Schätzbereich und Konfidenzintervall

In der beurteilenden Statistik wird entweder von der Grundmenge auf eine Teilmenge geschlossen (ein Anteil geschätzt) oder von einer Teilmenge auf die Gesamtmenge geschlossen (ein Konfidenzintervall des Anteils berechnet).

Damit das nicht zu abstrakt wird, kommt gleich ein Beispiel:

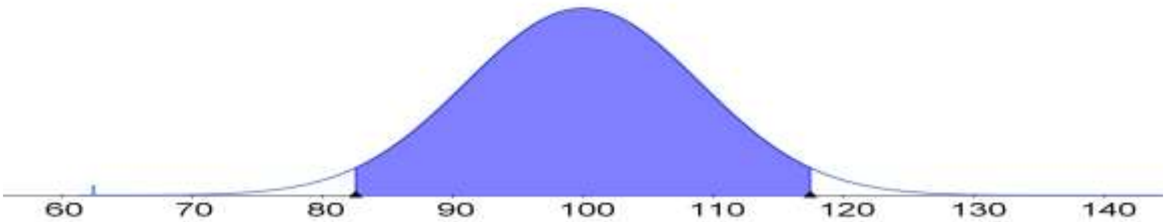
Beispiel 1: Schätzbereich bestimmen ⇨ **Von GROSS auf KLEIN schätzen**	**Beispiel 2: Konfidenzintervall bestimmen** ⇨ **Von KLEIN auf GROSS schätzen**
In einer Gemeinde von einer Million Einwohnern wählen 200 000 Personen die A−Partei (=20%) Wie groß wird bei einer Nachwahlbefragung von 500 Personen der Anteil der A−Partei−Wähler sein? Es werden wahrscheinlich auch 20% der Wähler sein – das wäre unsere Schätzung. Sicherheitshalber wird dafür ein Intervall angegeben, in dem mit einer Sicherheit von z.B. 95% der Anteil der A−Wähler liegt – Das ist der **SCHÄTZBEREICH**	Vor der Wahl der Parteien wird eine Wahlumfrage gemacht. Bei 1000 Wählern wollen 200 die A−Partei wählen (=20%) Wir wollen wissen, wie groß der Anteil der A−Wähler bei der Nationalratswahl sein wird. Es werden wahrscheinlich auch 20% der Wähler die A−Partei wählen. Sicherheitshalber wird dafür ein Intervall angegeben, in dem mit einer Sicherheit von z.B. 95% der Anteil der A−Wähler liegt – Das ist das **KONFIDENZINTERVALL**

Lösung 1) Wir wollen nun den **Schätzbereich** des 1.Beispiels berechnen:

Wir nehmen an, dass jeder Wähler mit der Einzelwahrscheinlichkeit von 20% die A−Partei gewählt hat. Es waren 500 Personen, die befragt wurden. Wie schaut die Verteilung der Anzahl der A−Wähler aus?

Bei dieser Fragestellung kann man von einer Binomialverteilung ausgehen, da die Fragestellung nur 2 Möglichkeiten kennt (A−Wähler oder Nicht−A−Wähler) und die Einzel−Wahrscheinlichkeit für alle Befragten gleich ist.

Wir fragen nun zuerst nach dem Erwartungswert der Binomialverteilung $E(X) = \mu = n \cdot p = 500 \cdot 0{,}2 =$ **100**
Und dann brauchen wir noch die Standardabweichung $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{500 \cdot 0{,}2 \cdot 0{,}8} =$ **8,9**
Damit können wir die Binomialverteilung durch die Normalverteilung approximieren:



Hier wurde schon ein symmetrisches Intervall mit 95% eingezeichnet
Wie kann man die Grenzen davon ausrechnen?

Dazu muss man überlegen, welche Wahrscheinlichkeit für den Bereich bis zur rechten Grenze gilt –
Dazu muss man die 95% und die Hälfte des Restes auf 100% (=2,5%) addieren – ergibt 97,5%

Schaut man in der Normalverteilungstabelle bei $\Phi(z) = 0{,}975$ nach, so steht da für z ➔ 1,96.

Dann ist die linke Grenze $x = \mu - z \cdot \sigma = 100 - 1{,}96 \cdot 8{,}9 =$ **82,6**

Und die rechte Grenze ist $x = \mu + z \cdot \sigma = 100 - 1{,}96 \cdot 8{,}9 =$ **117,4**

Der SCHÄTZBEREICH mit 95% Sicherheit für die Anzahl der A–Wähler ist [82,6; 117,4]

FORMEL

Jetzt wollen wir eine Formel für die schnellere Berechnung des Anteils h der A–Wähler allgemein entwickeln:

Wenn H die absolute Anzahl der A–Wähler ist, ist $h = \frac{H}{n}$ *(=0,20) der relative Anteil der A–Wähler.*

Wir wissen, dass die Wahrscheinlichkeit des symmetrischen Intervalls um den Mittelwert H mit Sicherheits–Wahrscheinlichkeit γ *so angeschrieben wird:*

$P(\mu - z \cdot \sigma < H < \mu + z \cdot \sigma) = \gamma$

Ersetzt man μ *durch* $n \cdot p$ *und* σ *durch* $\sqrt{n \cdot p \cdot (1-p)}$*, so ergibt sich die Formel:*

$P(n \cdot p - z \cdot \sqrt{n \cdot p \cdot (1-p)} < H < n \cdot p + z \cdot \sqrt{n \cdot p \cdot (1-p)}) = \gamma$

Dividiert man nun durch n, so ergibt sich:

$P(p - z \cdot \sqrt{\frac{p \cdot (1-p)}{n}} < h < p + z \cdot \sqrt{\frac{p \cdot (1-p)}{n}}) = \gamma$

Der γ–Schätzbereich für h ist also $[p - z \cdot \sqrt{\frac{p \cdot (1-p)}{n}}\ ;\ p + z \cdot \sqrt{\frac{p \cdot (1-p)}{n}}]$ mit $\Phi(z) = \frac{1+\gamma}{2}$

Beispiel 1a) Erfahrungsgemäß reagieren ca. 60% der Migränepatienten auf Entspannungsübungen positiv. Bei einer Studie erhalten 100 zufällig ausgewählte Migränepatienten diese Therapie. Wie viel Prozent positiv reagierender Patienten sind zu erwarten? Gib einen 95%-Schätzbereich an!

Lösung 1a): Wir stellen fest: n=100, p=0,6 γ=0,95

Damit suchen wir z aus der Normalverteilungstabelle mit $\Phi(z) = \frac{1+\gamma}{2} = 0{,}975$ ➔ z = 1,96

Das ergibt das Schätzintervall $[0{,}6 - 1{,}96 \cdot \sqrt{\frac{0{,}6 \cdot (1-0{,}6)}{100}}\ ;\ 0{,}6 + 1{,}96 \cdot \sqrt{\frac{0{,}6 \cdot (1-0{,}6)}{100}}] =$ **[0,504; 0,696]**

Netterweise ist die obige Formel auch die gleiche Formel, die wir zur Berechnung eines Konfidenzintervalls brauchen, nur wird hier p mit h vertauscht. Man kennt also die statistische Größe h (Anteil der A–Wähler in einer Stichprobe) und berechnet das Intervall für p (Wahrscheinlichkeit der A–Wähler)

Das γ–Konfidenzintervall für p ist also $[h - z \cdot \sqrt{\frac{h \cdot (1-h)}{n}}\ ;\ h + z \cdot \sqrt{\frac{h \cdot (1-h)}{n}}]$ mit $\Phi(z) = \frac{1+\gamma}{2}$

Lösung 2) für Beispiel 2: Das 95%-Konfidenzintervall wird gesucht für:

Vor der Wahl der Parteien wird eine Wahlumfrage gemacht. Bei 1000 Wählern wollen 200 die A–Partei wählen (=20%)

Das Konfidenzintervall ergibt sich, wenn zuerst die Buchstaben fixiert werden:

⇨ n=1000 h=0,2 γ=0,95 und z = 1,96 (wie vorher)

$[h - z \cdot \sqrt{\frac{h \cdot (1-h)}{n}}\ ;\ h + z \cdot \sqrt{\frac{h \cdot (1-h)}{n}}] = [0{,}2 - 1{,}96 \cdot \sqrt{\frac{0{,}2 \cdot (1-0{,}2)}{1000}}\ ;\ 0{,}2 + 1{,}96 \cdot \sqrt{\frac{0{,}2 \cdot (1-0{,}2)}{1000}}] =$ **[0,175 ; 0,225]**

➔ **Der relative Anteil der Wähler der A–Partei schwankt zwischen 17,5% und 22,5%**

Übungen:

Für das Konfidenzintervall mit γ =	90%	95%	99%
Ist z =	1,645	1,96	2,58

1) Der Hersteller eines Massenartikels geht davon aus, dass ca. 5% der Produktion fehlerhaft sind. Zur Qualitätskontrolle wird eine Stichprobe mit 150 Stück gemacht. Wie viel Prozent fehlerhafte Ware ist in der Stichprobe voraussichtlich enthalten? Gib einen 95%-Schätzbereich an!
2) In einem Betrieb haben 8% der Stoffballen einen Materialfehler und sind daher von der Weiterverarbeitung ungeeignet. Eine Stichprobe vom Umfang n=50 wird aus einer Lieferung entnommen.
   a) Berechne den 99%-Schätzbereich für den Anteil (%) der fehlerhaften Stoffballen in der Stichprobe. Gib auch die entsprechende absolute Anzahl der betroffenen Stoffballen an.
   b) Erkläre, wie sich die Bereichsgrenzen verändern, wenn eine Wahrscheinlichkeit von 95% gewählt wird
3) Eine Firma produziert USB−Sticks. Erfahrungsgemäß sind 3% der erzeugten Sticks fehlerhaft.
   a) Ermittle mit wie vielen fehlerhaften Sticks man bei 90% Sicherheit bei einer Stichprobe mit 60 Sticks rechnen kann
   b) Erkläre, wie sich die Bereichsgrenzen der Prozentwerte ändern, wenn der Umfang der Stichprobe erhöht wird
   c) Erkläre, wie sich die Bereichsgrenzen der Prozentwerte ändern, wenn der Prozentsatz der fehlerhaften USB−Sticks höher ist.

---

4) Vor einer Parlamentswahl ergibt eine Wahlumfrage mit 600 Personen, dass 150 Personen davon die Partei Z wählen wollen. Gib ein a) 95%- Konfidenzintervall b) 99%-Konfidenzintervall für den unbekannten relativen Anteil der Z−Partei−WählerInnen an!
5) Bei einer Umfrage wurde festgestellt, dass sich unter 320 Jugendlichen 12% Linkshänder befanden. Gib ein Konfidenzintervall für den unbekannten relativen Anteil der Linkshänder in diesem Schulsprengel an – mit der Sicherheit a) γ=0,95 b) γ=0,99
6) Bei einer Umfrage unter 100 Passagieren von Interrail gaben 72 Personen an, dann sie mit Interrail zufrieden sind. Berechne ein a) 95%-Konfidenzintervall b) 90%-Konfidenzintervall für den unbekannten relativen Anteil aller zufriedenen Fahrgäste von Interrail an.
7) Bei einer Überprüfung von 300 Platinen findet man 5 fehlerhafte. Ermittle ein 95%-Konfidenzintervall für den Anteil an fehlerhaften Platinen.

---

8) Bei einer Befragung von 1000 zufällig ausgewählten Männern gaben 30% an, sich elektrisch zu rasieren. Aufgrund dieses Ergebnisses wurde ein Konfidenzintervall [0,29; 0,31]für den unbekannten relativen Anteil der elektrisch rasierenden Männer in der männlichen Bevölkerung von einer Marktforschung publiziert. Berechne die dahinterliegende Sicherheit des Konfidenzintervalls an
9) Von welcher Sicherheit muss man ausgehen, wenn eine Wahlumfrage mit 500 Personen einen relativen Anteil der S−Wähler von 28% mit dem Konfidenzintervall [25%,31%] angibt.

---

10) Wie viel Personen muss man befragen, wenn man ein 95%-Konfidenzintervall der Länge 0,04 für die Partei A haben will, die bisher von 25% der Wähler gewählt wurde?
11) Für den relativen Anteil der Befürworter von längeren Öffnungszeiten der Supermärkte soll ein 95%-Konfidenzintervall der Länge 0,02 ermittelt werden.
    a) Wie viele Personen müssen mindestens befragt werden, wenn kein Schätzwert für den untersuchten Anteil bekannt ist (und man daher schlechtesten falls 50% annehmen muss)
    b) Nach einer Telefonumfrage ist bekannt, dass 70% der Personen die längeren Öffnungszeiten befürworten. Wie viele müssen jetzt befragt werden?

Lösungen:

1) Ca. [0,01; 0,09] = [1%; 9%]
2) a) [−0,019; 0,179] = [−1,9%; 17,9%] ➔ *50 ➔ 0−9 fehlerhafte Stoffballen

   b) Die unter Grenze wird größer, die obere kleiner. Insgesamt wird das Intervall kleiner (enger)
3) a) 0−4 fehlerhafte USB−Sticks

   b) wenn n größer wird, wird der Streubereich kleiner (enger)

   c) wenn der Prozentsatz der fehlerhaften USB−Sticks sich 50% annähert, wird der Streubereich größer (weiter)
4) a) [0,215; 0,285] b) [0,204; 0,296]
5) a) [0,084; 0,156] b) [0,073; 0,167]
6) a) [0,632; 0,808] b) [0,604; 0,836]
7) [0,002; 0,031]
8) z= 0,69 ➔ γ= 50,98% ➔ sehr unsicher
9) z=1,494 ➔ $\gamma = 86{,}38\%$ ➔ mittlere Sicherheit
10) mindestens 1801 Personen müssen befragt werden
11) a) mindestens 9604 Personen müssen befragt werden

    b) mindestens 8068 Personen müssen befragt werden