Serie 07: Schliessende Statistik - Vertrauensbereiche

Punktschätzung

Aufgabe 1

Wir betrachten eine Grundgesamtheit mit Erwartungswert μ und Varianz σ^2. Es sei X_1, X_2, X_3 eine einfache Zufallsstichprobe aus dieser Grundgesamtheit. Folgende drei Schätzfunktionen sind gegeben:

$$\Theta_1 = \frac{1}{3}(X_1 + X_2 + X_3),\ \Theta_2 = \frac{1}{4}(2X_1 + 2X_3),\ \Theta_3 = \frac{1}{3}(2X_1 + X_2)$$

a) Welche dieser Schätzfunktionen sind erwartungstreu?

b) Welche dieser Schätzfunktionen ist am effizientesten, welche am wenigsten effizient?

Hinweis: Verwenden Sie die Eigenschaften von Erwartungswert und Varianz aus 4.3 sowie den Satz zur Varianz einer Summe von stochastisch unabhängigen Zufallsvariablen aus 4.5.

Lösung:

a) $E(\Theta_1) = E\left(\frac{1}{3} \cdot (X_1 + X_2 + X_3)\right) = \frac{1}{3} \cdot \big(E(X_1) + E(X_2) + E(X_3)\big) = \frac{1}{3} \cdot (\mu + \mu + \mu) = \mu$

$E(\Theta_2) = E\left(\frac{1}{4} \cdot (2X_1 + 2X_3)\right) = \frac{1}{4} \cdot \big(2E(X_1) + 2E(X_3)\big) = \frac{1}{4} \cdot (2\mu + 2\mu) = \mu$

$E(\Theta_3) = E\left(\frac{1}{3} \cdot (2X_1 + X_2)\right) = \frac{1}{3} \cdot (2E(X_1) + E(X_2)) = \frac{1}{3} \cdot (2\mu + \mu) = \mu$

Alle drei Schätzer sind erwartungstreu.

b) $V(\Theta_1) = V\left(\frac{1}{3} \cdot (X_1 + X_2 + X_3)\right) = \frac{1}{9} \cdot V(X_1 + X_2 + X_3) = \frac{1}{9} \cdot \big(V(X_1) + V(X_2) + V(X_3)\big)$

$= \frac{1}{9} \cdot (\sigma^2 + \sigma^2 + \sigma^2) = \frac{\sigma^2}{3}$

$V(\Theta_2) = V\left(\frac{1}{4} \cdot (2X_1 + 2X_3)\right) = V\left(\frac{1}{2} \cdot (X_1 + X_3)\right) = \frac{1}{4} \cdot V(X_1 + X_3) = \frac{1}{4} \cdot \big(V(X_1) + V(X_3)\big)$

$= \frac{1}{4} \cdot (\sigma^2 + \sigma^2) = \frac{\sigma^2}{2}$

$V(\Theta_3) = V\left(\frac{1}{3} \cdot (2X_1 + X_2)\right) = \frac{1}{9} \cdot V(2X_1 + X_2) = \frac{1}{9} \cdot \big(V(2X_1) + V(X_2)\big)$

$= \frac{1}{9} \cdot \big(4 \cdot V(X_1) + V(X_2)\big) = \frac{1}{9} \cdot (4\sigma^2 + \sigma^2) = \frac{5\sigma^2}{9}$

Somit ist Θ_1 am effizientesten, Θ_3 am wenigsten effizient.

Aufgabe 2

Wir betrachten eine Grundgesamtheit, die nach der Verteilungsdichte (PDF) $f(x) = \lambda \cdot e^{-\lambda x},\ x \geq 0$ verteilt ist. Es wird daraus eine Zufallsstichprobe $X_1, X_2, \ldots, X_n$ entnommen, um den unbekannten Parameter $\lambda > 0$ zu schätzen. Definieren Sie die Likelihood-Funktion für dieses Problem und bestimmen Sie durch Differenzieren eine Schätzfunktion für den unbekannten Parameter $\lambda > 0$.

Lösung:

Likelihood-Funktion:

$L(\lambda) = \lambda \cdot e^{-\lambda x_1} \cdot \ldots \cdot \lambda \cdot e^{-\lambda x_n} = \lambda^n \cdot e^{-\lambda(x_1+\ldots+x_n)}$, $\lambda > 0$.

$$\frac{d}{d\lambda} L(\lambda) = e^{-\lambda(x_1+\ldots+x_n)} \cdot (n \cdot \lambda^{n-1} - \lambda^n \cdot (x_1 + \ldots + x_n)) = 0 \quad \Leftrightarrow \quad \lambda = \frac{n}{x_1 + \ldots + x_n}$$

ML-Schätzwert für den Parameter $\lambda > 0$: $\hat{\lambda}_{ML} = \frac{n}{x_1+\ldots+x_n}$

Aufgabe 3

Im Vorfeld einer Abstimmung soll mithilfe einer 0,1-wertigen Zufallsstichprobe $X_1, \ldots, X_n$ der unbekannte Anteil p an Ja-Stimmen geschätzt werden (1 für ja, 0 für nein). Definieren Sie eine Likelihood-Funktion für dieses Problem und bestimmen Sie durch Differenzieren eine Schätzfunktion für den unbekannten Parameter $0 < p < 1$.

Lösung:

Likelihood Funktion:

$L(p) = p^{x_1+\ldots+x_n} \cdot (1-p)^{n-x_1-\ldots-x_n}$

$$\frac{d}{dp} L(p) = (x_1 + \ldots + x_n) p^{x_1+\ldots+x_n-1} \cdot (1-p)^{n-x_1-\ldots-x_n} - (n - x_1 - \ldots - x_n) p^{x_1+\ldots+x_n} \cdot (1-p)^{n-x_1-\ldots-x_n-1}$$

$$= p^{x_1+\ldots+x_n-1} \cdot (1-p)^{n-x_1-\ldots-x_n-1} [(x_1 + \ldots + x_n) \cdot (1-p) - (n - x_1 - \ldots - x_n) p] = 0$$

$$\Leftrightarrow (x_1 + \ldots + x_n) \cdot (1-p) - (n - x_1 - \ldots - x_n) \cdot p = 0$$

$$\Leftrightarrow (x_1 + \ldots + x_n) - n \cdot p = 0 \Leftrightarrow p = \frac{x_1 + \ldots + x_n}{n}$$

ML-Schätzwert für den Parameter p : $\hat{p}_{ML} = \frac{x_1+\ldots+x_n}{n}$

Intervallschätzung

Aufgabe 4

Wir nehmen an, dass der Durchmesser X der auf einer Maschine hergestellten Schrauben eine normalverteilte Zufallsvariable ist. Eine Stichprobe vom Umfang $n = 100$, entnommen aus einer Tagesproduktion, ergab das folgende Ergebnis: $\bar{x} = 0.620$ cm, $s = 0.035$ cm. Bestimmen Sie die Vertrauensgrenzen für den unbekannten Mittelwert μ bei einer Irrtumswahrscheinlichkeit $\alpha = 5\%$.

Lösung:

(1) Zeile 1 aus Tabelle (da $n > 30$)

(2) $\bar{x} = 0.620$ cm, $s = 0.035$ cm (aus Aufgabenstellung)

(3) $\gamma = 1 - \alpha = 0.95$

Die standardisierte Zufallsvariable ist standardnormalverteilt (Spalte 5).

$p = \frac{1+\gamma}{2} = 0.975$, $c = u_p = 1.96$

(4) $e = c \cdot \frac{s}{\sqrt{n}} = 1.96 \cdot \frac{0.035}{\sqrt{100}} = 0.0069$cm

95%-Vertrauensintervall für μ: $[\bar{x} - e; \bar{x} + e] = [0.613; 0.627]$

Aufgabe 5

Gegeben ist eine normalverteilte Zufallsvariable X mit dem unbekannten Mittelwert μ und der ebenfalls unbekannten Varianz σ^2. Eine Stichprobe vom Umfang $n = 10$ ergab den arithmetischen Mittelwert $\bar{x} = 102$ und die empirische Varianz $s^2 = 16$. Bestimmen Sie für μ und σ^2 jeweils ein Vertrauensintervall zum Vertrauensniveau von $\gamma = 99\%$.

Lösung:

Rechnung für μ

(1) Zeile 2 aus Tabelle

(2) $\bar{x} = 102$, $s^2 = 16$ (aus Aufgabenstellung)

(3) $\gamma = 0.99$

Die standardisierte Zufallsvariable ist *t*-verteilt mit $f = n - 1 = 9$ (Spalte 5).

$p = \frac{1+\gamma}{2} = 0.995$, $c = t_{(p;f)} = 3.250$

(4) $e = c \cdot \frac{s}{\sqrt{n}} = 3.250 \cdot \frac{4}{\sqrt{10}} = 4.111$

99%-Vertrauensintervall für μ: $[\bar{x} - e; \bar{x} + e] = [97.89; 106.11]$

Rechnung für σ^2

(1) Zeile 3 aus Tabelle

(2) $\bar{x} = 102$, $s^2 = 16$ (aus Aufgabenstellung)

(3) $\gamma = 0.99$

Die standardisierte Zufallsvariable ist Chi-Quadrat-verteilt mit $f = n - 1 = 9$ (Spalte 5).

$p_1 = \frac{1-\gamma}{2} = 0.005$, $c_1 = z_{(p_1;f)} = 1.73$, $p_2 = \frac{1+\gamma}{2} = 0.995$, $c_2 = z_{(p_2;f)} = 23.59$

(4) $\theta_u = \frac{(n-1)\cdot s^2}{c_2} = \frac{9\cdot 16}{23.59} = 6.10, \quad \theta_o = \frac{(n-1)\cdot s^2}{c_1} = \frac{9\cdot 16}{1.73} = 83.24$

99%-Vertrauensintervall für σ^2: $[\theta_u; \theta_o] = [6.10; 83.24]$

Aufgabe 6

Für einen Autotyp wurde ein bestimmter Motor weiterentwickelt, dessen Leistung als eine normalverteilte Zufallsvariable betrachtet werden kann. Eine Stichprobenuntersuchung an $n = 8$ zufällig herausgegriffenen Motoren ergab das folgende Ergebnis:

i	1	2	3	4	5	6	7	8
x_i in PS	100.5	96.5	99.0	97.8	100.4	103.5	100.3	98.0

Bestimmen Sie für μ und σ^2 jeweils ein Vertrauensintervall zu einer Irrtumswahrscheinlichkeit von $\alpha = 5\%$.

<u>Lösung:</u>

Rechnung für μ

(1) Zeile 2 aus Tabelle

(2) $\bar{x} = 99.5$ Ps, $s^2 = 4.69$ Ps^2 (aus Tabelle bestimmen)

(3) $\gamma = 1 - \alpha = 0.95$

Die standardisierte Zufallsvariable ist *t*-verteilt mit $f = n - 1 = 7$ (Spalte 5).

$p = \frac{1+\gamma}{2} = 0.975$, $c = t_{(p;f)} = 2.365$

(4) $e = c \cdot \frac{s}{\sqrt{n}} = 2.365 \cdot \sqrt{\frac{4.69}{8}} = 1.81$

95%-Vertrauensintervall für μ: $[\bar{x} - e; \bar{x} + e] = [97.69;101.31]$

Rechnung für σ^2

(1) Zeile 3 aus Tabelle

(2) $\bar{x} = 99.5$ Ps, $s^2 = 4.69$ Ps^2 (s.o.)

(3) $\gamma = 0.95$

Die standardisierte Zufallsvariable ist Chi-Quadrat-verteilt mit $f = n - 1 = 7$ (Spalte 5).

$p_1 = \frac{1-\gamma}{2} = 0.025$, $c_1 = t_{(p_1;f)} = 1.69$, $p_2 = \frac{1+\gamma}{2} = 0.975$, $c_2 = t_{(p_2;f)} = 16.01$

(4) $\theta_u = \frac{(n-1)\cdot s^2}{c_2} = \frac{7\cdot 4.69}{16.01} = 2.05, \quad \theta_o = \frac{(n-1)\cdot s^2}{c_1} = \frac{7\cdot 4.69}{1.69} = 19.43$

95%-Vertrauensintervall für σ^2: $[\theta_u; \theta_o] = [2.05; 19.43]$

Aufgabe 7

Bei einer Qualitätskontrolle eines elektronischen Bauteils befanden sich 27 defekte Teile in einer Stichprobe vom Umfang $n = 500$. Bestimmen Sie den Schätzwert für den unbekannten Ausschussanteil p der Gesamtproduktion und ein Vertrauensintervall für diesen Parameter zum Vertrauensniveau

(a) $\gamma = 95\%$ und (b) $\gamma = 99\%$.

Lösung:

(1) Zeile 4 aus Tabelle

(2) $\hat{p} = \bar{x} = \frac{27}{500} = 0.054$

(3) (a) $\gamma = 0.95$ (b) $\gamma = 0.99$

Die standardisierte Zufallsvariable ist näherungsweise standardnormalverteilt (Spalte 5).

(a) $p = \frac{1+\gamma}{2} = 0.975, c = u_p = 1.96$ (b) $p = \frac{1+\gamma}{2} = 0.995, c = u_p = 2.576$

(4) (a) $e = c \cdot \sqrt{\frac{\bar{x}\cdot(1-\bar{x})}{n}} = 1.96 \cdot \sqrt{\frac{0.054\cdot 0.946}{500}} = 0.0198$

95%-Vertrauensintervall für p: $[\bar{x} - e; \bar{x} + e] = [0.034;0.074]$

(b) $e = c \cdot \sqrt{\frac{\bar{x}\cdot(1-\bar{x})}{n}} = 2.576 \cdot \sqrt{\frac{0.054\cdot 0.946}{500}} = 0.026$

99%-Vertrauensintervall für p: $[\bar{x} - e; \bar{x} + e] = [0.028;0.080]$

Aufgabe 8

Aus einer Sonderprägung wurden $n = 100$ Münzen nach dem Zufallsprinzip ausgewählt und ihre Masse m bestimmt. Man erhielt dabei den Stichprobenmittelwert $\bar{x} = 5.43$ g mit der Streuung $s^2 = 0.09$ g^2. Der Verteilungstyp der Zufallsvariablen ist jedoch unbekannt. Bestimmen Sie mit Hilfe einer Normalapproximation die Vertrauensintervalle für den unbekannten Mittelwert μ und die unbekannte Standardabweichung σ auf einem Vertrauensniveau von $\gamma = 95\%$.

Lösung:

Rechnung für μ

(1) Zeile 1 aus Tabelle (Approximation gemäss Zeile 5)

(2) $\bar{x} = 5.43$ g, $s^2 = 0.09$ g^2 (aus Aufgabenstellung)

(3) $\gamma = 0.95$

Die standardisierte Zufallsvariable ist näherungsweise standardnormalverteilt (Spalte 5).

$p = \frac{1+\gamma}{2} = 0.975, c = u_p = 1.96$

(4) $e = c \cdot \frac{s}{\sqrt{n}} = 1.96 \cdot \sqrt{\frac{0.09}{100}} = 0.0588$

95%-Vertrauensintervall für μ: $[\bar{x} - e; \bar{x} + e] = [5.37;5.49]$

Rechnung für σ^2

(1) Zeile 3 aus Tabelle (Approximation gemäss Zeile 5)

(2) $\bar{x} = 5.43$ g, $s^2 = 0.09$ g^2 (aus Aufgabenstellung)

(3) $\gamma = 0.95$

Die standardisierte Zufallsvariable ist Chi-Quadrat-verteilt mit $f = n - 1 = 99$ (Spalte 5).

$p_1 = \frac{1-\gamma}{2} = 0.025,$ $c_1 = t_{(p_1;f)} = 65.6 + 0.9 \cdot (74.2 - 65.6) = 73.34,$

$p_2 = \frac{1+\gamma}{2} = 0.975,$ $c_2 = t_{(p_2;f)} = 118.1 + 0.9 \cdot (129.6 - 118.1) = 128.45$

(interpoliert zwischen $f = 90$ und $f = 100$ oder genauer mit Python

scipy.stats.chi2.ppf(0.025, 99, loc=0, scale=1) und scipy.stats.chi2.ppf(0.975, 99, loc=0, scale=1)

(4) $\theta_u = \frac{(n-1)\cdot s^2}{c_2} = \frac{99\cdot 0.09}{128.45} = 0.069, \quad \theta_o = \frac{(n-1)\cdot s^2}{c_1} = \frac{99\cdot 0.09}{73.34} = 0.121$

95%-Vertrauensintervall für σ^2: $[\theta_u; \theta_o] = [0.069; 0.121]$

95%-Vertrauensintervall für σ : $[\sqrt{\theta_u}; \sqrt{\theta_o}] = [0.263; 0.349]$

Aufgabe 9

Ein Drehautomat fertigt Bolzen. Es ist bekannt, dass der Durchmesser der von dem Automaten gefertigten Bolzen normalverteilt ist mit Standardabweichung $\sigma = 0.5$ mm. Wie gross muss die Stichprobe mindestens sein, damit die Länge des 99%-Vertrauensintervalls für μ höchstens 0.4 mm beträgt?

Lösung:

(1) Zeile 1 aus Tabelle

(2) $\sigma = 0.5$ mm (aus Aufgabenstellung)

(3) $\gamma = 0.99$

Die standardisierte Zufallsvariable ist standardnormalverteilt (Spalte 5).

$p = \frac{1+\gamma}{2} = 0.995$, $c = u_p = 2.576$

(4) $e = c \cdot \frac{s}{\sqrt{n}} = 2.576 \cdot \frac{0.5}{\sqrt{n}} \leq \frac{0.4}{2} \quad \Leftrightarrow \quad \sqrt{n} \geq 2.576 \cdot \frac{0.5}{0.2} \quad \Rightarrow \quad n \geq 41.4736$

Damit das 99%-Vertrauensintervall höchstens 0.4 mm lang ist, muss die Stichprobengrösse n mindestens 42 betragen.

Aufgabe 10

Bei einer Stichprobe von $n = 5$ Robotern wird die maximale Dauergreifkraft eines Greifers gemessen. Es ergeben sich die folgenden Werte (in N): 200, 199, 198, 200, 198. Wir gehen davon aus, dass die Werte normalverteilt sind. Berechnen Sie ein 99%-Vertrauensintervall für den Mittelwert μ der maximalen Dauergreifkraft in der gesamten Produktion.

Lösung:

(1) Zeile 2 aus Tabelle

(2) $\bar{x} = 199$ N, $s^2 = 1$ N^2 (aus Tabelle bestimmen)

(3) $\gamma = 0.99$

Die standardisierte Zufallsvariable ist t-verteilt mit $f = n - 1 = 4$ (Spalte 5).

$p = \frac{1+\gamma}{2} = 0.995$, $c = t_{(p;f)} = 4.604$

(4) $e = c \cdot \frac{s}{\sqrt{n}} = 4.604 \cdot \sqrt{\frac{1}{5}} = 2.059$

99%-Vertrauensintervall für μ: $[\bar{x} - e; \bar{x} + e] = [196.94;201.06]$