Übung 11

-

Schätzverfahren
Musterlösung

Aktuelle Version: 30. August 2022

Hinweise:

- Übungen sind mit Vorteil alleine zu lösen.
- Benutzen Sie die Musterlösungen nur zur Korrektur.
- Die Übungen sind wichtige Vorbereitungen für die Prüfung. Lösen sie die Übungen sorgfältig und stellen Sie die Lösungswege übersichtlich dar.
- (Ergänzte) Vorlesungsunterlagen und Fachbücher helfen beim Lösen von Übungen und bringen gleichzeitig eine erweiterte Ansicht auf die Problemstellung.
- Wenn Sie die Übungen nicht verstehen, fragen Sie!

Übung 1. *Fragen*

1. Was sind Schätzverfahren?

   Wenn nur Stichproben zur Verfügung stehen, können die wahren, aber unbekannten Parameter der Grundgesamtheit nicht exakt bestimmt werden. In diesem Fall wird auf Schätzverfahren zurückgegriffen, welche diese Parameterwerte zu schätzen versuchen. Man versucht also, aufgrund der Stichprobe auf die Grundgesamtheit zu schliessen.

2. Welche zwei Arten von Schätzungen gibt es?

   Man unterscheidet:

   (a) Die *Punktschätzung*, bei der nur ein einziger Werte angegeben wird. Bei der Punktschätzung ist entsprechend die Präzision sehr hoch, die Zuverlässigkeit jedoch schlecht.

   (b) Die *Intervallschätzung*, bei der ein Intervall angegeben wird. Hier ist die Präzision tiefer, jedoch die Zuverlässigkeit besser. Oft wird die Präzision (Konfidenz) im Voraus festgelegt.

3. Wie hängen Schätzwerte für die Population mit den entsprechenden Parametern der Stichprobe zusammen?

   (a) Der Schätzwert für den *Mittelwert* der Population μ entspricht dem Stichprobenmittelwert $\bar{x}$.

   (b) Der Schätzwert für die *Varianz* der Population σ^2 entspricht der Stichprobenvarianz s_x^2 plus Korrekturfaktor $\frac{n}{n-1}$.

   (c) Der Schätzwert für die *Anteilswert* der Population p entspricht der Anteilswert der Stichprobe $\bar{p}$.

4. Was versteht man unter Erwartungstreue?

   Der Erwartungswert der Schätzfunktion stimmt mit dem Erwartungswert der Zufallsvariablen in der Grundgesamtheit überein.

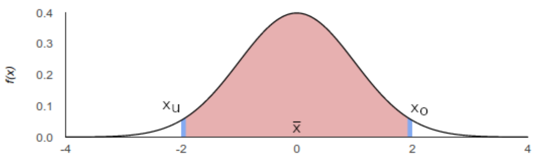
Übung 2. *Konfidenzintervalle von Mittelwerten*

Die Molkerei Alpmilch liefert an eine Lebensmittelkette werktäglich 40'000 Flaschen Milch mit einer Soll-Füllmenge von je 1000 ml. Der letzten Lieferung wurden 25 Flaschen entnommen; in dieser Stichprobe betrug die durchschnittliche Füllmenge 1000.55 ml. Aufgrund zahlreicher Kontrollen weiss man, dass die Ist-Füllmenge normalverteilt ist, mit einer Streuung von $\sigma = 1.2ml$.

1. Konfidenzintervalle

   (a) Erstellen Sie das zentrale 95%-Konfidenzintervall für die durchschnittliche Füllmenge der 40.000 Flaschen.

   Hier haben wir den Fall, dass wir eine konkrete Messung der Stichprobe des Mittelwertes vorliegen haben. Nun suchen wir den Mittelwert der Abfüllmenge aller abgefüllten Milchflaschen. Wir wissen, dass der Mittelwert wieder eine Verteilungsfunktion darstellt, die wir als normalverteilt angenehmen dürfen. Wir suchen daher also die Grenzen des Intervalls, indem die Werte zu 95% liegen:



Die Standardabweichung dieser Normalverteilung ist:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.2}{\sqrt{25}} = 0.24$$

Die Z-Werte lesen wir aus der Tabelle:

$$\begin{aligned} Z_{(1-\alpha/2)} &= Z_{(0.975)} = 1.96 \\ Z_{(\alpha/2)} &= Z_{(0.025)} = -1.96 \end{aligned}$$

Somit können wir das Konfidenzintervall für μ berechnen:

$$\begin{aligned} \mu = [x_u, x_o] &= \left[\bar{x} + \sigma_{\bar{x}} Z_{(1-\alpha/2)}, \bar{x} + \sigma_{\bar{x}} Z_{(\alpha/2)}\right] \\ &= [1000.55 - 0.24 \cdot 1.96, 1000.55 + 0.24 \cdot 1.96] \\ &= [1000.08, 1001.02] \end{aligned}$$

(b) Erstellen Sie das zentrale 99%-Konfidenzintervall für μ.

Eine höhere Sicherheit bedeutet eine grössere Fläche unter der Kurve, das Intervall vergrössert sich entsprechend:

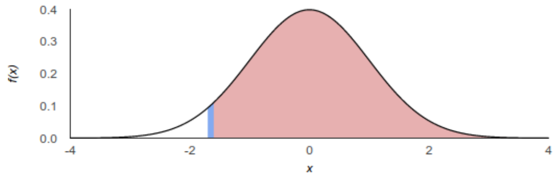
$$\begin{aligned} Z_{(1-\alpha/2)} &= Z_{(0.995)} = 2.58 \\ Z_{(\alpha/2)} &= Z_{(0.005)} = -2.58 \\ \mu &= \left[\bar{x} + \sigma_{\bar{x}} Z_{(1-\alpha/2)}, \bar{x} + \sigma_{\bar{x}} Z_{(\alpha/2)}\right] \\ &= [1000.55 - 0.24 \cdot 2.58, 1000.55 + 0.24 \cdot 2.58] \\ &= [999.93, 1001.16] \end{aligned}$$

(c) Erstellen Sie das zentrale 95% -Konfidenzintervall für μ, für den Fall, dass der Stichprobenumfang 36 Flaschen umfasst.

Durch eine grössere Probe wird die Standardabweichung kleiner und damit wird auch das Konfidenzintervall kleiner. Somit wird das Ergebnis sicherer:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.2}{\sqrt{36}} = 0.2$$
$$Z_{(1-\alpha/2)} = Z_{(0.975)} = 1.96$$
$$Z_{(\alpha/2)} = Z_{(0.025)} = -1.96$$
$$\mu = \left[\bar{x} + \sigma_{\bar{x}} Z_{(1-\alpha/2)}, \bar{x} + \sigma_{\bar{x}} Z_{(\alpha/2)}\right]$$
$$= [1000.55 - 0.2 \cdot 1.96, 1000.55 + 0.2 \cdot 1.96]$$
$$= [1000.16, 1000.94]$$

(d) Erstellen Sie das nach unten begrenzte 95%-Konfidenzintervall für μ.
Nun ist nur noch die untere Grenze des Intervalls gesucht:



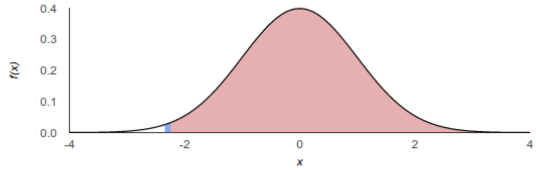
$$Z_{(\alpha)} = Z_{(0.05)} = -1.65$$
$$\mu \geq \bar{x} + \sigma_{\bar{x}} Z_{(\alpha)} = 1000.55 - 0.24 \cdot 1.65 = 1000.15$$

2. Konfidenz

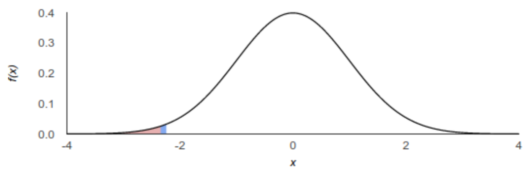
(a) Ermitteln Sie die Konfidenz für das mit 1000 ml nach unten begrenzte Intervall für μ.
Jetzt fragen wir nicht nach dem Intervall von μ sondern nach der Wahrscheinlichkeit, dass μ grösser gleich 1000ml ist. Damit ergibt sich ein Z von:

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{1000.55 - 1000}{0.24} = 2.29$$

Aus der Tabelle lesen wir eine entsprechende Wahrscheinlichkeit von $P(\mu \geq 1000) = 0.989$, d.h. der Mittelwert μ wird mit einer Wahrscheinlichkeit von 98.9% im Intervall $[1000, \infty]$ liegen.



(b) Ermitteln Sie die Konfidenz für das mit 1000 ml nach oben begrenzte Intervall für μ.
Nun suchen wir genau die den komplementären Teil der Fläche:

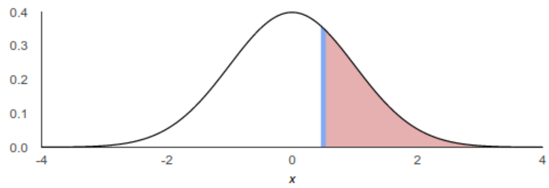


Damit ist $P(\mu < 1000) = 1 - P(\mu \geq 1000) = 1 - 0.989 = 0.011$.

(c) Ermitteln Sie die Konfidenz für das mit 1000 ml nach unten begrenzte Intervall für μ, für den Fall, dass die durchschnittliche Füllmenge in der Stichprobe nur 999.88 ml betragen hat!

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{999.88 - 1000}{0.24} = -0.5$$

Aus der Tabelle lesen wir die entsprechende Wahrscheinlichkeit von $P(\mu \geq 1000) = .309$.



3. Stichprobengrösse

(a) Wie viele Flaschen Milch müssen der Lieferung entnommen und geprüft werden, wenn das zentrale 95%-Konfidenzintervall für μ eine Genauigkeit von $e = \bar{x} - \mu = 0.25ml$ aufweisen soll?

$$Z_{(1-\alpha/2)} = Z_{(0.975)} = 1.96$$
$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\frac{e}{\sigma}$$
$$n = \left(\frac{Z\sigma}{e}\right)^2 = \left(\frac{1.96 \cdot 1.2}{0.25}\right)^2 = 88.5$$

Es müssen also mindestens 89 Flaschen Milch entnommen werden.

(b) Wie viele, wenn Sie mit einer Wahrscheinlichkeit von 99.5% und einer Genauigkeit von $e = \bar{x} - \mu = 0.25ml$ sicher sein möchte, dass die Sollfüllmenge in der Grundgesamtheit nicht unterschritten wird?

Hier ist nun ein einseitiges Intervall gesucht.

$$Z_{(1-\alpha)} = Z_{(0.995)} = 2.58$$
$$n = \left(\frac{Z\sigma}{e}\right)^2 = \left(\frac{2.58 \cdot 1.2}{0.25}\right)^2 = 153.4$$

Es müssen also mindestens 154 Flaschen Milch entnommen werden.

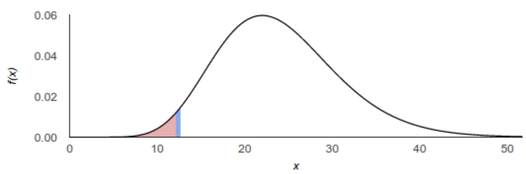
Übung 3. *Konfidenzintervalle von Varianzen*
Auf einer Anlage wird Zucker in Tüten abgefüllt. Das Soll-Füllgewicht beträgt 1000g. Aufgrund zahlreicher Messreihen ist bekannt, dass die Füllmenge der Tüten normalverteilt ist. Um die Anlage so einstellen zu können, dass höchstens 3% der Tüten das Soll-Füllgewicht unterschreiten, muss die Ungenauigkeit der Anlage in Form der Varianz bekannt sein.

Aus der Tagesproduktion von 90.000 Zuckertüten wurden 25 Tüten zufällig entnommen und gewogen. Die Varianz s^2 in dieser Stichprobe betrug $0.6g^2$.

1. Ermitteln Sie das zweiseitige 95%-Konfidenzintervall für die Varianz.

   Wir lesen die Werte aus der Tabelle der Quantile der χ^2 Verteilung und ermitteln das Konfidenzintervall:

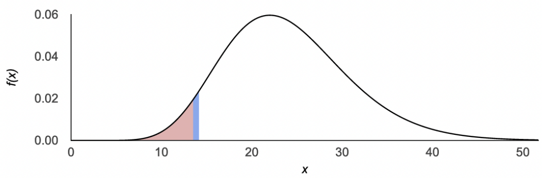
$$\chi^2_{(1-\alpha/2,n-1)} = \chi^2_{(0.975,24)} = 39.3641$$
$$\chi^2_{(\alpha/2,n-1)} = \chi^2_{(0.025,24)} = 12.4012$$



$$\sigma^2 = \left[\frac{s^2(n-1)}{\chi^2_{(1-\alpha/2,n-1)}}, \frac{s^2(n-1)}{\chi^2_{(\alpha/2,n-1)}}\right]$$
$$= \left[\frac{0.6\cdot 24}{39.3641}, \frac{0.6\cdot 24}{12.4012}\right] = [0.366, 1.161]$$

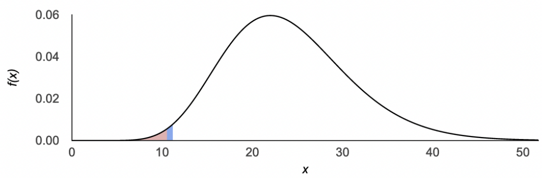
2. Ermitteln Sie das nach oben begrenzte 95%-Konfidenzintervall für die Varianz.

$$\chi^2_{(\alpha,n-1)} = \chi^2_{(0.05,24)} = 13.8484$$
$$\sigma^2 \leq \frac{s^2(n-1)}{\chi^2_{(\alpha/2,n-1)}} = \frac{0.6\cdot 24}{13.8484} = 1.040$$



3. Ermitteln Sie das nach oben begrenzte 99%-Konfidenzintervall für die Varianz.

$$\chi^2_{(\alpha,n-1)} = \chi^2_{(0.01,24)} = 10.8564$$
$$\sigma^2 \leq \frac{s^2(n-1)}{\chi^2_{(\alpha/2,n-1)}} = \frac{0.6 \cdot 24}{10.8564} = 1.326$$



Zusatzaufgaben

Übung 4. *Konfidenzintervalle*
In dem Produktionsvorgang von Schrauben werden zur Qualitätssicherung regelmässig 10 Schrauben entnommen und überprüft. In einer konkreten Entnahme werden die folgenden Grössen gemessen:

- Die Länge der Schraube ergab im Mittel $\bar{x} = 120.6mm$.
- Für die Stichprobe wurde die folgende Varianz ermittelt:

$$s_s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 = (5.44mm)^2$$

1. Wie gross schätzen Sie die Varianz der Population?

   Wir machen eine Punktschätzung der Varianz σ:

$$\sigma = s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$$

   mit

$$\sum(x_i - \bar{x})^2 = ns_s^2$$

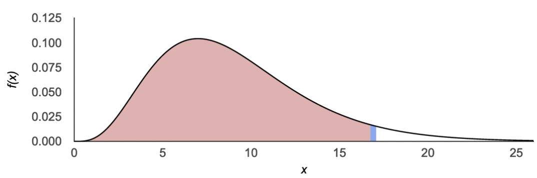
   folgt

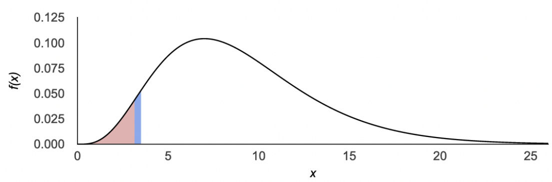
$$\sigma^2 = \frac{n}{n-1}s_s^2 = \frac{10}{9}(5.44)^2 = 32.88$$

2. Wie gross ist die Standardabweichung zu 90% respektive 99%?

   90% Konfidenzintervall:

$$\chi^2_{(0.95,9)} = 16.92$$
$$\chi^2_{(0.05,9)} = 3.33$$

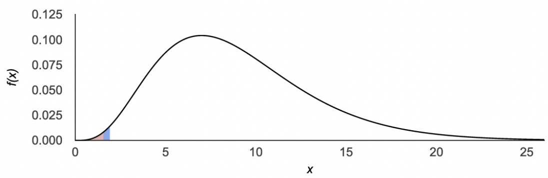


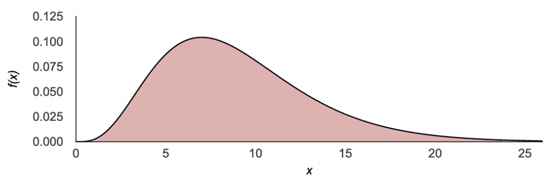


$$\sigma^2 = \left[\frac{s^2(n-1)}{\chi^2_{(1-\alpha/2,n-1)}}, \frac{s^2(n-1)}{\chi^2_{(\alpha/2,n-1)}}\right]$$
$$= \left[\frac{5.44^2 \cdot 9}{16.92}, \frac{5.44^2 \cdot 9}{3.33}\right] = [15.74, 79.98]$$

99% Konfidenzintervall:

$$\chi^2_{(0.995,9)} = 23.59$$
$$\chi^2_{(0.005,9)} = 1.73$$





$$\sigma^2 = \left[\frac{s^2(n-1)}{\chi^2_{(1-\alpha/2,n-1)}}, \frac{s^2(n-1)}{\chi^2_{(\alpha/2,n-1)}}\right]$$
$$= \left[\frac{5.44^2 \cdot 9}{23.59}, \frac{5.44^2 \cdot 9}{1.73}\right] = [11.29, 153.96]$$

3. Wie gross schätzen Sie den Mittelwert?

   Wir machen eine Punktschätzung des Mittelwertes aufgrund des Stichprobenmittelwerts:

   $$\mu = \bar{x} = 120.6$$

4. Wie gross ist der Mittelwert zu 90% respektive 99%?

   Auch hier schliessen oder schätzen wir ausgehend von einer Stichprobe auf die Gesamtheit. Mit nur 10 gemessenen Daten ist die Verwendung der t-Verteilung mit dem Freiheitgrad $\nu = 9$ angezeigt.

   90% Konfidenzintervall:

$$t_{(0.95,9)} = 1.83$$
$$\mu = \bar{x} \pm t\frac{\sigma}{\sqrt{n}} = 120.6 \pm 1.83\frac{32.88}{\sqrt{10}} = [101.6, 139.6] \qquad (1)$$

99% Konfidenzintervall:

$$t_{(0.995,9)} = 3.25$$
$$\mu = \bar{x} \pm t\frac{\sigma}{\sqrt{n}} = 120.6 \pm 3.25\frac{32.88}{\sqrt{10}} = [86.8, 154.4] \quad (2)$$

Übung 5. *Stichprobengrösse bei Anteilswerten*
Sie möchten den Anteil der über 40-Jährigen in der Bevölkerung ermitteln. Wie gross wählen Sie ihre Stichprobe, wenn Sie den Anteil auf $e = \pm 0.1\%$ genau mit 95% Konfidenz bestimmen wollen?

Die minimale Stichprobegrösse berechnet sich mit:

$$n \geq \frac{z^2\sigma^2}{e^2}$$
$$\sigma^2 = p(1-p)$$
$$Z_{(1-\alpha/2)} = Z_{(0.975)} = 1.96$$

σ^2 ist unbekannt, hängt aber nur von p ab. Wir können deshalb die maximale Varianz berechnen um damit eine konservative Schätzung der minimalen Stichprobengrösse zu berechnen. Wir leiten dazu ab und setzen gleich Null:

$$\frac{d\sigma^2}{dp} = \frac{d}{dp}\left(p(1-p)\right) = \frac{d}{dp}\left(p - p^2\right) = 1 - 2p = 0$$
$$p = 0.5$$
$$\sigma^2 = p(1-p) = 0.25$$

Wir erhalten:

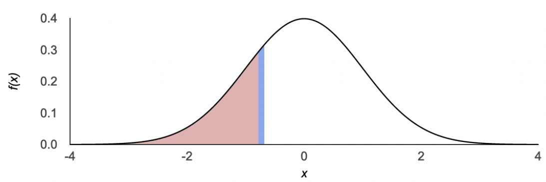
$$n \geq \frac{z^2\sigma^2}{e^2} = \frac{1.96^2 0.25}{0.1^2} = 96.04$$

Wir müssen also Stichprobe von mindestens 97 Personen nehmen.

Übung 6. *Konfidenzintervalle von Mittelwerten*
Es seien die Ergebnisse bei Prüfungen an einer Hochschule normalverteilt mit einer Standardabweichung von $\sigma = 1.5$ und einem Durchschnitt von $\mu = 4.5$ Notenpunkten. In der letzten Prüfungssession fanden $n = 30$ Prüfungen statt. Der Durchschnitt lag bei $\bar{x} = 4.3$ Notenpunkten.

1. Ermitteln Sie die Konfidenz für das mit $\bar{x} = 4.3$ nach oben begrenzte Intervall für μ. Was schliessen Sie daraus?

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{30}} = 0.27$$
$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{4.3 - 4.5}{0.27} = -0.73$$
$$P(\mu < 4.3) = 23.3\%$$



Alternativ können wir auch eine t-Verteilung mit $\nu = n - 1 = 29$ Freiheitgraden verwenden und erhalten mit dem bereits berechneten z-Wert $P(\bar{x} < 4) = F(-0.73) = 23.6\%$.

Es handelt sich bei den Ergebnissen der Prüfungsession als mit hoher Wahrscheinlichkeit tatsächlich um eine Stichprobe der Normalverteilung $N(\mu, \sigma)$.

2. Was würde sich bei einem unterschiedlichen Sessionsdurchschnitt $\bar{x}$ ändern?

   Je mehr der Durchschnitt $\bar{x}$ von μ abweicht, d.h. je grösser $|\bar{x} - \mu|$, desto kleiner ist die Wahrscheinlichkeit, dass es sich tatsächlich um eine Stichprobe der Normalverteilung $N(\mu, \sigma)$ handelt.

3. Was würde sich bei einer unterschiedlichen Anzahl Prüfungen n ändern?

   Mit höherem n wird die Verteilung der Mittelwerte schmaler, d.h. es wird erwartet, dass der Sessionsdurchschnitt $\bar{x}$ näher an μ sein muss. Die Wahrscheinlichkeit, dass es sich tatsächlich um eine Stichprobe der Normalverteilung $N(\mu, \sigma)$ handelt, wird also kleiner mit steigendem n.

   Umgekehrt wird der Bereich der erwarteten Stichprobenmittelwerte mit kleinem n grösser, die Wahrscheinlichkeit steigt.