Anleitung – Prognose- und Konfidenzintervalle

X sei die Anzahl der Treffer in einer Bernoulli-Kette der Länge n mit der Trefferwahrscheinlichkeit p. μ sei der Erwartungswert von X und σ die Standardabweichung von X. Dann fallen die Werte von X zu etwa P ins Intervall $[\mu - z \cdot \sigma;\ \mu + z \cdot \sigma]$, z. B.

P	68,3%	90%	95%	95,5%	99%	99,7%
z	1	1,64	1,96	2	2,58	3

Voraussetzung für die Gültigkeit dieser sogenannten Sigma-Regeln ist, dass die Laplace-Bedingung $\sigma = \sqrt{n \cdot p \cdot (1-p)} > 3$ erfüllt ist.

Ist die Trefferwahrscheinlichkeit p einer Grundgesamtheit n bekannt, so lässt sich eine Umgebung um den Erwartungswert μ der Trefferzahl X vorhersagen, in dem X mit der durch z festgelegten Sicherheitswahrscheinlichkeit liegen wird. Diese Umgebung nennt man Prognoseintervall. Bei der Bestimmung eines Prognoseintervalls erfolgt also ein Schluss von der Gesamtheit auf die Stichprobe.

Im Gegensatz dazu stehen Konfidenzintervalle (Vertrauensintervalle), bei der von einer Stichprobe auf die Gesamtheit geschlossen wird. Ausgehend von der in einer Stichprobe vom Umfang n erhobenen relativen Häufigkeit X/n wird mit der durch z vorgegebenen Sicherheitswahrscheinlichkeit die Trefferwahrscheinlichkeit p in der Grundgesamtheit ermittelt.

Die folgenden Formeln und Umformungen zeigen den Zusammenhang zwischen Prognose- und Konfidenzintervallen auf und bilden den Ausgangspunkt für die Bestimmung der Prognose- und Konfidenzintervalle mithilfe eines WTR oder CAS.

$\mu - z \cdot \sigma \leq X \leq \mu + z \cdot \sigma$	$\mid \mu = n \cdot p$	Formel 1a
$\Leftrightarrow n \cdot p - z \cdot \sigma \leq X \leq n \cdot p + z \cdot \sigma$	$\mid : n$	Formel 1b
$\Leftrightarrow p - z \cdot \frac{\sigma}{n} \leq \frac{X}{n} \leq p + z \cdot \frac{\sigma}{n}$	$\mid$ Aufteilen in zwei Ungleichungen	Formel 1c
$\Leftrightarrow \begin{cases} \frac{X}{n} \leq p + z \cdot \frac{\sigma}{n} \\ \frac{X}{n} \geq p - z \cdot \frac{\sigma}{n} \end{cases}$	$\mid -p$	
$\Leftrightarrow \begin{cases} \frac{X}{n} - p \leq +z \cdot \frac{\sigma}{n} \\ \frac{X}{n} - p \geq -z \cdot \frac{\sigma}{n} \end{cases}$	$\mid$ unten $\cdot\ (-1)$	
$\Leftrightarrow \begin{cases} \frac{X}{n} - p \leq z \cdot \frac{\sigma}{n} \\ -\left(\frac{X}{n} - p\right) \leq z \cdot \frac{\sigma}{n} \end{cases}$	$\mid$ Definition Betrag, $\sigma = \sqrt{n \cdot p \cdot (1-p)}$	
$\Leftrightarrow \left\lvert \frac{X}{n} - p \right\rvert \leq z \cdot \frac{\sqrt{n \cdot p \cdot (1-p)}}{n}$	$\mid \cdot n$	Formel 2a
$\Leftrightarrow \lvert X - n \cdot p \rvert \leq z \cdot \sqrt{n \cdot p \cdot (1-p)}$	$\mid \mu = n \cdot p$ und $\sigma = \sqrt{n \cdot p \cdot (1-p)}$	Formel 2b
$\Leftrightarrow \lvert X - \mu \rvert \leq z \cdot \sigma$	$\mid$ zurück zu Formel 1c	Formel 2c
$\Leftrightarrow p - z \cdot \frac{\sigma}{n} \leq \frac{X}{n} \leq p + z \cdot \frac{\sigma}{n}$	$\mid -\frac{X}{n}, -p$	
$\Leftrightarrow -\frac{X}{n} - z \cdot \frac{\sigma}{n} \leq -p \leq -\frac{X}{n} + z \cdot \frac{\sigma}{n}$	$\mid \cdot\ (-1)$	
$\Leftrightarrow \frac{X}{n} + z \cdot \frac{\sigma}{n} \geq p \geq \frac{X}{n} - z \cdot \frac{\sigma}{n}$	$\mid$ Vertauschen der beiden Seiten	
$\Leftrightarrow \frac{X}{n} - z \cdot \frac{\sigma}{n} \leq p \leq \frac{X}{n} + z \cdot \frac{\sigma}{n}$	$\mid \cdot n$	Formel 3a
$\Leftrightarrow X - z \cdot \sigma \leq n \cdot p \leq X + z \cdot \sigma$	$\mid \mu = n \cdot p$	Formel 3b
$\Leftrightarrow X - z \cdot \sigma \leq \mu \leq X + z \cdot \sigma$		Formel 3c

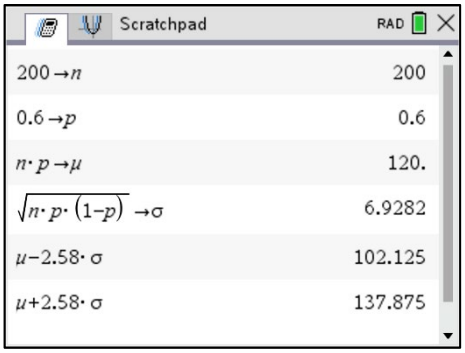
Schiemenz - www.hyperschlau.de

Mit der Formel 1 lassen sich Prognoseintervalle mithilfe eines WTR oder CAS bestimmen.

Beispiel: Eine Reißzwecke wird 200-mal geworfen. In 60% der Fälle kommt diese Reißzweckensorte auf der Seite zu liegen. X sei die Anzahl der Würfe, bei denen die Reißzwecke auf der Seite zu liegen kommt.

Bestimmen Sie mit einer Sicherheitswahrscheinlichkeit von 99% ein Prognoseintervall für die Anzahl an Seitenlagen.

Bei 200 Würfen kommt mit einer Wahrscheinlichkeit von 99% die Seitenlage der Reißzwecke zwischen 102- und 138-mal vor.



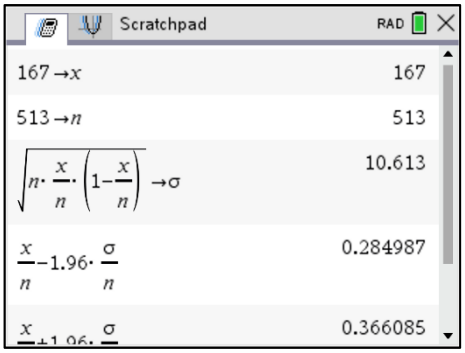
Mit der Formel 3 lassen sich Konfidenzintervalle mithilfe eines WTR oder CAS näherungsweise bestimmen. Da p unbekannt ist – genau das soll ja bestimmt werden – muss σ mit der relativen Häufigkeit abgeschätzt werden:

$$\sigma = \sqrt{n \cdot \frac{X}{n} \cdot \left(1 - \frac{X}{n}\right)}$$

Beispiel: Kurz vor einer Wahl geben 167 von 513 zufällig Befragten an, die Partei der Fleischesser wählen zu werden.

Bestimmen Sie mit einer Sicherheitswahrscheinlichkeit von 95% ein Konfidenzintervall für den Stimmenanteil, mit dem die Partei der Fleischesser (PdF) bei der Wahl rechnen kann.

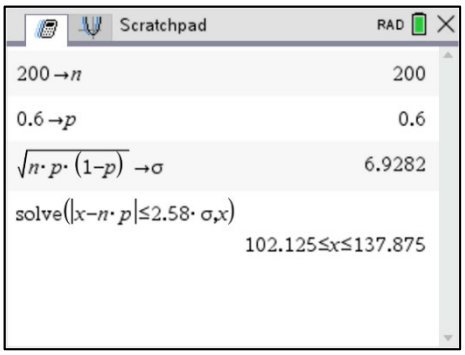
Die PdF kann mit einer Wahrscheinlichkeit von 95% mit einem Stimmanteil zwischen 28,4% und 36,7% rechnen.



Die Formel 2 ist die universelle Formel für den CAS. Mit ihr können sowohl Prognose- als auch Konfidenzintervalle berechnet werden.

Das erste Beispiel wird erneut betrachtet.

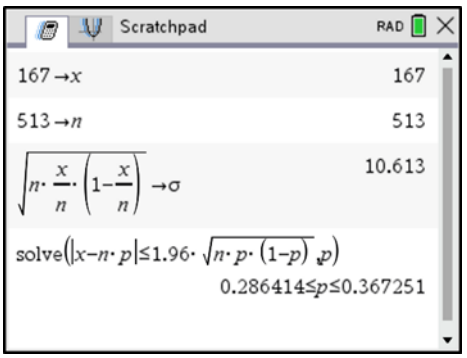
Der Eingabeaufwand bei der Nutzung des solve-Befehls ist nahezu unverändert. Das Ergebnis wird jedoch als Intervall ausgegeben.



Das zweite Beispiel wird erneut betrachtet.

Der Eingabeaufwand bei der Nutzung des solve-Befehls ist nahezu unverändert. Das Ergebnis ist jedoch genauer, da die Abschätzung von σ nur zur Prüfung der Laplace-Bedingung genutzt wird, nicht jedoch bei der Berechnung von p.

Die PdF kann mit einer Wahrscheinlichkeit von 95% mit einem Stimmanteil zwischen 28,6% und 36,8% rechnen.



Schiemenz - www.hyperschlau.de