Bestimmung von Vertrauensintervallen (Konfidenzintervallen) bei unbekannten Wahrscheinlichkeiten

Gegeben sein müssen *n, X* oder *h* einer Stichprobe und das Vertrauensniveau γ .

n: Umfang der Stichprobe; *X:* Absolute Häufigkeit eines Merkmals, *h*: Relative Häufigkeit eines Merkmals

γ : Sicherheit; Vertrauenszahl . *h* ergibt sich aus *n*: $h = \frac{X}{n}$. *k* ergibt sich aus γ :

γ	90%	95%	99%
k	1,64	1,96	2,58

Mit der Sicherheit γ liegt die unbekannte Wahrscheinlichkeit p im Vertrauensintervall.

Grafische Bestimmung der exakten Intervallgrenzen

p muss aus der Ungleichungskette

$$\boldsymbol{p - k\sqrt{\frac{p\cdot(1-p)}{n}} \leq h \leq p + k\sqrt{\frac{p\cdot(1-p)}{n}}}$$

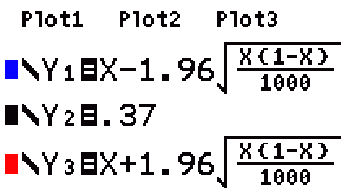
bestimmt werden.

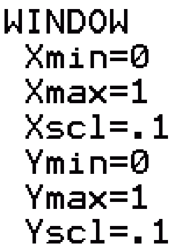
Dazu werden die drei Terme in den y-Editor des GTR eingegeben (p entspricht x im GTR) und die Schnittstellen von **y3** und y2 (untere Intervallgrenze) bzw. **y1** und y2 (obere Intervallgrenze) bestimmt.

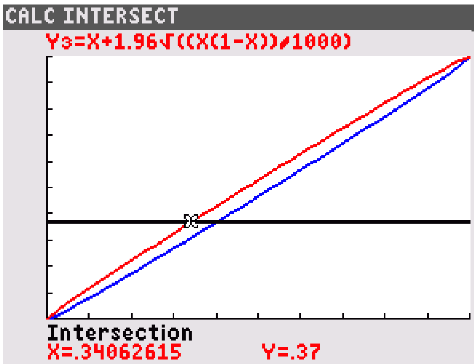
Beispiel

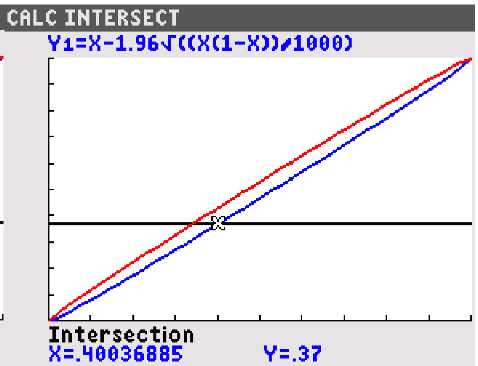
Die Partei CFP gibt eine Umfrage in Auftrag. Eine repräsentative Stichprobe von 1000 Wählern ergibt 370 Wähler der CFP. Bestimmen Sie das 95%-Vertrauensintervall für den unbekannten Anteil p der CFP-Wähler in der Bevölkerung.

$n = 1000; h = \frac{370}{1000} = 0{,}37; k = 1{,}96$









Mit 95% Sicherheit liegt die unbekannte Wahrscheinlichkeit p im Konfidenzintervall [0,34063; 0,40037].

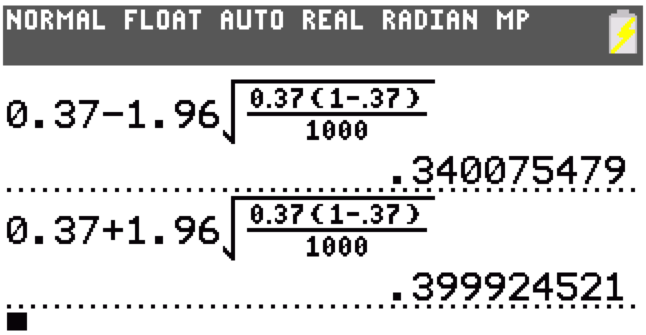
Interpretation im Sachzusammenhang:

Die Partei CFP kann bei der Wahl mit einer Sicherheit von 95 Prozent ein Ergebnis zwischen 34,06 % und 40,04 % der Wählerstimmen erwarten (eine Spanne von fast 6%).

Vermutlich unterscheidet sich die relative Häufigkeit h in der Stichprobe von dem Anteil p in der Gesamtheit nur wenig. Daher genügt es oft, mit folgender Näherungsformel zu rechnen:

Direkte Berechnung von p mit einer Näherungsformel

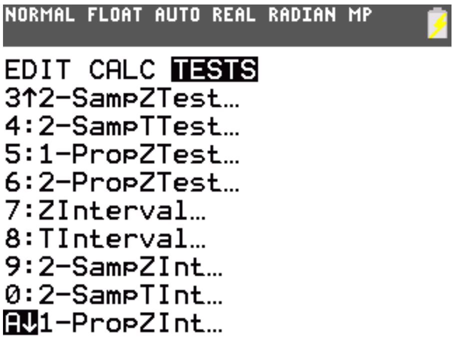
$$\left[h - k\cdot\sqrt{\frac{h\cdot(1-h)}{n}}; h + k\cdot\sqrt{\frac{h\cdot(1-h)}{n}}\right]$$

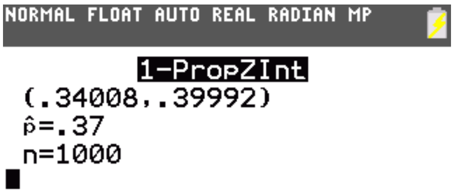


Mit den Werten aus dem Beispiel oben ($n = 1000$; $h = \frac{370}{1000} = 0{,}37; k = 1{,}96$) ergibt sich jetzt die Näherung für das Vertrauensintervall [0,34008; 0,39992].

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:

Aus der Näherungsformel oben lässt sich auch der benötigte Stichprobenumfang n zu einer gewünschten maximalen Intervalllänge berechnen. Die Fehlertoleranz $\varepsilon = k \cdot \sqrt{\frac{h\cdot(1-h)}{n}}$ 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.

Bei bekannter Größenordnung von h, z.B. aus einer vorherigen Umfrage, gilt:

Stichprobenumfang bei bekanntem h: $\boldsymbol{n > \left(\frac{k}{\varepsilon}\right)^2 \cdot h \cdot (1-h)}$

Beispiel: Zum Vertrauensniveau $\gamma = 0{,}95$ und einer gewünschten Fehlertoleranz des Vertrauensintervalls von $\pm$2 % soll die Anzahl n der zu befragenden Personen bestimmt werden. Aus früheren Untersuchungen ist bekannt, dass h höchstens 10% beträgt.

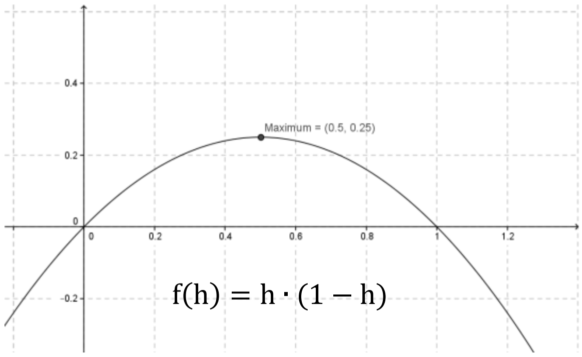
$k = 1{,}96,\ \varepsilon = 0{,}02, h = 0{,}1.$ (Beachte die Anmerkung unten zur Wahl von h.)

Damit ergibt sich für n: $n > \left(\frac{1{,}96}{0{,}02}\right)^2 \cdot 0{,}1 \cdot (1-0{,}1) \iff n > 864{,}36$

Es müssen mindestens 865 Personen befragt werden, um das Wahlergebnis mit einer Genauigkeit von $\pm$2 % vorherzusagen (bei einer Sicherheit von 95%).

Oft ist aber keine Größenordnung von h bekannt, so dass man den Faktor $h \cdot (1-h)$ durch den größtmöglichen Wert, den er annehmen kann, abschätzt. Dieser wird für h=0,5 erreicht, wie der Graph rechts zeigt. Ersetzt man also $h \cdot (1-h)$ durch 0,5 · 0,5 = 0,25 = $\frac{1}{4}$, so ergibt sich wegen $\left(\frac{k}{\varepsilon}\right)^2 \cdot h \cdot (1-h) = \left(\frac{k}{\varepsilon}\right)^2 \cdot \frac{1}{4} = \left(\frac{k}{2\varepsilon}\right)^2$ der

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



Für das Beispiel oben ergibt sich damit: $n > \left(\frac{k}{2\varepsilon}\right)^2 = \left(\frac{1{,}96}{2\cdot 0{,}02}\right)^2 = 2401$

Es müssen mindestens 2402 Personen befragt werden, um das Wahlergebnis mit einer Genauigkeit von $\pm$2 % vorherzusagen (bei einer Sicherheit von 95%).

Der Faktor h führte in der oberen Formel zu h(1-h)=0,1·0,9=0,09, in der unteren Formel beträgt er mit 0,25 fast das Dreifache.

Anmerkung zur Wahl von h: Wie der Graph oben zeigt, wird $\mathrm{h} \cdot (1-\mathrm{h})$ für h=0,5 maximal, so dass in der ersten Formel für h derjenige Wert gewählt werden sollte, der am dichtesten an 0,5 liegt. Wenn man weiß, dass h zwischen 0,3 und 0,4 liegt, sollte man h=0,4 einsetzen, bei möglichen Werten für h zwischen 0,75 und 0,85 wählt man h=0,75.

Hn 05/2013