

Mathematik 3: Übungsblatt - Statistik 3 (Schätzer)

1. Aufgabe:

Einer normalverteilten Grundgesamtheit wird eine Stichprobe vom Umfang $n = 50$ entnommen, und es werden die Stichprobenparameter ermittelt. Man erhält das Stichprobenmittel $\bar{x} = 15.57$ und die Stichprobenvarianz $s^2 = 0.09$.

- a) Geben Sie für den Mittelwert μ und die Varianz σ^2 der Grundgesamtheit Konfidenzintervalle für das Konfidenzniveau (Vertrauensniveau) $\gamma = 0.95$ an.
- b) Stellen Sie fest, ob bei einer Irrtumswahrscheinlichkeit $\alpha = 0.05$ signifikante Abweichungen vom Sollwert $\mu_0 = 15.50$ vorliegt
- c) Stellen Sie fest, wie sich eine Verringerung der statistischen Sicherheit auf die Konfidenzintervalle auswirkt!

Lösung:

- a) Es ergibt sich der Test

$$15.57 - t_{\frac{1+0.95}{2};49} \cdot \frac{0.3}{\sqrt{50}} < \mu < 15.57 + t_{\frac{1+0.95}{2};49} \cdot \frac{0.3}{\sqrt{50}}$$

bzw.

$$15.48 < \mu < 15.66 .$$

- b) Zu rechnen ist hier nichts, da $\alpha = 1 - \gamma$. Das erhaltene Vertrauensintervall schließt den Sollwert 15.50 ein, d.h., es liegt keine signifikante Abweichung des Stichprobenmittels von diesem Wert vor.

- c) Man kann das ausrechnen oder einfach ausprobieren. Ergebnis: Mit fallendem Konfidenzniveau γ fallen auch die Quantilwerte und damit wird das Vertrauensintervall kleiner. Das bedeutet auch, dass Tests weniger oft abgelehnt werden. Darin manifestiert sich das geringere Vertrauen praktisch.

2. Aufgabe:

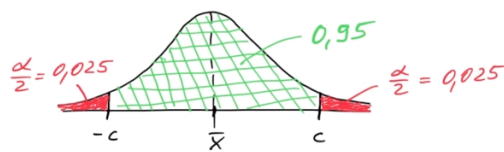
Eine Schokoladenfabrik interessiert sich für das 95 %-Konfidenzintervall des Erwartungswertes μ für das Gewicht einer Tafel Schokolade. Dafür wurden 30 Tafeln gewogen und es ergab sich ein Mittelwert von 101.08 Gramm. Die Varianz für die gesamte Produktion ist bekannt und liegt bei 4 g^2 .

a) Berechnen Sie das Konfidenzintervall.

b) Wie lautet das Konfidenzintervall, wenn die Varianz nicht bekannt ist? Verwenden Sie als Schätzer für die unbekannte Varianz $S^2 = 4 \text{ g}^2$

Lösung:

a)



Lösung: $n = 30$ $\bar{x} = 101,08 \text{ g}$ $\sigma^2 = 4 \Rightarrow \sigma = 2$
Vertrauensniveau $\gamma = 0,95$
Irrtumswahrscheinlichkeit $\alpha = 0,05$

Formel:
$$\bar{x} - c \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + c \cdot \frac{\sigma}{\sqrt{n}}$$

Außer c sind alle Werte bekannt, also

c ermitteln: Es gilt $P(-c \leq U \leq c) = \Phi(c) - \Phi(-c)$
 $= 2 \cdot \Phi(c) - 1$

$2 \cdot \Phi(c) - 1 \stackrel{!}{=} 0,95 \Leftrightarrow \Phi(c) = \frac{0,95 + 1}{2} = 0,975 \Rightarrow c = \underline{\underline{1,96}}$

Ablese in Tabelle Standard-normalverteil.

Damit ergibt sich das Konfidenzintervall (Werte in Formel einsetzen):

$$101,08 - 1,96 \cdot \frac{2}{\sqrt{30}} \leq \mu \leq 101,08 + 1,96 \cdot \frac{2}{\sqrt{30}}$$

$$\boxed{100,364 \text{ g} \leq \mu \leq 101,796 \text{ g}}$$

b)

hier $c = t_{0,975, 29}$

$\leftarrow n-1$ Freiheitsgrade

Ablese in der t -Verteilungstabelle bei $\alpha = 0,025$ und $q = 29$ (Freiheitsgrade): $c = \underline{\underline{2,045}}$

Einsetzen in Formel (gleiche Formel wie bei bekannter Varianz, wobei σ durch s ersetzt wird):

$$101,08 - 2,045 \cdot \frac{2}{\sqrt{30}} \leq \mu \leq 101,08 + 2,045 \cdot \frac{2}{\sqrt{30}}$$

$$\boxed{100,333 \text{ g} \leq \mu \leq 101,827 \text{ g}}$$

Fazit: bei unbekannter Varianz ist das Intervall etwas größer, da man etwas mehr Unsicherheit hat. Bei Stichproben mit $n > 30$ wird diese Unsicherheit vernachlässigbar und man kann die Standardnormalverteilungstabelle verwenden

3. Aufgabe:

An drei Ladesäulen eines Parkhauses wird pro Stunde jeweils die Anzahl der ladenden Elektrofahrzeuge gezählt. Die Anzahl X der Fahrzeuge an einer Ladesäule pro Stunde sei Poisson-verteilt mit dem unbekannten Parameter λ , d.h. $X \sim P_0(\lambda)$, und die drei Ladesäulen seien unabhängig voneinander. In einer bestimmten Stunde wurden beobachtet:

Ladesäule	1	2	3
Anzahl Fahrzeuge x_i	5	8	6

- Stellen Sie die Likelihood-Funktion $L(\lambda)$ für den unbekannten Parameter λ auf. Verwenden Sie dabei die Wahrscheinlichkeitsfunktion der Poisson-Verteilung $P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$.
- Stellen Sie die zugehörige **Log-Likelihood-Funktion** $\ell(\lambda) = \ln L(\lambda)$ auf und vereinfachen Sie diese so weit wie möglich.
- Bestimmen Sie durch Ableiten von $\ell(\lambda)$ nach λ und Nullsetzen den Maximum-Likelihood-Schätzwert $\hat{\lambda}$.
- Vergleichen Sie kurz den Rechenaufwand mit einer direkten Ableitung von $L(\lambda)$ (ohne Logarithmus). Was ist der entscheidende Vorteil der Log-Likelihood-Methode hier?

Lösung:

- a) Wegen der Unabhängigkeit der drei Ladesäulen ist die Likelihood-Funktion das Produkt der drei Einzelwahrscheinlichkeiten:

$$L(\lambda) = \prod_{i=1}^3 \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{x_1+x_2+x_3}}{x_1! x_2! x_3!} e^{-3\lambda} = \frac{\lambda^{19}}{5! 8! 6!} e^{-3\lambda}$$

(mit $x_1 + x_2 + x_3 = 5 + 8 + 6 = 19$).

- b) Durch Logarithmieren wird aus dem Produkt eine Summe, aus der Potenz ein Faktor:

$$\ell(\lambda) = \ln L(\lambda) = \ln \left(\frac{\lambda^{19}}{5! 8! 6!} \right) - 3\lambda = 19 \ln \lambda - \ln(5! 8! 6!) - 3\lambda$$

Allgemein für n Ladesäulen mit Beobachtungen $x_1, \dots, x_n$:

$$\ell(\lambda) = \left(\sum_{i=1}^n x_i \right) \ln \lambda - \ln \left(\prod_{i=1}^n x_i! \right) - n\lambda$$

- c) Ableiten nach λ (der konstante Term $\ln(5! 8! 6!)$ fällt weg) und Nullsetzen:

$$\ell'(\lambda) = \frac{19}{\lambda} - 3 \stackrel{!}{=} 0 \implies \hat{\lambda} = \frac{19}{3} \approx 6.33$$

Allgemein ergibt sich also $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$, d.h. der Stichprobenmittelwert ist der ML-Schätzwert für λ (im Beispiel: pro Ladesäule und Stunde werden im Mittel ≈ 6.33 ladende Fahrzeuge erwartet).

- d) Bei direkter Ableitung von $L(\lambda)$ müsste die Produktregel auf ein Produkt von Potenz- und Exponentialfunktion angewendet werden, was schnell unübersichtlich wird (vgl. Beispiel im Skript mit zwei Telefonisten). Da der Logarithmus streng monoton wachsend ist, liegt das Maximum von $\ell(\lambda)$ an derselben Stelle wie das von $L(\lambda)$ – man erhält also denselben Schätzwert $\hat{\lambda}$, muss dafür aber nur eine **Summe** statt eines **Produkts** ableiten, was insbesondere bei größerem n deutlich weniger fehleranfällig ist.

4. Aufgabe:

Die Lebensdauer T eines bestimmten Sensortyps sei exponentialverteilt mit dem unbekannten Parameter λ und der Dichte $f(t; \lambda) = \lambda e^{-\lambda t}$ für $t \geq 0$. Aus einer Testreihe liegen $n = 6$ unabhängige Realisierungen $t_1, \dots, t_6$ vor.

- Stellen Sie die Likelihood-Funktion $L(\lambda)$ für die Stichprobe $t_1, \dots, t_n$ allgemein auf.
- Leiten Sie daraus die Log-Likelihood-Funktion $\ell(\lambda)$ her.
- Bestimmen Sie allgemein (in Abhängigkeit von n und $\sum t_i$) den Maximum-Likelihood-Schätzwert $\hat{\lambda}$ mit Hilfe von $\ell(\lambda)$.
- Für die konkrete Stichprobe wurden folgende Lebensdauern (in Stunden) gemessen:

i	1	2	3	4	5	6
t_i	820	910	1005	760	1150	880

Berechnen Sie damit den konkreten Schätzwert $\hat{\lambda}$.

Lösung:

- Da die T_i unabhängig und identisch verteilt sind, ergibt sich die Likelihood-Funktion als Produkt der Dichten:

$$L(\lambda) = \prod_{i=1}^n f(t_i; \lambda) = \prod_{i=1}^n \lambda e^{-\lambda t_i} = \lambda^n e^{-\lambda \sum_{i=1}^n t_i}$$

- Logarithmieren macht aus Produkt und Exponentialfunktion eine Summe:

$$\ell(\lambda) = \ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n t_i$$

- Ableiten nach λ und Nullsetzen:

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n t_i \stackrel{!}{=} 0 \quad \implies \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^n t_i} = \frac{1}{\bar{t}}$$

(Die zweite Ableitung $\ell''(\lambda) = -n/\lambda^2 < 0$ bestätigt, dass es sich tatsächlich um ein Maximum handelt.)

- Für die gegebene Stichprobe ist $n = 6$ und

$$\sum_{i=1}^6 t_i = 820 + 910 + 1005 + 760 + 1150 + 880 = 5525, \quad \bar{t} = \frac{5525}{6} \approx 920.83$$

Damit ergibt sich der Schätzwert

$$\hat{\lambda} = \frac{6}{5525} \approx 0.00109 \text{ [1/h]}$$