \;=\; 0.33687\,.$$

d) Das $x - y$ Scattergramm gibt die folgende Abbildung wieder:



$\boxed{\textbf{4.}}$

Es ist $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 \;=\; \dfrac{1}{n-1}\Big[\sum_{i=1}^{n}x_i^2 - 2\overline{x}\sum_{i=1}^{n}x_i + n\cdot\overline{x}^2\Big]$

$$= \frac{1}{n-1}\Big[\sum_{i=1}^{n}x_i^2 - 2\overline{x}\,n\,\overline{x} + n\cdot\overline{x}^2\Big] \;=\; \frac{1}{n-1}\Big[\sum_{i=1}^{n}x_i^2 - n\cdot\overline{x}^2\Big] \quad \text{und}$$

$$s_{x,y} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})\cdot(y_i-\overline{y}) \;=\; \frac{1}{n-1}\Big[\sum_{i=1}^{n}x_iy_i - \overline{y}\sum_{i=1}^{n}x_i - \overline{x}\sum_{i=1}^{n}y_i + n\cdot\overline{x}\cdot\overline{y}\Big]$$

$$= \frac{1}{n-1}\Big[\sum_{i=1}^{n}x_iy_i - n\cdot\overline{x}\cdot\overline{y} - n\cdot\overline{x}\cdot\overline{y} + n\cdot\overline{x}\cdot\overline{y}\Big] \;=\; \frac{1}{n-1}\Big[\sum_{i=1}^{n}x_iy_i - n\cdot\overline{x}\cdot\overline{y}\Big].$$

$\boxed{\textbf{5.}}$

Aufgrund der Satzes in 9.2.4 gilt

$$Var(\overline{X}_n) = Var\big(\frac{1}{n}\sum_{i=1}^{n}X_i\big) \;=\; \frac{1}{n^2}Var\big(\sum_{i=1}^{n}X_i\big) \;=\; \frac{1}{n^2}\sum_{i=1}^{n}Var(X_i) \;=\; \frac{1}{n^2}\cdot n\cdot\sigma^2 \;=\; \frac{1}{n}\,\sigma^2$$

$$\mathbb{E}(S_n^2) = \mathbb{E}\Big(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\Big) \;=\; \frac{1}{n-1}\mathbb{E}\Big(\sum_{i=1}^{n}(X_i^2 - 2X_i\,\overline{X}_n + \overline{X}_n^2)\Big)$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}(X_i^2) - \frac{2}{n-1}\sum_{i=1}^{n}\mathbb{E}(X_i\overline{X}_n) + \frac{1}{n-1}n\,\mathbb{E}(\overline{X}_n^2).$$

Es ist

$$\mathbb{E}(X_i^2) = Var(X_i^2) + \mathbb{E}(X_i)^2 = \sigma^2 + \mu^2$$

$$\mathbb{E}(X_i\overline{X}_n) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(X_iX_j) \;=\; \frac{1}{n}\Big(\mathbb{E}(X_i^2) + \sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{E}(X_iX_j)\Big)$$

$$\underset{(9.19)}{=} \frac{1}{n}\Big(\sigma^2+\mu^2+\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{E}(X_i)\mathbb{E}(X_j)\Big) \;=\; \frac{1}{n}\big(\sigma^2+\mu^2+\mu^2(n-1)\big) \;=\; \frac{1}{n}(\sigma^2+n\mu^2)$$

$$\mathbb{E}(\overline{X}_n^2) = Var(\overline{X}_n) + \mathbb{E}(\overline{X}_n)^2 \;=\; \frac{1}{n}\sigma^2+\mu^2,$$

so dass man nach dem Einsetzen erhält

$$\mathbb{E}(S_n^2) = \frac{n}{n-1}(\sigma^2+\mu^2) \;-\; \frac{2}{n-1}(\sigma^2+n\mu^2) + \frac{n}{n-1}(\frac{1}{n}\sigma^2+\mu^2)$$

$$= \frac{n}{n-1}\sigma^2 - \frac{2}{n-1}\sigma^2 + \frac{1}{n-1}\sigma^2 \;=\; \sigma^2\big(\frac{n}{n-1} - \frac{1}{n-1}\big) \;=\; \sigma^2.$$

$\boxed{\textbf{6.}}$

Die Anzahl X der `grünen` Pflanzen der F_2-Generation ist $B(n,p)$-verteilt mit $p =$ „Einzelwahrscheinlichkeit“ für `grün` und $n = 353+125 = 478$. In der Stichprobe sind $x = 125$ `grüne`

Pflanzen.

a) Es ist $\quad \hat{p} = \frac{x}{n} \;=\; \frac{125}{478} = 0.2615, \qquad se(\hat{p}) = \sqrt{\frac{1}{n}\hat{p}\cdot(1-\hat{p})} \;=\; \sqrt{0.0004} = 0.0201.$

b) Mit $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.960$ lautet das (für große n gültige) 95%-Konfidenzintervall für p

$$\hat{p} - u_{1-\frac{\alpha}{2}} \cdot se(\hat{p}) \leq \; p \; \leq \hat{p} + u_{1-\frac{\alpha}{2}} \cdot se(\hat{p}),$$

also $\quad 0.2221 \leq p \leq 0.3009.$

c) Mit $p_0 = \frac{1}{4}$ lautet die (für große n gültige) Teststatistik

$$T \;=\; \sqrt{n}\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}} \;=\; \sqrt{478}\frac{0.0115}{\sqrt{0.1875}} \;=\; 0.5806.$$

i) $\quad |T| \not> 1.960 \;=\; u_{1-\frac{\alpha}{2}} \qquad \Longrightarrow \qquad H_0 \; (p = p_0)$ wird nicht verworfen
ii) $\quad T \not> 1.645 \;=\; u_{1-\alpha} \qquad \Longrightarrow \qquad H_0 \; (p \leq p_0)$ wird nicht verworfen.

$\boxed{7.}$ **a)**

Gemäß 10.3.3 bildet unter der dortigen Annahme ($X_1, ..., X_n$ unabhängig und $N(\mu, \sigma^2)$-verteilt)

$$\overline{x} - t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \leq \; \mu \leq \overline{x} + t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

ein Konfidenzintervall für μ zum Niveau $1-\alpha$. F'ür große n können wir $t_{n-1,1-\frac{\alpha}{2}}$ durch $u_{1-\frac{\alpha}{2}}$, das $1-\frac{\alpha}{2}$-Quantil der $N(0,1)$-Verteilung, ersetzen. Dann ist

$$\overline{x} - 1 \cdot \frac{s}{\sqrt{n}} \leq \; \mu \leq \overline{x} + 1 \cdot \frac{s}{\sqrt{n}}$$

ein Konfidenzintervall für μ zum Niveau $1-\alpha \approx 0.682$, denn für α gilt $u_{1-\frac{\alpha}{2}} = 1$, also $\alpha \approx 0.318$ (beachte: $u_{1-\frac{\alpha}{2}} = 1 \iff 1 - \frac{\alpha}{2} = \Phi(1) = 0.841$, mit Φ die Verteilungsfunktion der $N(0,1)$-Verteilung).

b) Die Angabe $\overline{x} \pm s$ bezieht sich auf eine einzelne Beobachtung X_i („Schwankung" von X_i), $\overline{x} \pm \frac{s}{\sqrt{n}}$ bezieht sich auf den Mittelwert $\overline{x}$ („Schwankung" von $\overline{x}$).

$\boxed{8.}$ **a)**

In der Situation von 10.3.3 ($X_1, ..., X_n$ unabhängig und $N(\mu, \sigma^2)$-verteilt) lautet die univariate Stichprobe (x_i) vom Umfang $n = 12$

$$(x_i) = (4.10, 4.31, 4.05, 4.38, 4.23, 4.24, 4.63, 4.29, 4.10, 3.91, 4.51, 4.86).$$

Man berechnet den Mittelwert und die Standardabweichung der Stichprobe zu

$$\bar{x} \;=\; 4.3008\,, \qquad s_x \;=\; 0.2656\,.$$

a) Zum Prüfen der Hypothese $H_0 : \mu = 4.0$ berechnet man die Teststatistik

$$t \;=\; \sqrt{n}\,\frac{\bar{x} - 4.0}{s_x} \;=\; \sqrt{12}\,\frac{0.3008}{0.2656} \;=\; 3.923\,.$$

Dieser Wert übersteigt (betragsmäßig) das Quantil $\;t_0 \equiv t_{n-1,1-0.05/2} = t_{11,0.975} \equiv 2.2010.$ Die Hypothese H_0 wird verworfen.

b) Mit $\bar{x}$, s_x und t_0 wie in a) wird die halbe Intervallbreite

$$c_0 \equiv t_0 \cdot \frac{s_x}{\sqrt{n}} \;=\; 0.1688\,.$$

Also erhalten wir das Konfidenzintervall für μ zum 95 % Niveau

$$4.1320 \;=\; \bar{x} - c_0 \;\leq\; \mu \;\leq\; \bar{x} + c_0 \;=\; 4.4696\,,$$

das – konform mit Teil a) – die Zahl 4.0 nicht einschließt.

$\boxed{\textbf{9.}}$

In der Situation von 10.3.4 ($X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2}$ unabhängig und $N(\mu_1, \sigma^2)$- bzw $N(\mu_2, \sigma^2)$-verteilt) lauten die zwei Stichproben, jeweils vom Umfang $n_1 = n_2 = 12$,

$$(x_i) = (4.10, 4.31, 4.05, 4.38, 4.23, 4.24, 4.63, 4.29, 4.10, 3.91, 4.51, 4.86)$$

$$(y_i) = (4.37, 4.25, 4.22, 5.30, 5.05, 4.95, 4.25, 4.68, 4.43, 4.35, 4.88, 4.49).$$

Mittelwerte, Standardabweichungen und pooled Varianz s^2

$$\bar{x} \;=\; 4.3008\,, \qquad \bar{y} \;=\; 4.6017$$

$$s_x \;=\; 0.2656\,, \qquad s_y \;=\; 0.3625$$

$$s^2 \;=\; \frac{1}{22}\left(11 \cdot 0.2656^2 + 11 \cdot 0.3625^2\right) \;=\; \frac{1}{2} \cdot 0.2019 \;=\; 0.1010\,, \quad s \;=\; 0.3178\,.$$

a) Die Teststatistik zum Prüfen von $H_0 : \mu_1 = \mu_2$ berechnet sich wie folgt:

$$t = \sqrt{\frac{12 \cdot 12}{2 \cdot 12}} \cdot \frac{4.3008 - 4.6017}{0.3178} \;=\; \sqrt{6} \cdot \frac{-0.3009}{0.3178} \;=\; -2.319\,.$$

Als zugehöriges Quantil der t-Verteilung hat man

$$t_0 \equiv t_{22,0.975} \;=\; 2.074\,.$$

Wegen $|t| = 2.319 > 2.074$ wird H_0 verworfen zugunsten von $\mu_1 \neq \mu_2$ (sogar zugunsten von $\mu_1 < \mu_2$).

b) Mit $\bar{x}$, $\bar{y}$, s und t_0 wie in Teil a) wird die halbe Breite des Intervalls

$$c_0 \equiv t_0 \cdot \frac{s}{\sqrt{6}} \;=\; 0.2691$$

und das das Konfidenzintervall für $\mu_1 - \mu_2$ zum 95 % Niveau

$$-0.570 \;=\; (\bar{x} - \bar{y}) - c_0 \;\leq\; \mu_1 - \mu_2 \;\leq\; (\bar{x} - \bar{y}) + c_0 \;=\; -0.032\,,$$

das – konform mit Teil a) – die Zahl 0 nicht einschließt.

$\boxed{10.}$ **a)**

Es ist $n = 21$ (die 3 Datensätze mit `Bu = -` wurden weggelassen).

Kontingenztafel:

	$Y=0$	$Y=1$	Summe $n_{i\bullet}$
$X=0$	3	5	8
$X=1$	8	5	13
Summe $n_{\bullet j}$	11	10	21

Tafel der „erwarteten Häufigkeiten“ e_{ij}:

	$Y=0$	$Y=1$	
$X=0$	4.19	3.81	8
$X=1$	6.81	6.19	13
	11	10	21

b) Wie in 10.4.2 bilden wir die Teststatistik

$$\hat{\chi}^2_n \;=\; \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} \;=\; n\cdot\Big(\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}}-1\Big).$$

Mit den Werten aus der Kontingenztafel in a) ergibt sich

$$\hat{\chi}^2_{21} \;=\; 21\cdot\Big[\frac{3^2}{8\cdot 11}+\frac{5^2}{8\cdot 10}+\frac{8^2}{13\cdot 11}+\frac{5^2}{13\cdot 10}-1\Big] = 21\cdot 0.0546 \;=\; 1.1473.$$

Da $\hat{\chi}^2_{21} \not> 3.842 = \chi^2_{1,1-\alpha}$ für $\alpha = 0.05$ (siehe Tabelle 10.3), kann die Hypothese der Unabhängigkeit von X und Y nicht verworfen werden.

b) Es ist

$$C = \sqrt{\frac{\hat{\chi}^2_{21}}{\hat{\chi}^2_{21}+21}} \;=\; 0.2276 \qquad \text{und} \qquad V = \sqrt{\frac{\hat{\chi}^2_{21}}{21\cdot K}} \underset{K=1}{=} 0.2337.$$

$\boxed{11.}$ **a)**

Es ist

$$\frac{\big(n_j - n\cdot p_j^{(0)}\big)^2}{n\cdot p_j^{(0)}} = \frac{n_j^2 - 2nn_jp_j^{(0)} + (np_j^{(0)})^2}{np_j^{(0)}} = \frac{n_j^2}{np_j^{(0)}} - 2n_j + np_j^{(0)},$$

woraus

$$\begin{aligned}\hat{\chi}^2_n &= \sum_{j=1}^{m}\frac{\big(n_j - n\cdot p_j^{(0)}\big)^2}{n\cdot p_j^{(0)}} \;=\; \sum_{j=1}^{m}\frac{n_j^2}{np_j^{(0)}} - 2\sum_{j=1}^{m}n_j + n\sum_{j=1}^{m}p_j^{(0)}\\ &= \sum_{j=1}^{m}\frac{n_j^2}{np_j^{(0)}} - 2n + n \;=\; \sum_{j=1}^{m}\frac{n_j^2}{np_j^{(0)}} - n\end{aligned}$$

folgt. Ferner ist nach Definition von e_{ij}

$$\begin{aligned}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} &= \sum_{i=1}^{I}\sum_{j=1}^{J}\Big(\frac{n_{ij}^2}{e_{ij}}-2n_{ij}+e_{ij}\Big)\\ &= \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}}\cdot n-2\sum_{i=1}^{I}\sum_{j=1}^{J}n_{ij}+\frac{1}{n}\sum_{i=1}^{I}n_{i\bullet}\sum_{j=1}^{J}n_{\bullet j}\\ &= n\cdot\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}}\;-\;2n\;+\;\frac{1}{n}n^2\\ &= n\cdot\Big(\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}}-1\Big).\end{aligned}$$

b) Es ist $\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{\bullet j}} \;\le\; \sum_{j=1}^{J}n_{ij} \;=\; n_{i\bullet}$, woraus

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}} \;\le\; I$$

folgt. Analog ist auch $\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}} \le J$, also

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}} \;\le\; \min\{I,J\}=K+1. \tag{+}$$

Also ist $1-\frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}}} \;\le\; 1-\frac{1}{K+1} \;=\; \frac{K}{K+1}$, woraus wegen

$$1-\frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}}} = \frac{n\Big(\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}}-1\Big)}{n\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}}} \underset{a)}{=} \frac{\hat{\chi}_n^2}{\hat{\chi}_n^2+n} \;=\; C^2$$

die Behauptung $\;0\le C\le\sqrt{K/(K+1)}$ folgt.
Die zweite Ungleichung, nämlich $\;V^2\le 1,$ d.h. $\;\hat{\chi}_n^2\le n\cdot K,\;$ folgt wegen

$$\frac{\hat{\chi}_n^2}{n} \underset{a)}{=} \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}}-1 \underset{(+)}{\le} K.$$

Sei nun o.E. $I\le J$. Dann ergibt sich aus obigem Beweis (siehe (+))

$$V=1 \quad\Longleftrightarrow\quad \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{n_{ij}^2}{n_{i\bullet}\cdot n_{\bullet j}} \;=\; I.$$

Letzteres ist wegen $\sum\limits_{j=1}^{J} \frac{n_{ij}^2}{n_{i\bullet} \cdot n_{\bullet j}} \leq 1$ genau dann der Fall, wenn

$$\sum_{j=1}^{J} \frac{n_{ij}^2}{n_{\bullet j}} = n_{i\bullet} \qquad \text{für alle } i = 1, ..., I,$$

also wenn $\frac{n_{ij}^2}{n_{\bullet j}} = n_{ij}$, $j = 1, ..., J$, $i = 1, ..., I$ (beachte, dass stets $\frac{n_{ij}^2}{n_{\bullet j}} \leq n_{ij}$). Damit muss $n_{ij} = 0$ oder $n_{ij} = n_{\bullet j}$ gelten. Da bei festem j nicht für zwei verschiedene i

$$n_{ij} = n_{\bullet j}$$

gelten kann, folgt, dass in jeder Spalte nur an einer einzigen Stelle eine von 0 verschiedene Häufigkeit stehen kann.
Analog argumentiert man im Fall $I \geq J$.