Übungsaufgaben Kapitel 13
Lösungen

Prof. Dr. Torben Kuhlenkasper

Aufgabe 1

Die durch die Werbeblöcke erzielten täglichen Werbeeinnahmen eines Fernsehsenders können als unabhängige und normalverteilte Zufallsvariablen angesehen werden, deren Erwartungswert davon abhängt, ob ein Werktag vorliegt oder nicht. Für die weitere Auswertung wurden folgende Statistiken berechnet (alle Angaben in Euro) :

Werktage (Mo-Fr) ($n = 36$): $\bar{x} = 72750$, $s = 16350$
Wochenende (Sa-So) ($n = 25$) : $\bar{x} = 187750$, $s = 26350$

Geben Sie ein Schätzverfahren zur Berechnung von 99%-Konfidenzintervallen für die wahren täglichen Werbeeinnahmen an Werktagen bzw. Wochenende an und berechnen Sie die zugehörigen Intervalle.

Gesucht ist ein 99% Konfidenzintervall für μ unter Normalverteilung und unbekannter Varianz:

$$\left[\overline{X} - t_{0.995;n-1} \cdot \frac{s}{\sqrt{n}}, \overline{X} + t_{0.995;n-1} \cdot \frac{s}{\sqrt{n}}\right]$$

Für das Intervall für die Werktage kann wegen $n > 30$ die Standardnormalverteilung verwendet werden und man benutzt die Approximation:

$$\left[\overline{X} - z_{0.995} \cdot \frac{s}{\sqrt{n}}, \overline{X} + z_{0.995} \cdot \frac{s}{\sqrt{n}}\right]$$

und man erhält mit dem Wert aus der Tabelle

$$\left[72750 - 2.58 \cdot \frac{16350}{6}, 72750 + 2.58 \cdot \frac{16350}{6}\right] = [65719.5, 79780.5]$$

Für das Wochenende erhält man mit dem Wert der t-Verteilung bei 24 Freiheitsgraden das Intervall:

$$\left[187750 - 2.7969 \cdot \frac{26350}{5}, 187750 + 2.7969 \cdot \frac{26350}{5}\right] = [173010.34, 202489.66]$$

Aufgabe 2

Bei der Analyse der Dauer von Arbeitslosigkeit wurde der Zusammenhang zwischen Ausbildungsniveau und Dauer der Arbeitslosigkeit untersucht. Unter den 123 Arbeitslosen ohne Ausbildung waren 86 Kurzzeit-, 19 mittelfristige und 18 Langzeitarbeitslose.

(a) Schätzen Sie die Wahrscheinlichkeit, daß ein Arbeitsloser ohne Ausbildung kurzzeitig, mittelfristig oder langfristig arbeitslos ist, und geben Sie für jede der Schätzungen ein 95%- und 99%-Konfidenzintervall an.

Es gilt für die Schätzungen:

$$\begin{aligned} \hat{p}_{kurz} &= \frac{86}{123} \approx 0.699 \\ \hat{p}_{mittel} &= \frac{19}{123} \approx 0.154 \\ \hat{p}_{lang} &= \frac{18}{123} \approx 0.146 \end{aligned}$$

Ein (approximatives) Konfidenzintervall für die Anteilwerte p_i ist gegebn durch

$$\left[\hat{p}_i \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n}}\right]$$

So ergeben sich mit $n = 123$ und $z_{0.975} = 1.96$ bzw. $z_{0.995} = 2.58$ folgende Konfidenzintervalle:

	p_{kurz}	p_{mittel}	p_{lang}
95%	$[0.61814, 0.78024]$	$[0.09060, 0.21835]$	$[0.08388, 0.20881]$
99%	$[0.59250, 0.80587]$	$[0.07039, 0.23855]$	$[0.06412, 0.22857]$

(b) Wieviel größer müsste der Stichprobenumfang sein, um die Länge der Konfidenzintervalle zu halbieren?

Für die Länge l der Konfidenzintervalle gilt

$$l = 2 \cdot z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n}}$$

und somit:

$$\boxed{n \geq \frac{z^2_{1-\alpha/2}}{l^2}}$$

für die neue Länge l^* soll nun gelten $l^* = \frac{1}{2} \cdot l$ und man erhält somit durch einsetzen:

$$\begin{aligned} n &\geq \frac{z^2_{1-\alpha/2}}{(\frac{1}{2}l)^2} \\ n &\geq \frac{z^2_{1-\alpha/2}}{\frac{1}{4}l^2} \\ n &\geq \frac{z^2_{1-\alpha/2}}{l^2} \cdot 4 \end{aligned}$$

Um die Länge zu halbieren, muß n vervierfacht werden.

Aufgabe 3

Für die Durchführung eines Entwicklungshilfeprojekts soll in einem Entwicklungsland zunächst der Anteil der Personen ermittelt werden, die unter dem Existenzminimum leben. In einer Pilotstudie mit $n = 50$ Personen wurden 30 als "arm", d.h. als "unter dem Existenzminimum lebend" eingestuft.

(a) Schätzen Sie aus obigen Angaben den Anteil der Armen in diesem Land.

In der Stichprobe vom Umfang $n = 50$ werden 30 Personen als arm eingestuft.

Für die Schätzung des Anteil der Armen ergibt sich $\hat{p} = \frac{30}{50} = 0.6$

(b) Berechnen Sie ein näherungsweises 90%-Konfidenzintervall für den Anteil der armen Bevölkerung in diesem Entwicklungsland.

Da $n = 50 \geq 30$ ist die Faustregel für die Approximation der t-Verteilung durch die Standardnormalverteilung erfüllt und es gilt:

$$\left[\hat{p}_i \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n}}\right]$$

Mit den Werten $\hat{p} = 0.6$, $\alpha = 0.1$, $z_{0.95} = 1.65$, $n = 50$ sowie $\sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n}} = \sqrt{\frac{0.6 \cdot 0.4}{50}} = 0.06928$ ergibt sich das Intervall

$$[0.486, 0.736]$$

(c) Berechnen Sie ein 95%-Konfidenzintervall für den Anteil der Armen, und vergleichen Sie es mit dem in (b) berechneten.

Nun ist $\alpha = 0.05$, d.h. es ist eine größere Sicherheit verlangt. Mit $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$ erhält man das Intervall

$$[0.464, 0.736]$$

Dieses Konfidenzintervall ist etwas breiter als das unter (b) berechnete, d.h. für den Wunsch nach mehr Sicherheit „zahlt“ man mit größerer Schätzungenauigkeit.

(d) In einer weiteren Zufallsstichprobe werden $n = 200$ Personen befragt. Auch bei dieser größeren Stichprobe ergab sich ein Anteil von 0.6 an Personen, die unter dem Existenzminimum leben. Geben Sie ebenfalls ein 95%-Konfidenzintervall an, und vergleichen Sie es mit dem in (c) berechneten. Womit läßt sich der Unterschied erklären?

Nun ergibt sich aus den vorliegenden Daten $n = 200$, $\hat{p} = 0.6$, $\alpha = 0.05$, $z_{0.975} = 1.96$ sowie $\sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n}} = \sqrt{\frac{0.6 \cdot 0.4}{200}} = 0.034641$. Mit diesen Werten erält man das Intervall

$$[0.532, 0.668]$$

Dieses Konfidenzintervall ist viel kürzer als das unter (c) errechnete. Diese Erhöhung der Schätzgenauigkeit wird durch die Vergrößerung des Stichprobenumfangs erzielt.

Aufgabe 4

Ein Student hat in den ersten drei Semestern an der Hochschule Pforzheim die Punktezahlen in seinen geschriebenen Klausuren erfasst. Er geht von Normalverteilung aus und hat folgende Werte beobachtet:

```
47 51 39 59 45 50
```

Wie viele Punkte kann der Student mindestens in 90% der Klausuren mit einer Wahrscheinlichkeit von 95% erwarten? (Rechenhilfe= $s^2 = 44.7$)

Die Varianz wird hier aus den Daten geschätzt und ist somit unbekannt. Somit ist ein einseitiges Toleranzintervall mit unbekannter Varianz sowie dem Anteil $\gamma = 0.9$, der Sicherheit $1 - \alpha = 0.95$ und $n = 6$ gesucht. Es gilt allgemein:

$$\left[\overline{x} - k^{ue}_{\gamma;1-\alpha;n} \cdot s, \infty\right)$$

Hier ergibt sich aus den Daten $\overline{x} = 48.5$ sowie $s^2 = 44.7$ und somit $s = 6.69$. Aus der Tabelle (S. 12 der Tabellensammlung) erhält man hier $k^{ue}_{0.9;0.95;6} = 3.006$. Es ergibt sich:

$$[48.5 - 3.006 \cdot 6.69, \infty) = [28.39, \infty)$$

Mit einer Wahrscheinlichkeit von 95% werden 90% der Klausuren mindestens eine Punktzahl von 28.39 haben.

Aufgabe 5

Wir betrachten Konfidenzintervalle zum Niveau 95% für μ bei einem normalverteilten Merkmal mit ebenfalls unbekannter Varianz. Welche der folgenden Aussagen sind richtig?

Allgmein ist das 95% Konfidenzintervall für μ bei normalverteiltem Merkmal mit unbekannter Varianz gegeben durch

$$\left[\overline{X} - t_{0.975;n-1} \cdot \frac{s}{\sqrt{n}}, \overline{X} + t_{0.975;n-1} \cdot \frac{s}{\sqrt{n}}\right]$$

(a) Die Breite der Konfidenzintervalle ist zufällig. d.h. bei wiederholter Durchführung des Experiments sind die realisierten Intervalle unterschiedlich breit.

Die in der Breite des Intervalls enthaltene Standardabweichung s ist eine Zufallsvariable und damit auch die Breite des Konfidenzintervalls. Die Aussage ist also richtig.

(b) Bei wiederholter Durchführung des Experiments fällt der Parameter μ mit 95% Wahrscheinlichkeit in das Konfidenzintervall.

Der wahre, aber unbekannte Parameter μ ist **nicht** zufällig, kann also auch nicht in das Konfidenzintervall fallen. Zufällig ist per Konstruktion das Konfidenzintervall. Die Aussage ist somit falsch.

(c) Bei wiederholter Durchführung des Experiments überdeckt das Konfidenzintervall den Parameter μ mit 95% Wahrscheinlichkeit.

Dies ist die korrekte Häufigkeitsinterpretation von Konfidenzintervallen. Das zufällige Intervall überdeckt in 95 Prozent der Fälle den unbekannten Parameter.

(d) Mit wachsendem Stichprobenumfang nimmt die Länge der Konfidenzintervalle im Mittel ab.

Diese Aussage ist richtig, da die Breite reziprok von n abhängt und darüberhinaus die Standardabweichung im Durchschnitt bzw. tendenziell mit wachsendem Stichprobenumfang kleiner wird.

(e) Unter- und Obergrenze eines Konfidenzintervalls sind zufällig. Die Grenzen des Konfidenzintervalls hängen von den zufälligen Größen $\overline{x}$ und s ab, so dass diese tatsächlich zufällig sind. Die Aussage ist somit korrekt.