Konfidenzintervalle mit CAS

Reimund Vehling

28. März 2017

Inhaltsverzeichnis

1 Weniger Berechnung - mehr Verständnis

Ausgangspunkt bei Prognoseintervallen ist ein bekannter Parameter p einer Grundgesamtheit. Man will wissen, wie sich Stichprobenergebnisse h um den Parameter p verteilen. Es geht also um die Richtung „*von der Grundgesamtheit zur Stichprobe.*“
Beispiel: Nach Wikipedia haben 35% aller Bundesbürgern die Blutgruppe 0+. Wir nehmen nun an, dass dieser Anteil korrekt ist. In welcher Umgebung um p liegen 95% aller Zufallsstichproben vom Umfang 400?
Es muss das folgende Intervall bestimmt werden: $\left[p - 1{,}96 \cdot \sqrt{\frac{p\cdot(1-p)}{n}}\ ;\ p + 1{,}96 \cdot \sqrt{\frac{p\cdot(1-p)}{n}}\right]$.
Mit $p = 0{,}35$ und $n = 400$ erhält man das Prognoseintervall [0,3032... ; 0,3967...].
Interpretationen:

- Mit 95%-iger Wahrscheinlichkeit liegt ein Stichprobenergebnis h in [0,3032... ; 0,3967...].
- 95% aller Stichprobenergebnisse h fallen in [0,3032... ; 0,3967...].

Wichtiger ist die umgekehrte Richtung „*von der Stichprobe zur Grundgesamtheit*“. Hier kommen nun die so genannten Konfidenzintervalle ins Spiel. Ausgangspunkt bei Konfidenzintervallen ist ein einziges Stichprobenergebnis h. Damit will man etwas über den Anteil p in der Grundgesamtheit erfahren.
Beispiel:
Da wir nicht alle Personen befragen können, treffen wir eine bestimmte Auswahl, die befragt werden. Eine Zufallsstichprobe unter allen Personen in Deutschland hat ergeben, dass 124 von 400 Personen die Blutgruppe 0+ besitzen. Dies entspricht einem Anteil von 31%.
Erste Idee: 31% aller Menschen in Deutschland haben Blutgruppe 0+. Mathematischer ausgedrückt: Die Punktschätzung h ist gleich dem unbekannten Parameter p der Grundgesamtheit. Doch halt: Wir wissen, dass Zufallsstichproben eine natürliche Variabilität aufweisen. Eine andere Zufallsstichprobe wird vermutlich eine andere Punktschätzung hervorbringen, aber irgendwie doch ähnliche Ergebnisse liefern. Müssen wir uns dieser Variabilität beugen und aufgeben? Oder können wir vielleicht doch noch etwas über p sagen? Die Statistiker sind ja schon immer mit einer Unsicherheit von 5% zufrieden. Wir geben also nicht sofort auf. Um diese Frage zu beantworten, ist es wichtig zu wissen, wie Stichprobenergebnisse um den Parameter p schwanken, wenn mehrere Stichproben gezogen werden (siehe Abbildung 1)

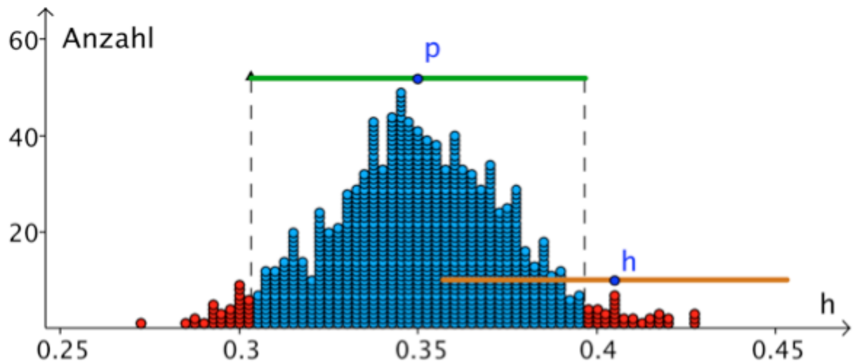


Abbildung 1: Simulation der Stichprobenverteilung

Aber halt – wir gehen nicht hinaus und nehmen weitere Zufallsstichproben. Wir stellen es uns nur vor. Das ist die entscheidende Idee: Was würde passieren, wenn wir viele Stichproben nehmen würden? Das ist übrigens ein zentrales Vorgehen in der Stochastik. Man kann es auch so formulieren: Wir wollen etwas über die Stichprobenverteilung der ermittelten Anteile in Zufallsstichproben wissen. Was wissen wir über unser Modell der Stichprobenverteilung? Zuerst einmal einige Fakten.
Die Stichprobenverteilung

- ist näherungsweise normalverteilt, wenn n hinreichend groß ist,
- hat den (leider uns unbekannten) Erwartungswert aller Menschen in Deutschland mit Blutgruppe 0+,
- hat die Standardabweichung $\sqrt{\frac{p\cdot(1-p)}{n}}$.

Da wir aber p nicht kennen, kennen wir auch nicht den genauen Wert der Standardabweichung. Bei hinreichend großem Wert für n kann man aber den Term durch $\sqrt{\frac{h\cdot(1-h)}{n}}$ abschätzen.
Zur Illustration ein kleines Beispiel: $n = 400$; $p = 0{,}35$; $\sqrt{\frac{p\cdot(1-p)}{n}} = 0{,}0238...$.

h	0,25	0,30	0,32	0,40	0,50
$\sqrt{\frac{h\cdot(1-h)}{n}}$	0,0216...	0,0229...	0,0233...	0,0244...	0,025

Durch die Tatsache, dass sich die Stichprobenverteilung näherungsweise so wie eine Normalverteilung verhält, kann man nun sagen:

- Etwa 68 % aller Stichprobenergebnisse werden Werte für haben, die eine Standardabweichung von entfernt sind.
- Etwa 95 % aller Stichprobenergebnisse werden Werte für haben, die 1,96 Standardabweichungen von entfernt sind.

Aber wo genau ist unser p in der Verteilung? Das wissen wir leider nicht, werden es auch niemals erfahren. Hier die zündende Idee: Den Blickwinkel verändern und alles von unserem bekannten h aus sehen, etwa so:
Falls ich das h bin, dann gibt es eine 95%-Chance, dass p nicht mehr als 1,96-Standardabweichungen von mir entfernt ist. Dann habe ich den Wert p eingefangen – vielleicht. Sicher kann ich mir nicht sein.
Das Beste, was ich machen kann ist es, ein Intervall anzugeben. Und auch dann kann ich mir nicht sicher sein, ob ich p damit eingefangen habe. Ich kann aber sagen, dass diese Methode in 95% aller Fälle dazu führt, dass p vom Intervall eingefangen wird. Man kann auch von „*p wird vom Intervall überdeckt*“ sprechen.
Beispiel: Der Graph der eigentlich niemals bekannten Stichprobenverteilung (genauer: die Approximation mithilfe einer Normalverteilung) sowie die Berechnung von 100 Konfidenzintervallen (genauer: Realisationen). Das 95%-Intervall um p wurde eingezeichnet. Fünf von 100 Intervallen überdecken nicht.

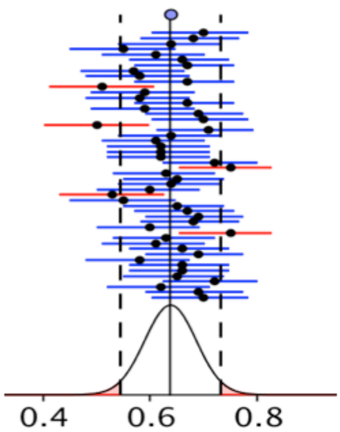


Abbildung 2: 100 Realisationen

Zusammenfassung:
Mit dem Stichprobenergebnis h erhalten wir ein so genanntes 95%-Konfidenzintervall

$$\left[h - 1{,}96 \cdot \sqrt{\frac{h \cdot (1-h)}{n}}\ ;\ h + 1{,}96 \cdot \sqrt{\frac{h \cdot (1-h)}{n}}\right].$$

Was bedeutet es nun, wenn wir sagen, wir haben ein 95%-iges Vertrauen darin, dass unser Intervall den unbekannten Parameter beinhaltet?
Streng genommen meinen wir das Folgende: „*95% aller Stichproben vom gleichen Umfang n werden ein Konfidenzintervall produzieren, dass den unbekannten Parameter p überdeckt (einfängt).*“
Übrigens noch genauer wird es, wenn wir „*Realisation eines Konfidenzintervalls*“ sagen. Die Zufallsgröße H der Stichprobenverteilung „realisiert“ h-Werte und zugehörige Intervallgrenzen. Man kann sagen, dass h und die beiden Intervallgrenzen Realisationen der Zufallsgröße H sind. Etwas ungenau sagt man oft: h und die beiden Intervallgrenzen sind Zufallsgrößen. Übrigens ist der unbekannte Parameter p keine Zufallsgröße, p ist nur unbekannt, hat aber einen festen Wert. Entweder liegt p in dem Konfidenzintervall oder nicht. Mit 95%-iger Wahrscheinlichkeit kann p nicht darin liegen. Wenn man unbedingt will, kann man dem Parameter p nur die Werte 0 oder 1 zuordnen.
Das alles wirkt ziemlich sperrig. Also sagen wir auch oft: „*Wir sind zu 95 % sicher, dass der unbekannte Parameter p in unserem Intervall liegt.*“
Stolpersteine – oder besser: So darf man nicht argumentieren:

- 31 % aller Menschen in Deutschland haben Blutgruppe 0+.
- Es ist wahrscheinlich wahr, dass 31 % aller Menschen in Deutschland die Blutgruppe 0+ haben.
- Wir kennen nicht den genauen Anteil aller Menschen in Deutschland, die Blutgruppe 0+ haben. Aber wir wissen, dass dieser Anteil in unserem berechneten Intervall liegt.
- Mit einer Wahrscheinlichkeit von 95 % liegt der gesuchte Anteil in dem berechneten Intervall.
- Wir kennen nicht den genauen Anteil aller Menschen in Deutschland die Blutgruppe 0+ haben. Aber höchstwahrscheinlich wird dieser Anteil in unserem berechneten Intervall liegen.

Die letzte Äußerung ist übrigens gar nicht mehr weit von der korrekten Interpretation entfernt. Es gibt eine Vielzahl anderer Verfahren, um ein Konfidenzintervall zu berechnen. Andere Methoden können zwar bessere Intervalle liefern, aber die Grundidee ist immer die Gleiche. Man hat nur eine Stichprobe, die (noch) kein Muster zeigt. Aber sehr viele (gedachte) Stichproben ergeben ein Muster, eine Verteilung, mit der wir Aussagen über den unbekannten Parameter p der Grundgesamtheit machen können. Für kleine Stichproben wird diese Methode zu ungenau. Professionelle Umfrageinstitute *arbeiten* oft mit einer Stichprobengröße von ca. 1000. Da ist diese Methode sehr gut geeignet. Es gibt noch einige Fragezeichen. Konfidenzintervalle funktionieren nur wenn echte Zufallsstichproben gezogen werden. Das kann streng genommen eigentlich niemals realisiert werden. Also muss man sich behelfen und sich mit Näherungen bei der Auswahl von Stichproben begnügen.

2 Konfidenzintervalle mit dem TI-Nspire CAS

Mit dem TI-Nspire CAS lässt sich das Gesagte ohne große Mühe erstaunlich einfach umsetzen. Abbildung 3 zeigt als Punktdiagramm das Ergebnis einer Simulation von 400 Stichproben vom Umfang 200 für $p = 0{,}6$.

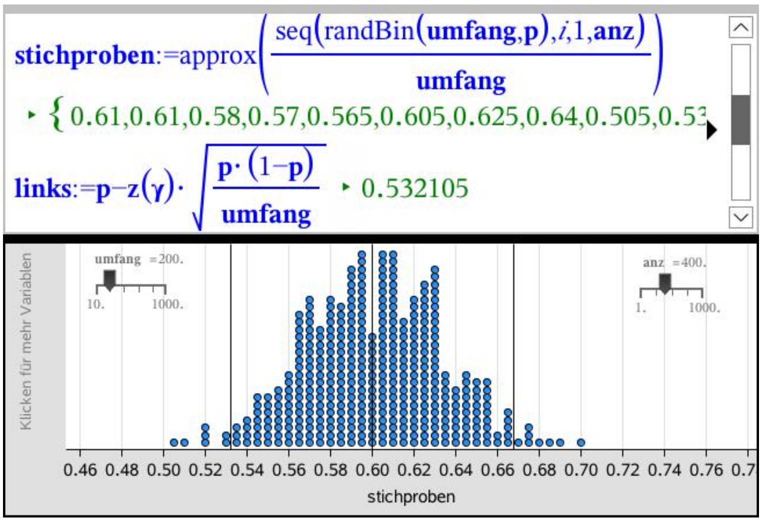
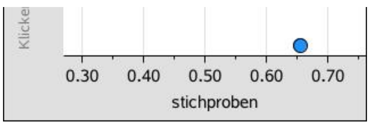


Abbildung 3: Simulation von Prognoseintervallen

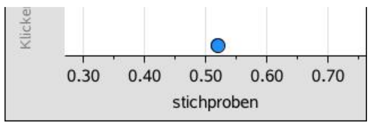
Zusätzlich wurden die beiden Grenzen des 95%-Prognoseintervalls eingezeichnet. Man erkennt, dass insgesamt 15 Stichprobenergebnisse außerhalb des Intervalls liegen. Man erwartet theoretisch zwar 20 Stichprobenergebnisse außerhalb des Prognoseintervalls, aber Zufall und Variabiltät gehören untrennbar zusammen. Schon mit dieser kleinen Lernumgebung können sämtliche Eigenschaften eines Konfidenzintervalls entdeckt werden - und zwar vor den meist umfangreichen Berechnungen:
Nach Konstruktion liegen etwa 95% aller Stichprobenergebnisse innerhalb der $1{,}96\frac{\sigma}{n}$-Umgebung von p, nur etwa 5% außerhalb. Diese Ergebnisse sind mehr als $1{,}96\frac{\sigma}{n}$ von p entfernt. Umgekehrt ist auch p von diesen Ergebnissen mehr als $1{,}96\frac{\sigma}{n}$ entfernt. Diese Erkenntnis ist zwar trivial, aber für den Übergang zu Konfidenzintervallen zentral.
Wir haben ja immer nur eine einzige Stichprobe. Unserer Rettungshaken ist die Erkenntnis, dass in vielen Fällen das Stichprobenergebnis in der Nähe des unbekannten Parameters liegt.
Anschaulich kann das mit einem Rollenspiel simuliert werden, in dem ein Zollstock einen wichtigen Part übernimmt:
Das einzige Stichprobenergebnis - beziehungsweise ein Schüler - erhält den Zollstock, der das Prognoseintervall symbolisieren soll. In 95% aller Fälle sollte p von diesem Intervall überdeckt werden. So kann die erste Annäherung aussehen - ohne Rechnung.
Die beiden Fehler sollten angesprochen werden:

- Da p unbekannt ist, kann man auch kein exaktes Prognoseintervall (keinen Zollstock) kennen.
- Die Symmetrie ist nur für $h = 0{,}5$ gegeben.

Aber mit den Näherungen aus 2 ist der Weg zu einer tragfähigen Grundvorstellung gelegt. Ein Konfidenzintervall ist eine bestimmte Methode, die nur funktioniert, wenn man in Gedanken sich sehr viele Realisationen vorstellt. Der Unterschied von H und h kann angesprochen werden. Auch wird die Interpretation einer Berechnung deutlich.
Unterstützend kann der Rechner wieder eingesetzt werden: Zuerst immer nur ein Stichprobenergebnis und dann wie in Abbildung 3 sehr viele (s. Abb. 4).



(a) ein Stichprobe



(b) eine zweite Stichprobe

Abbildung 4: jeweils eine einzige Stichprobe

3 Theorie

Die Berechnung eines Konfidenzintervalls kann mit dem Rechner auf unterschiedliche Art und Weise geschehen. Da mit dem TI-Nspire CAS auch die Fisherverteilung zur Verfügung steht, können die Grenzen sogar exakt berechnet werden. Das ist aber nur etwas für interessierte Lehrerinnen und Lehrer und nichts für die Schule. Trotzdem soll an dieser Stelle kurz darauf eingegangen werden.
Zwischen der Binomialverteilung und F_{α,f_1,f_2}, dem α-Quantil der F-Verteilung bei f_1 und f_2 Freiheitsgraden, bestehen die folgenden Zusammenhänge:

$$P(X \leq k) = \alpha \Leftrightarrow \frac{(n-k)\cdot p}{(k+1)\cdot(1-p)} = F_{1-\alpha,2\cdot(k+1),2\cdot(n-k)} \tag{1}$$

$$P(X \geq k) = \alpha \Leftrightarrow \frac{k\cdot(1-p)}{(n-k+1)} = F_{1-\alpha,2\cdot(n-k+1),2\cdot k} \tag{2}$$

Diese beiden Beziehungen können jeweils nach p aufgelöst werden. Dann liefert Gleichung 1 für eine Realisation eines Konfidenzintervalls die rechte Grenze p_o, Gleichung 2 die linke Grenze p_u:

$$p_u = \frac{k}{k+(n-k+1)\cdot F_{1-\alpha,2\cdot(n-k+1),2\cdot k}} \tag{3}$$

$$p_o = \frac{(k+1)\cdot F_{1-\alpha,2\cdot(k+1),2\cdot(n-k)}}{(n-k)+(k+1)\cdot F_{1-\alpha,2\cdot(k+1),2\cdot(n-k)}} \tag{4}$$

Mit GeoGebra, EXCEL und CAS-Rechnern können diese Quantile ermittelt werden. Somit können Realisationen von Konfidenzintervallen für Anteile mithilfe des Binomialansatzes ohne große Mühe ermittelt werden. Das macht man aber nicht. Für hinreichend große Stichprobengrößen liefern Näherungen mithilfe der Normalverteilung akzeptable Realisationen. Mehr noch: Es lassen sich damit besser Abhängigkeiten erkennen. Die einzelnen Berechnungen machen mit einem CAS-Rechner keinen Unterschied mehr (s. Abb. 5).

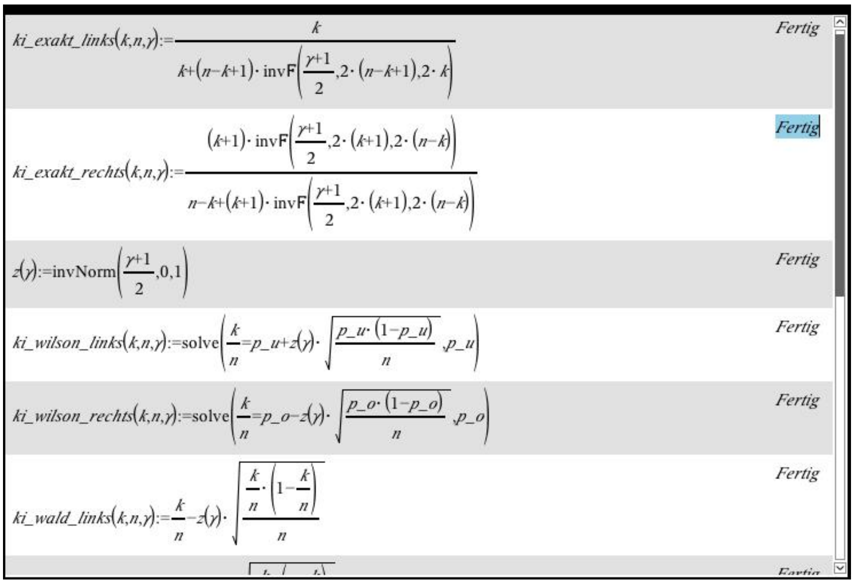


Abbildung 5: Drei Verfahren - umgesetzt mit dem TI_Nspire

Mit diesen Funktionsdefinitionen können nun für beliebige Stichproben die Intervallgrenzen ermittelt werden. Zusätzlich werden mit den einzelnen Funktionsdefinitionen die zugehörigen Berechnungen deutlich. Übrigens wird für WALD-Intervalle keine Funktionsdefinition benötigt. Dieses Berechnung ist schon im Rechner integriert. Der Befehl lautet 1-Prop-z-Intervall. Übrigens existiert keine Funktion für das WILSON-Intervall.