Aufgabe 36 –2 (S. 346)**: Schätzverfahren für Mittelwert und Standardabweichung**

a) Punktschätzungen für m aufgrund der Werte der kleinen Stichprobe aus Aufgabe 32

Bei einer Punktschätzung wird für den zu schätzenden Parameter der Grundgesamtheit aufgrund des Ergebnisses der Stichprobe lediglich ein einziger Schätzwert angegeben, also ist

$$\dot{m} = \bar{x} = 1600,\!-DM$$

b) Bestimmung von Konfidenzbereichen für m**:**

mit $(1-a_0) = 0{,}95$ bzw. $0{,}99$.

$$\bar{x} = 1600,\!-DM$$
$$\hat{s} = 250,\!-DM$$
$$n = 10$$
$$\Rightarrow \frac{n}{N} = \frac{10}{50000} = 0{,}0002 \leq 0{,}05$$, also kein Endlichkeitsfaktor

s ist unbekannt und $n = 10 \leq 30$, damit ist

$$\tilde{m} = \bar{x} \pm t_{\left(a_0/2;f\right)} \cdot \frac{\hat{s}}{\sqrt{n}}$$

b1) Einsetzen **bei** $(1 - a_0) = 0{,}95$**:**

$$K\left(\bar{x} - t_{0,025;9} \cdot \frac{\hat{s}}{\sqrt{n}} \leq m \leq \bar{x} - t_{0,025;9} \cdot \frac{\hat{s}}{\sqrt{n}}\right) = 0{,}95$$
$$K\left(1600 - 2{,}262 \cdot \frac{250}{\sqrt{10}} \leq m \leq 1600 + 2{,}262 \cdot \frac{250}{\sqrt{10}}\right) = 0{,}95$$
$$K(1421{,}17DM \leq m \leq 1778{,}83DM) = 0{,}95$$

Bei 95%iger Konfidenz wäre der Mittelwert der GG zwischen 1421,17DM und 1778,83DM zu erwarten.

b2) Einsetzen **bei** $(1 - a_0) = 0{,}99$**:**

$$K\left(\bar{x} - t_{0,005;9} \cdot \frac{\hat{s}}{\sqrt{n}} \le m \le \bar{x} + t_{0,005;9} \cdot \frac{\hat{s}}{\sqrt{n}}\right) = 0,99$$

$$K\left(1600 - 3,250 \cdot \frac{250}{\sqrt{10}} \le m \le 1600 + 3,250 \cdot \frac{250}{\sqrt{10}}\right) = 0,99$$

$$K(1343,06DM \le m \le 1856,94DM) = 0,99$$

Bei 99%iger Konfidenz wäre der Mittelwert der GG zwischen 1343,06DM und 1856,95DM zu erwarten.
Anmerkung: Ein prozentualer Vergleich der Konfidenzbereiche zeigt, daß eine Erhöhung des Konfidenzniveaus von 95% auf 99% zu einer Vergrößerung des Unsicherheitsbereiches um 44% führt. Dies entspricht dem prozentualen Anwachsen des $t_{\frac{a_o}{2}}$ - Wertes.

a) Konfidenzintervalle für m für die größere Stichprobe (aus Aufgabe 32)

Ausgangsdaten für die Bereichsschätzungen mit einem Konfidenzniveau von 0,95 :

$N = 10000 \qquad n = 550 \qquad \bar{x} = 1550,-DM \qquad \hat{s} = 220,-DM$
$(1 - a_0) = 0,95$ bzw. 0,99
Der Endlichkeitsfaktor musste angewandt werden, weil
$\frac{n}{N} = \frac{550}{10000} = 0,055 > 0,05$

s ist unbekannt und $n = 550 > 30$, daraus folgt

$$\tilde{m} = \bar{x} \pm z_{a_0/2} \cdot \frac{\hat{s}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

$K(m_1 \le m \le m_2) = 1 - a_0$, daraus folgt für die Berechnung der Konfidenzbereiche bei $K = 0,95 \Rightarrow z = 1,96$

$$K\left(\bar{x} - z_{a_{0,05}/2} \cdot \frac{\hat{s}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \le m \le \bar{x} + z_{a_{0,05}/2} \cdot \frac{\hat{s}}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}\right) = 0,95$$

$$K\left(1550 - 1,96 \cdot \frac{220}{\sqrt{550}} \cdot \sqrt{\frac{10000-550}{10000-1}} \le m \le 1550 + 1,96 \cdot \frac{220}{\sqrt{550}} \cdot \sqrt{\frac{10000-550}{10000-1}}\right) = 0,95$$

$$K(1532,13DM \le m \le 1567,87DM) = 0,95$$

Bei 95%iger Konfidenz wäre der Mittelwert der GG zwischen 1532,13DM und 1567,87DM zu erwarten.

Aufgabe 37: Schätzverfahren für Mittelwerte

$N = 10000 \qquad n = 25 \qquad \overline{x} = 250,\!-DM \qquad \hat{s} = 50,\!-DM$

a) Berechnung eines Konfidenzintervalls für m **bei n=25** und einer 95%igen Konfidenz

s ist unbekannt, $n = 25 \leq 30$ $\Rightarrow$ Das Intervall ist t – verteilt. Daraus folgt

$K\left(m_1 \leq m \leq m_2 \right) = 1 - a$

$\tilde{m} = \overline{x} \pm t_{\left(a_0/2;f\right)} \cdot \frac{\hat{s}}{\sqrt{n}}$

Auf den Endlichkeitsfaktor kann verzichtet werden, weil
$\frac{n}{N} = \frac{25}{10000} = 0,0025 \leq 0,05$ ist:

$$K\left(\overline{x} - t_{\left(a_{0,05}/2;24\right)} \cdot \frac{\hat{s}}{\sqrt{n}} \leq m \leq \overline{x} + t_{\left(a_{0,05}/2;24\right)} \cdot \frac{\hat{s}}{\sqrt{n}} \right) = 0,95$$

$$K\left(250 - 2,064 \cdot \frac{50}{\sqrt{25}} \leq m \leq 250 + 2,064 \cdot \frac{50}{\sqrt{25}} \right) = 0,95$$

$$K\left(229,36DM \leq m \leq 270,64DM \right) = 0,95$$

Der Durchschnitt der Ausgaben für Fleisch pro Haushalt in diesem Stadtteil bewegt sich bei 95%igem Vertrauen im Intervall von 229,36DM bis 270,64DM.

b) Konfidenzintervall für m **bei n=100:**

Dabei ergeben sich die selben Stichprobenwerte. Wie groß ist das Konfidenzintervall jetzt?

$N = 10000 \qquad n = 100 \qquad \overline{x} = 250,\!-DM \qquad \hat{s} = 50,\!-DM$

s ist unbekannt, $n = 100 > 30$.

Daraus folgt

$K\left(m_1 \leq m \leq m_2 \right) = 1 - a \qquad \tilde{m} = \overline{x} \pm z_{\left(a_0/2\right)} \cdot \frac{\hat{s}}{\sqrt{n}}$

Auf den Endlichkeitsfaktor kann verzichtet werden, weil
$\frac{n}{N} = \frac{100}{10000} = 0,01 \leq 0,05$ ist.

Einsetzen:

$$K\left(\overline{x} - z_{\left(a_{0,05}/2\right)} \cdot \frac{\hat{s}}{\sqrt{n}} \leq m \leq \overline{x} + z_{\left(a_{0,05}/2\right)} \cdot \frac{\hat{s}}{\sqrt{n}}\right) = 0,95$$

$$K\left(250 - 1{,}96 \cdot \frac{50}{\sqrt{100}} \leq m \leq 250 + 1{,}96 \cdot \frac{50}{\sqrt{100}}\right) = 0,95$$

$$K(240{,}20DM \leq m \leq 259{,}80DM) = 0,95$$

Der Durchschnitt der Ausgaben für Fleisch pro Haushalt in diesem Stadtteil liegt bei 95%igem Vertrauen im Intervall von 240,20DM bis 259,80DM.

Aufgabe 38: Schätzverfahren für s:

a) Punktschätzung für s:

Bei einer Punktschätzung wird für den zu schätzenden Parameter der Grundgesamtheit aufgrund des Ergebnisses der Stichprobe lediglich ein einziger Schätzwert angegeben, also ist

$$\dot{s} = \hat{s} = 250DM$$

$$n = 10 \leq 30,$$

damit ist

b) Konfidenzintervalle für s bei kleinen Stichproben:

GG = NV, n = 10 und $\hat{s}$ = 250,
da $n = 10 \leq 30$ ergeben sich die Konfidenzintervalle wie folgt:

b1) bei $(1 - a_0) = 0{,}95$:

$$K\left(\sqrt{\frac{n-1}{c^2_{a_{0,05}/2;f}} \cdot \hat{s}^2} \leq s \leq \sqrt{\frac{n-1}{c^2_{1-a_{0,05}/2;f}} \cdot \hat{s}^2}\right) = 0,95$$

$$K\left(\sqrt{\frac{10-1}{19{,}023} \cdot 250^2} \leq s \leq \sqrt{\frac{10-1}{2{,}700} \cdot 250^2}\right) = 0,95$$

$$K(171{,}96DM \leq s \leq 456{,}44DM) = 0,95$$

Bei 95%iger Konfidenz wäre die Standardabweichung der GG zwischen 171,96DM und 456,44DM zu erwarten.

b2) bei $(1 - a_0) = 0{,}99$:

$$K\left(\sqrt{\frac{n-1}{c^2_{\alpha_{0,01}/2;f}}\cdot\hat{s}^2} \le s \le \sqrt{\frac{n-1}{c^2_{1-\alpha_{0,01}/2;f}}\cdot\hat{s}^2}\right)=0,99$$

$$K\left(\sqrt{\frac{10-1}{23,589}\cdot 250^2} \le s \le \sqrt{\frac{10-1}{1,735}\cdot 250^2}\right)=0,99$$

$$K=(154,42DM \le s \le 569,39DM)=0,99$$

Bei 99%iger Konfidenz wäre eine Standardabweichung der GG zwischen 154,47DM und 569,39DM zu erwarten.
Anmerkung: Bei der Standardabweichung wächst der Konfidenzbereich mit steigendem Konfidenzniveau um 46%. Eine einfache Proportionalität zwischen der Veränderung des Konfidenzbereiches und der c^2 - Werte ist nicht gegeben.

c) Konfidenzintervalle für s bei großen Stichproben:

c1) bei $(1-a_0)=0,95$:

Aus $n=550>30$ folgt ein 95%-Konfidenzbereich mit:

$$K\left(\frac{\hat{s}\cdot\sqrt{2\cdot f}}{\sqrt{2\cdot f-1}+z_{a_0/2}} \le s \le \frac{\hat{s}\cdot\sqrt{2\cdot f}}{\sqrt{2\cdot f-1}-z_{a_0/2}}\right)=1-a$$

für K = 0,95 $\Rightarrow$ z = 1,96 und $f=n-1=550-1=549$

$$K\left(\frac{\hat{s}\cdot\sqrt{2\cdot f}}{\sqrt{2\cdot f-1}+Z_{a_0/2}} \le s \le \frac{\hat{s}\cdot\sqrt{2\cdot f}}{\sqrt{2\cdot f-1}-Z_{a_0/2}}\right)=0,95$$

$$K\left(\frac{220\cdot\sqrt{2\cdot 549}}{\sqrt{2\cdot 549-1}+1,96} \le s \le \frac{220\cdot\sqrt{2\cdot 549}}{\sqrt{2\cdot 549-1}-1,96}\right)=0,95$$

$$K(207,80DM \le s \le 233,95DM)=0,95$$

Bei 95%iger Konfidenz wäre die Standardabweichung in der GG zwischen 207,80DM und 233,95DM zu erwarten.

c2) bei $(1-a_0)=0,99$:

Aus $n=550>30$ folgt ein 99%-Konfidenzbereich mit:

$$K\left[\frac{220\sqrt{2\cdot 549}}{\sqrt{2\cdot 549-1}+2,58} \le s \le \frac{220\sqrt{2\cdot 549}}{\sqrt{2\cdot 549-1}-2,58}\right]=$$

$$K[204,19 \le s \le 238,69]=0,95$$

Bei 99%iger Konfidenz wäre also eine Standardabweichung in der GG zwischen 204,19 DM und 238,69 DM zu erwarten.

zusätzlich zu den gestellten Aufgaben werden die Konfidenzintervalle für s aus den Aufgabe 37 a) und b) berechnet:

a) Konfidenzintervalle für s bei einer kleinen Stichproben

n= 25, $\hat{s}$ = 50 DM, Konfidenzniveau = 0,95

$n = 25 \leq 30$, daraus folgt:

$$K\left(\sqrt{\frac{n-1}{c^{2}_{\left(a/2;f\right)}}\cdot\hat{s}^{2}} \leq s \leq \sqrt{\frac{n-1}{c^{2}_{\left(1-a/2;f\right)}}\cdot\hat{s}^{2}}\right)=1-a$$

$$K\left(\sqrt{\frac{25-1}{39{,}364}\cdot 50^{2}} \leq s \leq \sqrt{\frac{25-1}{12{,}401}\cdot 50^{2}}\right)=0{,}95$$

$$K\left(39{,}04DM \leq s \leq 69{,}56DM\right)=0{,}95$$

Die Standardabweichung der Ausgaben für Fleisch pro Haushalt in diesem Stadtteil bewegt sich bei 95%igem Vertrauen im Intervall von 39,04DM bis 69,56DM.

b) Konfidenzintervall für s bei einer großen Stichprobe:

$n = 100 > 30$, daraus folgt:

$$K\left(\frac{\hat{s}\cdot\sqrt{2f}}{\sqrt{2f-1}+z_{\left(a/2\right)}} \leq s \leq \frac{\hat{s}\cdot\sqrt{2f}}{\sqrt{2f-1}-z_{\left(a/2\right)}}\right)=1-a$$

Einsetzen:

$$K\left(\frac{50\cdot\sqrt{2\cdot 99}}{\sqrt{2\cdot 99-1}+1{,}96} \leq s \leq \frac{50\cdot\sqrt{2\cdot 99}}{\sqrt{2\cdot 99-1}-1{,}96}\right)=0{,}95$$

$$K\left(43{,}98DM \leq s \leq 58{,}26DM\right)=0{,}95$$

Die Standardabweichung der Ausgaben für Fleisch pro Haushalt in diesem Stadtteil liegt bei 95%igem Vertrauen im Intervall von 43,98DM bis 58,26DM.

Aufgabe 39: Konfidenzschätzung für Anteilswerte

n = 2000 k = 80 a_0 = 0,01 bzw. a_0 = 0,05

$$\Rightarrow p = \frac{k}{n} = \frac{80}{2000} = 0{,}04$$

$\Rightarrow VAR(k) = n \cdot p \cdot (1-p) = 2000 \cdot 0{,}04 \cdot 0{,}96 = 76{,}8 \geq 9$, also Approximation über $N_{(0;1)}$

Der Endlichkeitsfaktor muß nicht verwendet werden, weil
$\frac{n}{N} = \frac{2000}{50000} = 0{,}04 < 0{,}05 \Rightarrow$
bei a_0 = 0,05 $\Rightarrow$ $Z_{\frac{a_0}{2}}$ = 1,96
bei a_0= 0,1 $\Rightarrow$ $Z_{\frac{a_0}{2}}$= 2,58
Formel:

$$\tilde{p} = p \pm z_{\frac{a_0}{2}} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Die $z_{a_0/2}$-Werte ergeben sich wie folgt:

$1 - a_{0,05} = 0{,}95 \Rightarrow z_{a_{0,05}/2} = 1{,}96$
Ergebnis:

$$K\left(p - z_{\frac{a_0}{2}} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \leq p \leq p + z_{\frac{a_0}{2}} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \right) = 1 - a_0$$

$$K\left(0{,}04 - 1{,}96 \cdot \sqrt{\frac{0{,}04 \cdot 0{,}96}{2000}} \leq p \leq 0{,}04 + 1{,}96 \cdot \sqrt{\frac{0{,}04 \cdot 0{,}96}{2000}} \right) = 1 - 0{,}05$$

$$K(0{,}031 \leq p \leq 0{,}049) = 0{,}95$$

Mit 95%iger Konfidenz (Vertrauen) ist der Anteilswert in der GG zwischen 3,1% und 4,9% zu erwarten.

Aufgabe 40: Bestimmung des Stichprobenumfangs für die Schätzung von m

a) Zusammenhang zwischen n und dem Konfidenzbereich

Je größer der Stichprobenumfang, desto kleiner die Konfidenzbereiche für μ und σ.

Vergleichende Analyse der Wirkung einer Stichprobenumfangserhöhung auf die Aussagegenauigkeit des Schätzverfahrens:

Bei $n = 25$ ergaben sich folgende Konfidenzintervalle:

$K(229{,}36DM \le m \le 270{,}64DM) = 0{,}95$; das entspricht einer Spannweite von $41{,}28\,DM$
$K(39{,}04DM \le s \le 69{,}56DM) = 0{,}95$; das entspricht einer Spannweite von $30{,}52\,DM$

Bei $n = 100$ ergaben sich folgende Konfidenzintervalle:

$K(240{,}20DM \le m \le 259{,}80DM) = 0{,}95$; das entspricht einer Spannweite von $19{,}60\,DM$
$K(43{,}98DM \le s \le 58{,}26DM) = 0{,}95$; das entspricht einer Spannweite von $14{,}28\,DM$

Eine Vervierfachung des Stichprobenumfangs bewirkt hier etwa eine Halbierung der Spannweite der Konfidenzbereiche.

b) minimale Größe der Stichprobe

Wie groß muss „n" gewählt werden, damit bei 95%iger Konfidenz und $s = 49$ der maximale Schätzfehler ($e = |\bar{x} - m|$) nicht mehr als 5DM ($e \le 5DM$) ist?

Aus $z_{\left(a_{0,05}/2\right)} = 1{,}96$ folgt durch Einsetzen in

$$\Leftrightarrow n = \frac{z^2_{\left(a_0/2\right)} \cdot s^2}{e^2}$$

$$\Leftrightarrow n = \frac{1{,}96^2 \cdot 49^2}{5^2} = 368{,}95$$

Es müssen mindestens 369 Personen befragt werden.

Aufgabe 41: Bestimmung des Stichprobenumfangs für die Schätzung von p

Verwendung der Daten aus Aufgabe 33c)
Wie groß muss die Stichprobe (n) sein, damit der Stichprobenfehler $(e = |p - p|)$ nicht größer als 0,5% ist?

für a_0 = 0,05. Vermutung p=π_0=0,05 (5%-Hürde).

Formel:

$$Z_{\frac{a_0}{2}} = \frac{p - p_0}{\sqrt{\frac{p \cdot (1 - p)}{n}}}$$

umgestellt nach n:

$$n = \frac{Z^2_{a/2} \cdot [p_0 \cdot (1 - p_0)]}{e^2}$$

laut Tabelle ist $z_{\frac{a_0}{2}} = 1{,}96$ bei $a_0 = 0{,}05$.

Eingesetzt ergibt sich:

$$n = \frac{1{,}96^2 \cdot 0{,}05 \cdot (1 - 0{,}05)}{0{,}005^2} = 7299{,}04 \approx 7300$$ Personen bei 95%iger Konfidenz.

Achtung: Bei diesem n ist sicherheitshalber zu prüfen, ob der Endlichkeitsfaktor hätte berücksichtigt werden müssen!

den Auswahlsatz testen: $\frac{n}{N} > 0{,}05$? $\frac{7300}{50000} = 0{,}145981 \Rightarrow Ef!$

Formel mit *Ef*: $n = \frac{Z^2_{a/2} \cdot p_0 \cdot (1 - p_0) \cdot (N - n)}{e^2 \cdot (N - 1)}$

Eingesetzt: $n = \frac{1{,}96^2 \cdot 0{,}05 \cdot 0{,}95 \cdot (50000 - 7300)}{0{,}005^2 \cdot 49999} = 6233{,}51 \approx 6234$

Wichtig: Es wird immer aufzurunden, weil es sich um eine Mindestanforderung handelt.

Die Stichprobe müsste 6234 Menschen umfassen, damit ein solch geringer Stichprobenfehler erwartet werden kann.