

Teil 3

Schätzungen nach Stichproben
Hochrechnung von der Stichprobe auf das Endergebnis:
Vertrauensintervalle
Abhängigkeit der Genauigkeit vom Umfang der Stichprobe

Berücksichtigung der CAS-Rechner:

CAS O ClassPad 330
+ TI Nspire CAS

Datei Nr. 34 013

Stand: 26. Juli 2018

Friedrich W. Buckel

Inhalt

§ 10 Hochrechnungen und Schätzungen	3
10.1 Schluss von der Gesamtheit auf die Stichprobe	3
Berechnung der Wahrscheinlichkeiten zu X-Intervallen, die zum Erwartungswert symmetrisch liegen	
mit TI Nspire CAS	4
mit CASIO ClassPad	5
Zusammenfassung	7
10.2 Genauigkeit der Interpretation von Stichprobenergebnissen	9
10.3 Vertrauensintervalle für relative Häufigkeiten	11
Grafische Darstellung der h-Intervall-Radien, Wurzeltrichter	13
Starkes Gesetz der großen Zahlen	13
Trainingsaufgaben 1 und 2	14
10.4 Schätzen und Hochrechnen von Wahrscheinlichkeiten	15
Das Ellipsendiagramm	16
Lösung der quadratischen Ungleichung mit der Parabelmethode	17
Zusammenfassung	20
Unterschiede zwischen den Sigma-Intervallen	21
Hochrechnung auf absolute Häufigkeiten	22
10.5 Berechnung des notwendigen Umfangs einer Stichprobe	24
Trainingsaufgaben 3 bis 8	26
10.6 Lösungen der Trainingsaufgaben	28 - 42

§10 Hochrechnungen und Schätzungen

10.1 Schluss von der Gesamtheit auf die Stichprobe

Liegt einer Zufallsvariablen X in einem n -stufigen Experiment eine **Binomialverteilung** zugrunde, dann kann man den **Erwartungswert** durch $E(X) = np$ berechnen. Er sagt den Mittelwert voraus, den man erwarten kann. Der Zufall sorgt jedoch dafür, dass die Ergebnisse für X um diesen Wert schwanken. Diese Schwankungen oder Abweichungen vom Erwartungswert kann man mit der **Standardabweichung** $\sigma = \sqrt{np(1-p)}$ ganz gut abschätzen. Wenn $\sigma > 3$ ist, kann man die Verteilung der Werte von X rund um den Erwartungswert mit den sogenannten **Sigma-Intervallen** vorhersagen. Beispielsweise weiß man, dass dann X mit 95,5% Wahrscheinlichkeit um höchstens 2σ vom Erwartungswert abweicht. Es gibt sogar mehrere nützliche Sigma-Intervalle.

Das Folgende steht ausführlich in 34012 Sto BinVerteilung 2.

Zunächst nochmals ausführlich das 2σ -Intervall:

X liegt zu etwa 95,5% Wahrscheinlichkeit in	$E - 2\sigma \leq X \leq E + 2\sigma$.
Das kann man so schreiben:	$P(E - 2\sigma \leq X \leq E + 2\sigma) \approx 0,955$
Oder in Form einer Betragsungleichung:	$P(X - E \leq 2\sigma) \approx 0,955$

Dann das 3σ -Intervall:

X liegt zu etwa 99,7% Wahrscheinlichkeit in	$E - 3\sigma \leq X \leq E + 3\sigma$.
Das kann man so schreiben:	$P(E - 3\sigma \leq X \leq E + 3\sigma) \approx 0,997$
Oder in Form einer Betragsungleichung:	$P(X - E \leq 3\sigma) \approx 0,997$

Für viele Anwendungen sind „glattere“ Prozentzahlen günstiger. Wir wollen uns jetzt an Hand einer Anwendungsaufgabe an solche Werte herantasten und beobachten, was uns dabei begegnet.

AUFGABE 1 für CAS-Rechner (wer keinen hat, sollte dennoch weiter lesen!)

Die Wahlbeteiligung an einer Parlamentswahl betrug 75 %. Das heißt, dass drei von vier Wahlberechtigten ihre Stimme abgegeben haben. Meinungsforschungsinstitute versuchen durch Befragen von Wählern, schon vor dem offiziellen Endergebnis herauszufinden, wie die Wahl ausgegangen „sein könnte“. Sie befragen also Personen nach ihrem Wahlverhalten. Das Institut „Vorwissen“ möchte in Berlin 2000 Personen befragen. *Das heißt aber nun nicht, dass alle 2000 angesprochenen auch zur Wahl gegangen sind.*

Berechne den Erwartungswert der Anzahl Wähler unter diesen 2000 und berechne die zu E symmetrischen **Intervalle, die zu 70%, 80% und 90% Wahrscheinlichkeit** gehören.

1. Lösung mit TI-Nspire CAS

Erwartungswert für die Zahl X der Wähler: $E(X) = n \cdot p = 2000 \cdot 0,75 = 1500$

Zugehörige Standardabweichung: $\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{2000 \cdot 0,75 \cdot 0,25} \approx 19,36$

Ein zu E symmetrisches Intervall wird durch diese Doppelungleichung beschrieben:

$$E - x \leq X \leq E + x$$

bzw. $1500 - x \leq X \leq 1500 + x$

Gesucht sind drei Werte für x so, dass diese drei Bedingungen erfüllt sind:

- (1) $P(1500 - x \leq X \leq 1500 + x) \approx 0,7$
- (2) $P(1500 - x \leq X \leq 1500 + x) \approx 0,8$
- (3) $P(1500 - x \leq X \leq 1500 + x) \approx 0,9$

Ich habe mit dem CAS-Rechner **TI Nspire** diese Funktion definiert (siehe rechts):

$$p(x) = P(1500 - x \leq X \leq 1500 + x)$$

Mit der Verteilungsfunktion der Binomialverteilung kann man eine Wertetafel dieser Funktion erstellen.

Für $x = 19$, das ist ungefähr σ , bekommen wir:

$$P(1481 \leq X \leq 1519) \approx \boxed{70\%}$$

Für $x = 24$, das ist etwa $1,25 \sigma$, bekommen wir

$$P(1476 \leq X \leq 1524) \approx \boxed{80\%}$$

Für $x = 31$, das sind etwa $1,6 \sigma$, bekommen wir $P(1469 \leq X \leq 1531) \approx \boxed{90\%}$.

Weitere wichtige Werte sind:

Für $x = 1,96 \sigma \approx 38$ erhält man 95% und für $x = 2,58 \sigma \approx 50$ sogar 99% der Werte.

Für $x = 1,96 \sigma \approx 38$ erhält man $1500 - 38 \leq X \leq 1500 + 38 \Leftrightarrow 1462 \leq X \leq 1538$. Das sind 95,3 %.

für $x = 2,58 \sigma \approx 50$ erhält man $1500 - 50 \leq X \leq 1500 + 50 \Leftrightarrow 1450 \leq X \leq 1550$, Das sind 99,1 %.

Kontrollieren wir das mit unserem Rechner:

x	binomCdf(2000, 0.75, 1500-x, 1500+x)
38	0.953248
50	0.990903

Auf den nächsten Seiten folgen zwei Berechnungsmethoden dazu mit **CASIO ClassPad**.

Define	Value
$p(x)$	$\text{binomCdf}(2000, 0.75, 1500-x, 1500+x)$

x	p(x)
16	0.605371
17	0.631449
18	0.660602
19	0.69068
20	0.710243
21	0.73134
22	0.754749
23	0.775106
24	0.794226
25	0.812138
26	0.828872
27	0.844464
28	0.858954
29	0.872383
30	0.884797
31	0.89624
32	0.906762
33	0.916409