Statistik I für Human- und Sozialwissenschaften

7. Übung

Lösungsvorschlag

Gruppenübung

G 22 Bei der Größenmessung in einer Gruppe von 9 Personen ergaben sich folgende Körpergrößen [in cm]:

`184.2, 182.6, 185.3, 184.5, 186.2, 183.9, 185.0, 187.1, 184.4.`

Diese Messwerte werden als Realisationen der Zufallsvariablen $X_1, \ldots, X_9$ angenommen, die unabhängig und identisch $N(\mu, \sigma^2)$-verteilt seien.

(a) Geben Sie ein Konfidenzintervall zum Konfidenzniveau 0.99 für den Erwartungswert μ an, falls die Standardabweichung bekannt ist und $\sigma = 2.4$ [cm] beträgt.

(b) Welches Konfidenzintervall ergibt sich in (a) für dasselbe Konfidenzniveau, falls die Standardabweichung als unbekannt angenommen wird?

(c) Ermitteln Sie im letzteren Fall ein Konfidenzintervall zum Konfidenzniveau 0.9 für die Varianz σ^2.

(a) Der Mittelwert der Messwerte ist $\bar{x} = 184.8$ *und das* 0.995*-Quantil der* $N(0,1)$*-Verteilung* $z_{0.995} \approx 2.58$.
Daher lautet das gesuchte Konfidenzintervall

$$\left[\bar{x} - z_{0.995}\frac{\sigma}{\sqrt{9}}, \bar{x} + z_{0.995}\frac{\sigma}{\sqrt{9}}\right] \approx [182.736, 186.864].$$

(b) Die Standardabweichung der Messwerte ist $\bar{s} \approx 1.3134$ *und das* 0.995*-Quantil der* t_8*-Verteilung* $t_{8;0.995} \approx 3.36$.
Daher lautet das gesuchte Konfidenzintervall

$$\left[\bar{x} - t_{8;0.995}\frac{\bar{s}}{\sqrt{9}}, \bar{x} + t_{8;0.995}\frac{\bar{s}}{\sqrt{9}}\right] \approx [183.329, 186.27].$$

(c) Das 0.95*-Quantil und das* 0.05*-Quantil der* χ^2_8*-Verteilung sind* $\chi^2_{8;0.95} \approx 15.51$ *und* $\chi^2_{8;0.05} \approx 2.37$. *Daher lautet das gesuchte Konfidenzintervall*

$$\left[\frac{(n-1)\bar{s}^2}{\chi_{8;0.995}}, \frac{(n-1)\bar{s}^2}{\chi_{8;0.005}}\right] \approx [0.89, 5.005].$$

G 23 Eine Apparatur füllt eine bestimmte Menge (in g) eines pulverförmigen Medikaments ab. Es wird angenommen, dass diese Menge durch eine normalverteilte Zufallsvariable X mit unbekanntem Erwartungswert μ und unbekannter Varianz σ^2 beschrieben werden kann. Zehn Messungen ergaben folgende Werte $x_1, \ldots, x_{10}$ für die abgefüllte Menge:

18.3 20.2 20.7 19.8 19.5 20.9 18.1 20.5 20.4 17.6.

(a) Geben Sie die konkreten Intervallgrenzen des Konfidenzintervalls für μ zum Niveau $1 - \alpha = 0.9$ an.
Hinweis: $\sum_{i=1}^{10} x_i^2 = 3\,854.3$

(b) Wie verändern sich die Intervallgrenzen des Konfidenzintervalls für μ zum Niveau $1-\alpha = 0.9$, falls zusätzlich $\sigma^2 = 1$ bekannt ist?

(a) Das Konfidenzintervall lautet:

$$\left[\bar{X}_n - t_{n-1;1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{S}_X^2}{n}}\ ,\ \bar{X}_n + t_{n-1;1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{S}_X^2}{n}}\right] .$$

Zur Berechnung des konkreten Schätzintervalls müssen wir zunächst die empirische Varianz der Messreihe bestimmen:

$$\begin{aligned} \bar{x}_{10} &= \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1}{10}(18.3 + 20.2 + \ldots + 17.6) = \frac{196}{10} = 19.6, \\ \bar{s}_x^2 &= \frac{1}{9}\sum_{i=1}^{10}(x_i - \bar{x}_{10})^2 = 1.411. \end{aligned}$$

Mit diesen Werten und $t_{9,0.95} = 1.83$ *erhält man als konkretes Schätzintervall für* μ

$$\begin{aligned} &\left[\bar{x}_{10} - t_{9;1-\frac{0.1}{2}} \cdot \sqrt{\frac{1.411}{10}}\ ,\ \bar{x}_{10} + t_{9;1-\frac{0.1}{2}} \cdot \sqrt{\frac{1.411}{10}}\right] \\ &= \left[19.6 - 1.83 \cdot \sqrt{\frac{1.411}{10}}\ ,\ 19.6 + 1.83 \cdot \sqrt{\frac{1.411}{10}}\right] \\ &\approx [18.913, 20.287]\ . \end{aligned}$$

(b) Das Konfidenzintervall ergibt sich in diesem Fall, da σ^2 *als bekannt vorausgesetzt wird, durch*

$$\left[\bar{X}_n - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma^2}{n}}\right].$$

Mit $\sigma^2 = 1$ *und* $u_{0.95} = 1.65$ *ergibt sich als konkretes Schätzintervall für* μ

$$\begin{aligned} &\left[\bar{x}_{10} - z_{1-\frac{0.1}{2}} \cdot \sqrt{\frac{\sigma^2}{10}}, \bar{x}_{10} + z_{1-\frac{0.1}{2}} \cdot \sqrt{\frac{\sigma^2}{10}}\right] \\ &= \left[19.6 - 1.65 \cdot \frac{1}{\sqrt{10}}\ ,\ 19.6 + 1.65 \cdot \frac{1}{\sqrt{10}}\right] \\ &\approx [19.078, 20.122]\ . \end{aligned}$$

G 24 Die Zufallsvariablen $X_1, ..., X_n$ seien unabhängig und identisch verteilt mit der Verteilungsfunktion

$$F(x) = \begin{cases} 0 & -\infty < x < \theta - 1 \\ \frac{1}{2}x - \frac{1}{2}\theta + \frac{1}{2} & \theta - 1 \leq x \leq \theta + 1 \\ 1 & x > \theta + 1 \end{cases}$$

wobei $\theta \in \mathbb{R}$ unbekannt ist.
Zeigen Sie, dass das arithmetische Mittel $T_n(X_1, \ldots, X_n) = \bar{X}$ ein erwartungstreuer Schätzer für θ ist.
Hinweis: Bestimmen Sie als erstes den Erwartungswert der Zufallsvariablen.

Sei X eine Zufallsvariable mit obiger Verteilungsfunktion, dann gilt (Dichte berechnen!)

$$EX = \int_{\theta-1}^{\theta+1} \tfrac{1}{2} x dx = \tfrac{1}{4} x^2 \Big|_{\theta-1}^{\theta+1} = \theta$$

Der Schätzer T_n ist erwartungstreu, falls $\forall \theta \in \mathbb{R} : E_\theta(T_n(X_1, ..., X_n)) = \theta$.
Es gilt

$$E_\theta(\bar{X}) = E_\theta \left(\frac{1}{n}(X_1 + \ldots + X_n) \right) = \frac{1}{n} \cdot n \cdot E_\theta(X_1) = \theta.$$

Somit ist der betrachtete Schätzer erwartungstreu.
Oder: Wie in der Vorlesung gezeigt wurde, ist das arithmetische Mittel ein erwartungstreuer Schätzer für den Erwartungswert.

G 25 (a) Bestimmen Sie die folgenden Quantile:

$$\chi_{6;0.95},\ \chi_{14;0.01},\ z_{0.975},\ z_{0.8}, t_{10;0.05},\ t_{3;0.975}.$$

(b) Gegeben sei die folgende Stichprobe

0.085, 1.08, 0.35, 3.28, 1.24, 2.58, 0.02, 0.13, 0.22, 0.52.

Bestimmen Sie den Median und das harmonische Mittel dieser Stichprobe.

(a) $\chi_{6;0.95} = 12.59$, $\chi_{14;0.01} = 4.66$, $z_{0.975} = 1.960$, $z_{0.8} = 0.84$, $t_{10;0.05} = -1.81$, $t_{3;0.975} = 3.18$
(b) $Med = 0.5 \cdot (0.52 + 0.85) = 0.685$, $\bar{x}_{HM} = 0.014$

Hausübung

H 13 In einer Stadt liegen für 121 Jahre die Niederschlagsmengen im Monat April vor. Die Messreihe $x_1, \ldots, x_{121}$ (x_i = Niederschlagshöhe in mm im i-ten Jahr) hat das arithmetische Mittel $\bar{x}_{121} = 53.68$ und die empirische Standardabweichung $\bar{s}_x = 6.13$. Es wird angenommen, dass die Werte $x_1, \ldots, x_{121}$ eine Realisierung von 121 unabhängigen, identisch $\mathrm{N}(\mu, \sigma^2)$–verteilten Zufallsvariablen sind. Mit Konfidenzschätzverfahren zum Niveau $1-\alpha = 0.98$ bestimme man je ein konkretes Schätzintervall

(a) für μ,
(b) für σ^2,
(c) für μ unter der Voraussetzung $\sigma^2 = 6.13^2$,

Hinweis: $\chi^2_{120;0.99} = 158.962$ und $\chi^2_{120;0.01} = 86.909$

(a) Das Konfidenzintervall lautet:

$$\left[\bar{X}_n - t_{n-1;1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{S}_X^2}{n}} , \bar{X}_n + t_{n-1;1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\bar{S}_X^2}{n}} \right].$$

Mit $t_{120,0.99} = 2.36$ erhält man das konkrete Schätzintervall

$$\begin{aligned} & \left[\bar{x}_{121} - t_{120;1-\frac{0.02}{2}} \cdot \frac{\bar{s}_x}{\sqrt{121}} , \bar{x}_{121} + t_{120;1-\frac{0.02}{2}} \cdot \frac{\bar{s}_x}{\sqrt{121}} \right] \\ = & \left[53.68 - 2.36 \cdot \frac{6.13}{\sqrt{120}} , 53.68 + 2.36 \cdot \frac{6.13}{\sqrt{120}} \right] \\ \approx & [52.369, 54.99] \end{aligned}$$

für μ. Man beachte, dass in der Aufgabenstellung die empirische Streuung s_x und nicht etwa die empirische Varianz s_x^2 angegeben ist.
(b) Das Konfidenzintervall ergibt sich durch

$$\left[\frac{(n-1)\bar{S}_X^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)\bar{S}_X^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right].$$

Mit den im Hinweis angegebenen Quantilen erhält man für σ^2 das konkrete Schätzintervall

$$\begin{aligned}&\left[\frac{120\bar{s}_x^2}{\chi^2_{120;1-\frac{0.02}{2}}}, \frac{120\bar{s}_x^2}{\chi^2_{120;\frac{0.02}{2}}}\right]\\ = &\left[\frac{120\cdot 6.13^2}{158.962}, \frac{120\cdot 6.13^2}{86.909}\right]\\ \approx &\ [28.336, 51.88]\ .\end{aligned}$$

(c) Das Konfidenzintervall ergibt sich in diesem Fall, da σ^2 als bekannt vorausgesetzt wird, durch

$$\left[\bar{X}_n - z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\sigma^2}{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}}\cdot\sqrt{\frac{\sigma^2}{n}}\right].$$

Mit $\sigma^2 = 6.13^2$ und $u_{0.99} = 2.33$ ergibt sich das konkrete Schätzintervall

$$\begin{aligned}&\left[\bar{x}_{121} - z_{1-\frac{0.02}{2}}\cdot\sqrt{\frac{\sigma^2}{121}}, \bar{x}_{121} + z_{1-\frac{0.02}{2}}\cdot\sqrt{\frac{\sigma^2}{121}}\right]\\ = &\left[53.68 - 2.33\cdot\frac{6.13}{11}, 53.68 + 2.33\cdot\frac{6.13}{11}\right]\\ \approx &\ [52.38, 54.98]\ .\end{aligned}$$

Verglichen mit dem Ergebnis in a) wird die Länge des Konfidenzintervalls also etwas kleiner; dafür lagen hier aber auch mehr Informationen vor.

H 14 Um die Präzision einer Waage zu überprüfen, wird n-mal das Gewicht eines Kilogramm-Prototyps gemessen. Die entstehende Messreihe soll als Realisierung von unabhängigen, identisch $N(1,\theta)$-verteilten Zufallsvariablen $X_1,\ldots,X_n$ mit unbekannter Varianz $\theta > 0$ aufgefasst werden.
Bestimmen Sie den Maximum-Likelihood-Schätzer T_n für $\tau(\theta) = \theta$.

Für $x_1,\ldots,x_n \in \mathbb{R}$ ist die Likelihoodfunktion gegeben durch

$$\begin{aligned}L(\theta;x_1,\ldots,x_n) &= \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}}\exp(-(x_i-1)^2/(2\theta))\\ &= (2\pi\theta)^{-n/2}\cdot\exp\left(-\frac{1}{2\theta}\sum_{i=1}^n (x_i-1)^2\right).\end{aligned}$$

Logarithmieren ergibt

$$g(\theta) := \ln L(\theta;x_1,\ldots,x_n) = -\frac{n}{2}\ln(2\pi\theta) - \frac{1}{2\theta}\sum_{i=1}^n (x_i-1)^2.$$

Diese Funktion gilt es nun zu maximieren, also:

$$
\begin{aligned}
& \frac{dg}{d\theta} = 0 \\
\Rightarrow \quad & -\frac{n}{2} \cdot \frac{1}{\theta} + \frac{1}{2\theta^2} \sum_{i=1}^{n} (x_i - 1)^2 = 0 \\
\Rightarrow \quad & \frac{n}{2\theta} = \frac{1}{2\theta^2} \sum_{i=1}^{n} (x_i - 1)^2 \\
\Rightarrow \quad & \theta = \frac{1}{n} \sum_{i=1}^{n} (x_i - 1)^2 =: \hat{\theta}.
\end{aligned}
$$

Wegen

$$
\begin{aligned}
\left.\frac{d^2 g}{d\theta^2}\right|_{\theta=\hat{\theta}} &= \frac{n}{2\hat{\theta}^2} - \frac{1}{\hat{\theta}^3} \sum_{i=1}^{n} (x_i - 1)^2 \\
&= \frac{n}{2\hat{\theta}^2} - \frac{1}{\hat{\theta}^3} \cdot n\hat{\theta} = -\frac{n}{2\hat{\theta}^2} < 0
\end{aligned}
$$

liegt an der Stelle $\hat{\theta}$ tatsächlich ein Maximum vor, d.h.

$$
T_n(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} (x_i - 1)^2
$$

ist der Maximum-Likelihood-Schätzer für θ.