Bestimmung von Vertrauensintervallen bei unbekannten Wahrscheinlichkeiten

Gegeben sein müssen *n* und *X* bzw. h_n.

n:	Umfang der Stichprobe
X:	Absolute Häufigkeit eines Merkmals
h_n :	Relative Häufigkeit eines Merkmals
γ :	Sicherheit; Vertrauenszahl
k:	ergibt sich aus γ :

γ	90%	95%	99%
k	1,64	1,96	2,58

Näherung für das Vertrauensintervall: $\left[h_n - k \cdot \frac{\sqrt{n \cdot h_n \cdot (1-h_n)}}{n}; h_n + k \cdot \frac{\sqrt{n \cdot h_n \cdot (1-h_n)}}{n}\right]$

Mit $\sigma_n = \sqrt{n \cdot h_n \cdot (1-h_n)}$ und $\varepsilon = k \cdot \frac{\sigma_n}{n}$ schreibt man kurz $[h_n - \varepsilon; h_n + \varepsilon]$.

Andere Schreibweise (Anlehnung an die exakte Lösung unten):

Näherungsformel: $\left[h_n - k \cdot \sqrt{\frac{h_n \cdot (1-h_n)}{n}}; h_n + k \cdot \sqrt{\frac{h_n \cdot (1-h_n)}{n}}\right]$

Beispiel

Die Partei CFP gibt eine Umfrage in Auftrag. Eine repräsentative Stichprobe von 1000 Wählern ergibt 370 Wähler der CFP. Bestimmen Sie das 95%-Vertrauensintervall für den unbekannten Anteil p der CFP-Wähler in der Bevölkerung.

Näherungslösung:

$n = 1000$, $h_n = \frac{370}{1000} = 0{,}37$,

Untere Intervallgrenze: $0{,}37 - 1{,}96\sqrt{\frac{037 \cdot (1-0{,}37)}{1000}}$ = 0,37 - 0,02992

Obere Intervallgrenze: : $0{,}37 + 1{,}96\sqrt{\frac{037 \cdot (1-0{,}37)}{1000}}$ = 0,37 + 0,02992

Vertrauensintervall $[0{,}37 - 0{,}02992; 0{,}37 + 0{,}02992] = [0{,}34008; 0{,}39992]$

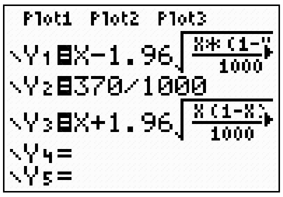
Interpretation im Sachzusammenhang:

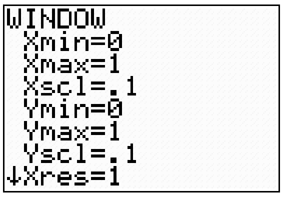
Die Partei CFP kann bei der Wahl mit einer Sicherheit von 95 Prozent ein Ergebnis zwischen 34,01% und 39,99% der Wählerstimmen erwarten (eine Spanne von fast 6%).

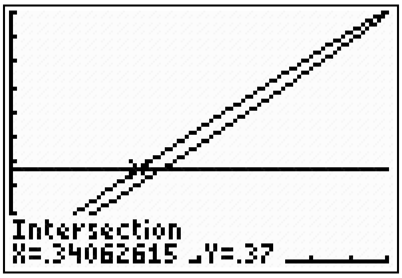
Die **exakte Lösung** besteht in der Eingabe der Ungleichungskette (p durch x ersetzen!)

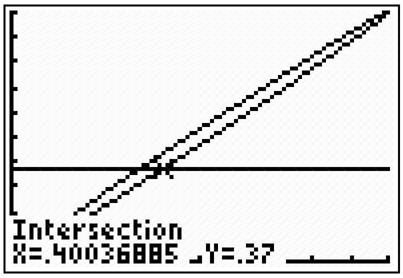
$p - k\sqrt{\frac{p \cdot (1-p)}{n}} \leq h_n \leq p + k\sqrt{\frac{p \cdot (1-p)}{n}}$ als drei Funktionen Y1 ≤ Y2 ≤ Y3 in den GTR

und ergibt das Intervall $[0{,}34063; 0{,}40037]$. Bild links: intersect(Y2,Y3), rechts: intersect(Y1,Y2)

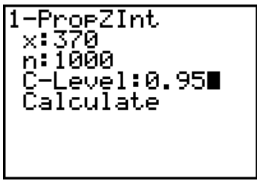


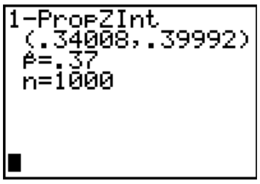






Näherungslösung mit dem GTR:
Mit dem GTR wählt man den Test STAT→TESTS→A:1-propZInt…





und erhält das gleiche Resultat wie mit der Näherungsformel oben.

Berechnung des Stichprobenumfangs:

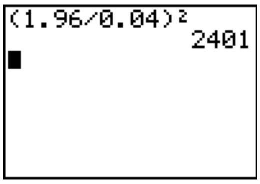
ε gibt die halbe Länge des Intervalls an. Wenn man eine Wahrscheinlichkeit auf $\pm$ 0,5 % genau haben möchte, so ist ε=0,005 und die gesamte Intervalllänge 2ε = 0,01.

Formeln für den Stichprobenumfang n:

Stichprobenumfang**:** $$\boldsymbol{n > \left(\frac{k}{2\varepsilon}\right)^2}$$

bei bekannter Größenordnung von p: $$\boldsymbol{n > \left(\frac{k}{\varepsilon}\right)^2 \cdot p \cdot (1-p)}$$

Beispiel: Zum Vertrauensniveau $\gamma = 0{,}95$ und einer gewünschten Länge 0,04 (4%) des Vertrauensintervalls soll die Anzahl der zu befragenden Personen bestimmt werden:



$$k = 1{,}96,\ \varepsilon = \frac{0{,}04}{2} = 0{,}02 \ \Rightarrow\ n > \left(\frac{k}{2\varepsilon}\right)^2 = \left(\frac{1{,}96}{0{,}04}\right)^2 = 2401$$

Es müssen mindestens 2402 Personen befragt werden, um das Wahlergebnis mit einer Genauigkeit von 4% vorherzusagen (bei einer Sicherheit von 95%).

Hier wird die unbekannte Wahrscheinlichkeit p durch p = 0,5 abgeschätzt, das gibt den größtmöglichen Wert des Terms p(1-p). Ist im Beispiel oben bekannt, dass p maximal 10%, also 0,1 ist, so kann man die zweite Formel benutzen und kommt zu einem wesentlich geringeren Stichprobenumfang:

$$n > \left(\frac{k}{\varepsilon}\right)^2 \cdot p \cdot (1-p)$$
$$n > \left(\frac{1{,}96}{0{,}02}\right)^2 \cdot 0{,}1 \cdot (1-0{,}1)$$
$$n > 864{,}36$$

Somit muss man jetzt nur noch 865 Personen befragen.

Hn 03/2012