

Konfidenzintervall für den Anteilswert π am Beispiel einer Meinungsumfrage

Nach einer Meinungsumfrage der Wochenzeitung „Bezirksblatt“ vom März 2005, ein halbes Jahr vor den Landtagswahlen im Burgenland (Stichprobenumfang $n = 402$) ergab folgende Umfrageergebnisse (siehe Bild rechts): Bei sofortigen Landtagswahlen würde die SPÖ 55%, die ÖVP 35%, die FPÖ 5% und die Grünen 5% der Stimmen erhalten.

Wir wollen eine „Hochrechnung“ für den Wahlausgang durchführen und berechnen Konfidenzintervalle für den Stimmenanteil der vier Parteien.

Weiters berechnen wir die Schwankungsbreite e für alle berechneten Konfidenzintervalle und vergleichen die berechneten Konfidenzintervalle mit den tatsächlichen Ergebnissen der Wahl: SPÖ: 52.2% ; ÖVP: 36.3% ; FPÖ: 5.8%; Grüne: 5.2%.

Bei Meinungsumfragen ist es üblich mit einem Konfidenzniveau $c = 95\%$ zu rechnen.

Die Anzahl X der Wähler der Partei A in einer Stichprobe vom Umfang n ist binomialverteilt mit dem **bekanntem** Parameter n und einem **unbekanntem** π ($\mu = n \cdot \pi$).

- x Wähler von Partei A in der Stichprobe
- $n = 402$ Umfang der Stichprobe
- $c = 95\%$ Niveau des Konfidenzintervalls
- $\hat{p} = \frac{x}{402}$ Wähleranteil (bekannt) von Partei A in der Stichprobe

Für das Konfidenzintervall des Anteilswertes π einer Grundgesamtheit gilt für $0,3 < \hat{p} < 0,7$ die Näherungsformel:

$$\hat{p} - z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \leq \pi \leq \hat{p} + z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \quad \text{bild2.}$$

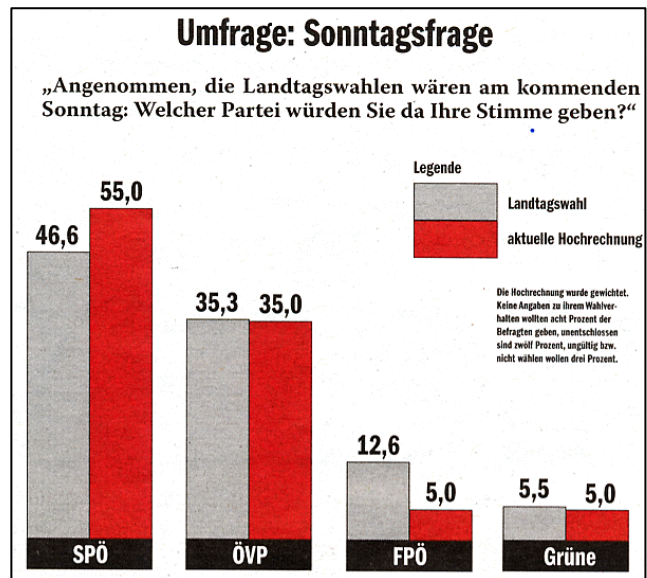
mit $z = \Phi^{-1}\left(\frac{c+1}{2}\right)$ (für $c = 0,95$ gilt $z \approx 1,96$)

oder

$$[\hat{p} - e ; \hat{p} + e] \quad \text{mit } e = z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \quad \text{bild3.gif}$$

e = Fehlertoleranz; Schwankungsbreite

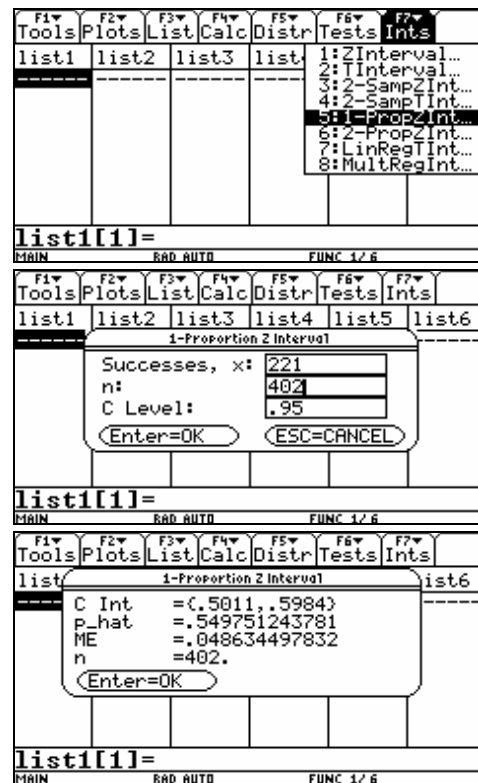
Alle TI-Produkte berechnen Konfidenzintervalle für Anteilswerte nach dieser Formel im **1-PropZInt**.



Wir berechnen zunächst ein Konfidenzintervall für die SPÖ auf „klassische“ Weise mit dem V200, mit dem TI-84 Plus, mit TI-InterActive! und mit TI-Nspire.

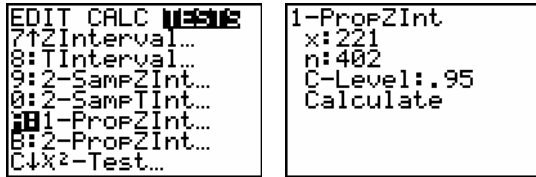
$$\hat{p} = 0,55 = \frac{x}{402} \Rightarrow x \approx 221$$

Arbeiten Sie beim **V200** mit der Listenstatistik.

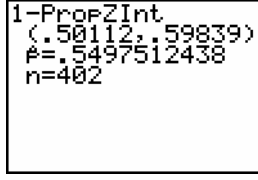


$p_hat = \hat{p}$ Wähleranteil in Stichprobe
 $ME = e$ Schwankungsbreite

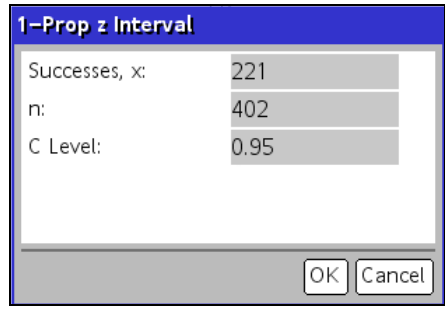
TI-84 Plus



screen05/06

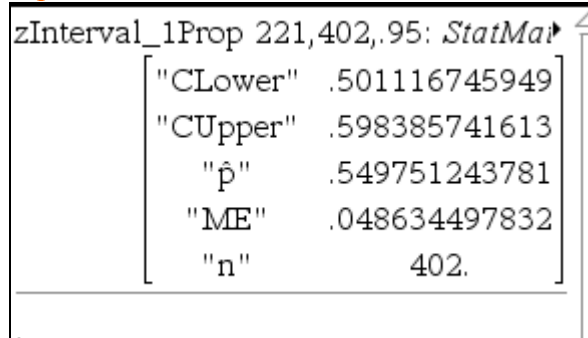


screen07



nspire2

nspire3



TI-InterActive!

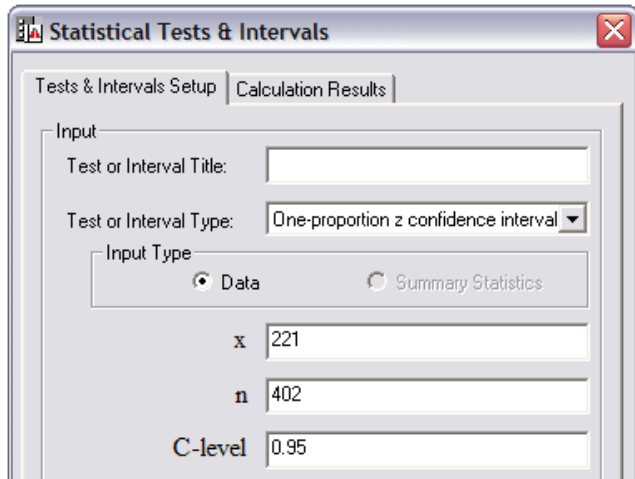


Bild4.gif

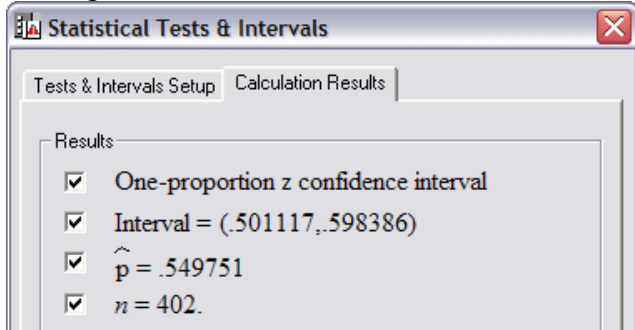
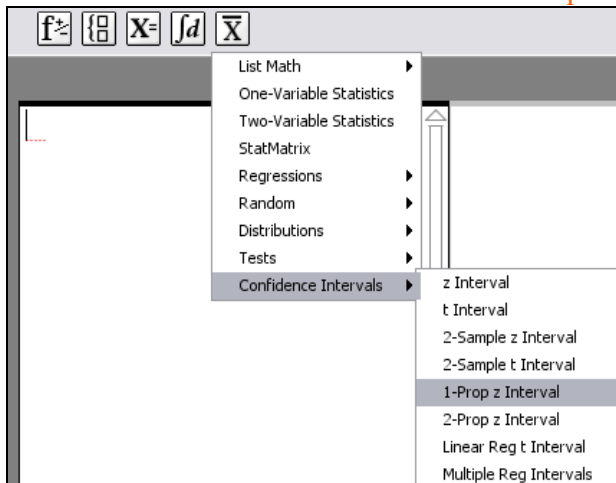


Bild5.gif

TI-Nspire CAS

Arbeiten Sie im Calculator um Konfidenzintervalle direkt zu berechnen. nspire1



Interpretation der Ergebnisse:

Mit einer Wahrscheinlichkeit von 95% liegt der Wähleranteil der SPÖ zwischen 50,1% und 59,8% der abgegebenen Stimmen.

Konfidenzintervall SPÖ:

[0,5011 ; 0,5984] c = 0,95

Der Stimmenanteil der SPÖ bei Wahl war 52,2% und liegt somit im Konfidenzintervall.

Auf analoge Weise können die Konfidenzintervalle auch für die anderen Parteien berechnet werden.

Interessanter ist es allerdings die Konfidenzintervalle grafisch als „Konfidenzellipse“ darzustellen.

Grafische Darstellung von Konfidenzintervallen

Für die Veranschaulichung von Konfidenzintervallen werden zuerst die Gleichungen der begrenzenden Funktionen definiert. Diese Funktionen werden dann für konkrete Werte von n und \hat{p} dargestellt.

In diesem Beispiel gilt: p steht für \hat{p} .

Zunächst werden die begrenzenden Funktionen $\text{piU}(n,p)$ und $\text{piO}(n,p)$ sowie die Funktion für die Schwankungsbreite $\text{e}(n,p)$ definiert.

Das vertikale Geradenstück zwischen den Ellipsenhälften stellt das Konfidenzintervall für den Wert \hat{p} (hier p) dar.

In TIII definieren Sie zunächst Werte und Funktionen mit **MathBoxes** im Arbeitsblatt.

$c := 0.95 = .95$	Konfidenzniveau
$z := \text{invNorm}\left(\frac{c+1}{2}\right) = 1.95996$	
$e(n, p) := z \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$	Schwankungsbreite
$\text{piU}(n, p) := p - e(n, p)$	untere Intervallgrenze
$\text{piO}(n, p) := p + e(n, p)$	obere Intervallgrenze

Bild8.gif

Geben Sie die Funktionsbezeichnungen in den **Funktionseditor** ein.

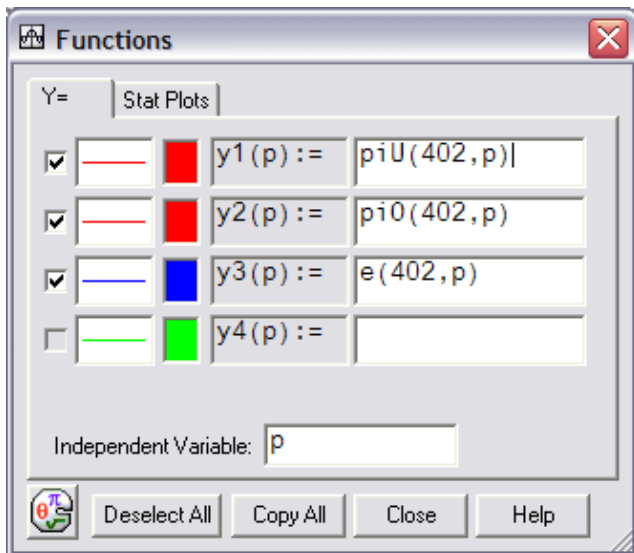


Bild7.gif

Bei den Handhelds geben Sie die Funktionsgleichungen direkt im Funktionseditor ein.

F1 F2 F3 F4 F5 F6

Zoom Edit All Style

Plot 1:

$y1 = 1.96 \cdot \sqrt{\frac{x \cdot (1-x)}{402}}$

$y2 = x - y1(x)$

$y3 = x + y1(x)$

$y4 =$

$y5 =$

$y6 =$

$y7 =$

$y4(x) =$

screen10

Screen12

Screen15

In N-Spire geben Sie die Definitionen der Werte und Funktionen zunächst in den **Calculator** ein.

$c := .95$.95
$z := \text{invNorm}\left(\frac{c+1}{2}\right)$	1.95996398599
$e(n,p) := z \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$	Done
$\text{piu}(n,p) := p - e(n,p)$	Done
$\text{pio}(n,p) := p + e(n,p)$	Done
$\text{piu}(402, .55)$.501367939973
$\text{pio}(402, .55)$.598632060027

Nspire4

Im nächsten Schritt werden die Grafen der soeben definierten Funktionen in einem geeigneten Koordinatensystem für den vorliegenden Stichprobenumfang $n = 402$ dargestellt.

Mit Hilfe der TRACE-Funktion werden an der Stelle $p = 0,55$ die Funktionswerte ermittelt und die Werte angezeigt. Die y-Koordinaten der berechneten Punkte ergeben die Grenzen des Konfidenzintervalls.

Bild9.gif

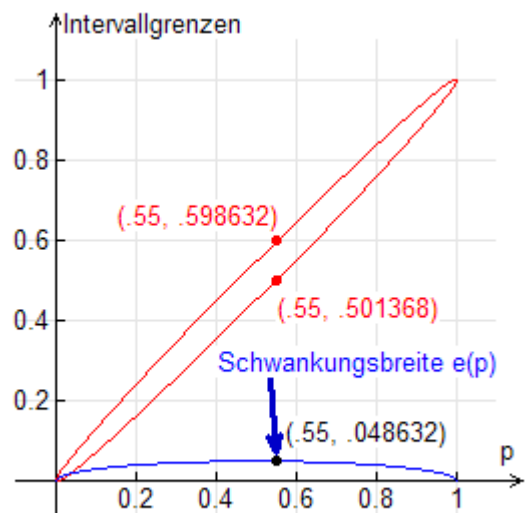
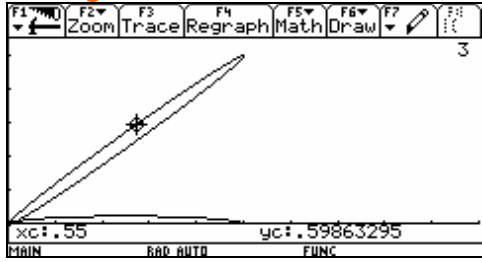
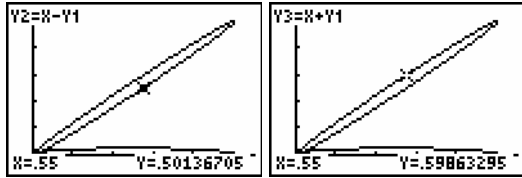


Bild9.gif

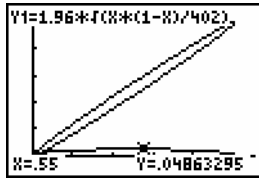


Screen11



Screen13

Screen13a



screen14

Aus den Grafiken wird für $p = 0,55$ das Konfidenzintervall $[0,501 ; 0,599]$ und die Schwankungsbreite $e = 0,0486$ (5%) abgelesen. In TII können die Konfidenzintervalle optisch noch besser aufbereitet werden:

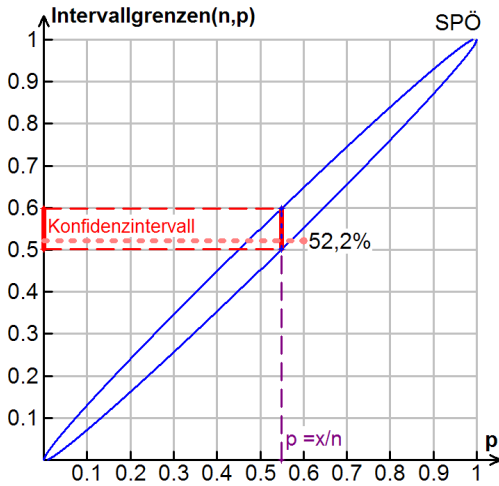


Bild10.gif

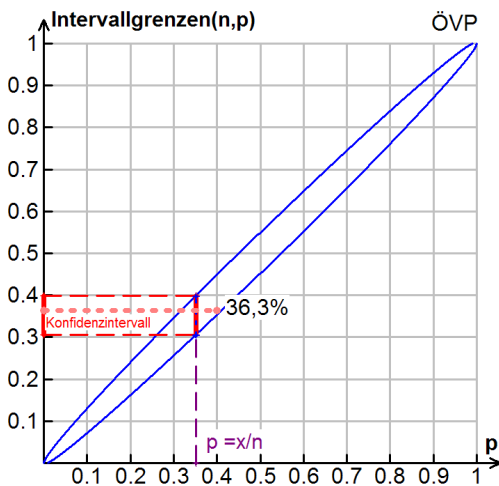


Bild11.gif

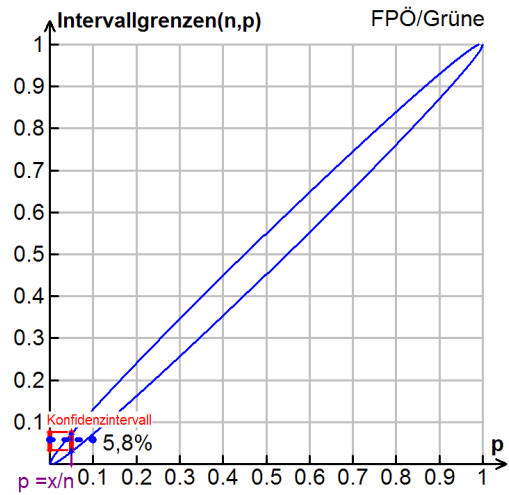
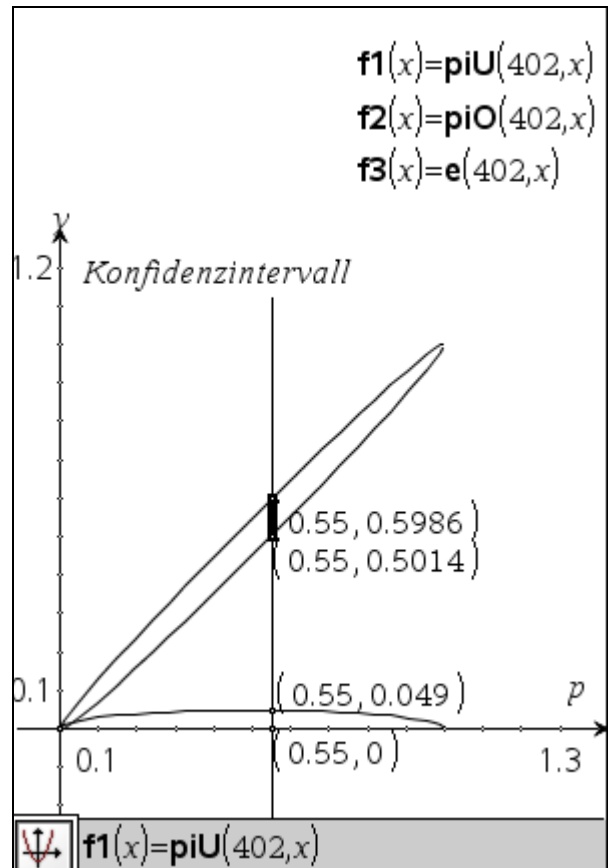


Bild12.gif

Bei der Darstellung des „Konfidenzintervalls“ mit dem N-Spire haben Sie sogar den Vorteil der *Interaktivität der Grafik*. Durch Ziehen der vertikalen Geraden, kann \hat{p} (hier p) und damit das Konfidenzintervall verändert werden.



Nspire5

Man erhält folgende weitere Konfidenzintervalle:

ÖVP: $[0,3041 ; 0,3974]$

FPÖ/Grüne: $[0,0285 ; 0,071]$

Alle Konfidenzintervalle enthalten den tatsächlichen Stimmenanteil der jeweiligen Partei nach der Wahl. Die „Hochrechnung“ hat also richtige Vorhersagen geliefert.

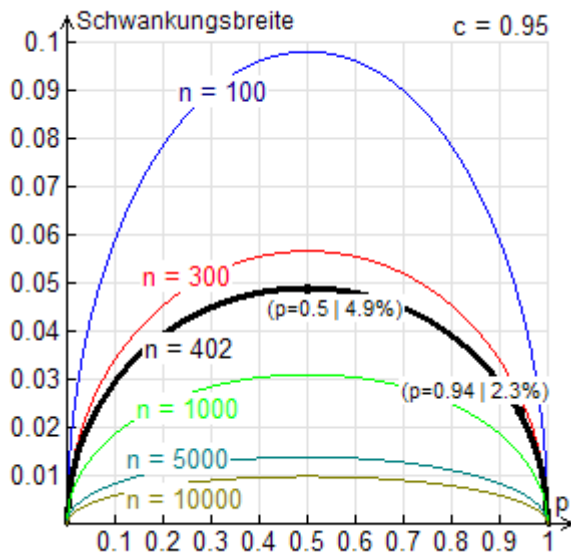
Die Berechnungen zeigen weiters, dass die Schwankungsbreite e sehr stark vom Anteilswert \hat{p} (in der Grafik p) und vom Umfang n der Stichprobe abhängt.

Man kann allgemein zeigen, dass die maximale Schwankungsbreite bei $\hat{p} = 0.5$ auftritt.

Aus der Grafik ($c = 0,95$) erkennt man weiters, dass beispielsweise Umfragen mit $n = 100$ eine Schwankungsbreite von fast 10% (Breite des Konfidenzintervalls $\approx 20\%$) haben und somit für Schätzungen unbrauchbar sind.

Sinnvoll sind erst Umfragen mit einem Stichprobenumfang von $n > 300$.

Für den Stichprobenumfang $n = 402$ ergibt sich eine maximale Schwankungsbreite von ca. 4,9%.



Überlegungen zum Umfang einer Stichprobe für den Anteilswert π

Geht man vom Konfidenzintervall mit

$$e = z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

aus und berechnen Sie daraus durch elementare Umformungen den

$$\text{Stichprobenumfang } n = \frac{z^2 \cdot \hat{p} \cdot (1 - \hat{p})}{e^2} \text{ mit}$$

$$z = \Phi^{-1}\left(\frac{c+1}{2}\right).$$

Das Problem ist allgemein, dass bei der Berechnung des Stichprobenumfangs der Anteilswert π der Grundgesamtheit nicht bekannt ist. Man kann dieses Problem dadurch lösen, indem man eine **Vorerhebung** mit kleinem Stichprobenumfang macht und daraus

$$\hat{p} = \frac{x}{n} \approx \pi \text{ ermittelt.}$$

Fortsetzung der Berechnungen:

Die Blitzumfrage unter 402 Wahlberechtigten ergab, dass bei der nächsten Wahl $x = 221$ von den fragten Personen die SPÖ wählen würden. Der Rest der Stimmen verteilt sich auf andere Parteien. Es wurde ein 95% Konfidenzintervall für den unbekanntem Wähleranteil π der SPÖ unter allen Wahlberechtigten ermittelt: $[0,501 ; 0,599]$. Die Schwankungsbreite ist allerdings ca. 5%.

Dem Kandidaten der SPÖ ist diese Schätzung zu unpräzise. Er möchte ein Ergebnis auf $\pm 1\%$ (dh. $e = 0.01$) genau. Wie groß muss dafür der Stichprobenumfang n sein ($c = 0.95$)?

$$\hat{p} = 221/402 = 0.549 \approx 55\%$$

$$n = \frac{z^2 \cdot \hat{p} \cdot (1 - \hat{p})}{e^2} = \frac{1.96^2 \cdot 0.55 \cdot (1 - 0.55)}{0.01^2} =$$

$$9507.96; \text{ Stichprobenumfang } n = 9508$$

Die Blitzumfrage mit $n = 402$ kann als **Vorerhebung** aufgefasst werden.

Daraus wird $\hat{p} = 55\%$ berechnet.

Alternativ kann man den „worst case“ annehmen:

Dabei wird für $\hat{p} = 0.5$ verwendet, da $0.5 \cdot (1 - 0.5) = 0.25$ ein Maximum der Funktion mit $f(p) = p \cdot (1 - p)$ ist.

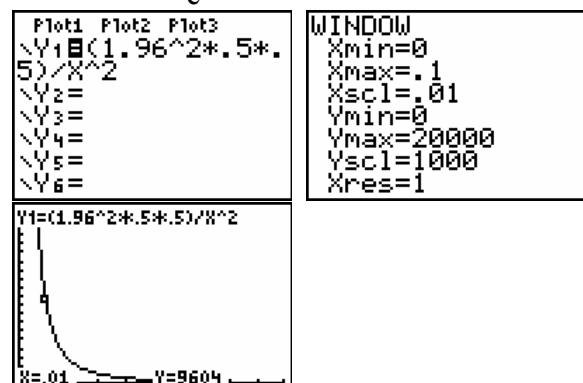
$$n = \frac{1.96^2 \cdot 0.5 \cdot (1 - 0.5)}{0.01^2} = 9604; \text{ Der Stich-$$

probenumfang ist mit $n = 9604$ größer als im vorigen Fall mit einer Vorerhebung. Dies ist in der Praxis aus Kostengründen interessant.

Grafische Darstellung des Stichprobenumfanges n , in Abhängigkeit von der Fehlertoleranz e .

$c = 0.95$ und $\hat{p} = 0.5$

$$n(e) = \frac{1.96^2 \cdot \hat{p} \cdot (1 - \hat{p})}{e^2}$$



Friedrich Tinhof
Bundeshandelsakademie Eisenstadt
T3 Österreich fritz.tinhof@t3oesterreich.at
Literatur: Mathematik Band V für HAK
Trauner Verlag