Prognoseintervalle mit CAS-Rechner

Dr. Wilfried Zappe, Ilmenau
Illustrationen von Dr. Wilfried Zappe

© *johnwoodcock/DigitalVision Vectors/Getty Images Plus*

Wirft man eine „ideale" Münze n-mal und betrachtet das Ergebnis „Wappen" als Treffer, so geht man davon aus, dass die Trefferwahrscheinlichkeit $p = 0{,}5$ ist.
Trotzdem wird es in einer konkreten Stichprobe des Umfangs n häufig passieren, dass nicht genau die Hälfte der Ergebnisse „Wappen" lautet. Vielmehr wird man erwarten dürfen, dass die Anzahl der Treffer zufallsbedingt in einem Intervall um den Erwartungswert liegt. Im Mathematikunterricht der Oberstufe lassen sich solche Prognoseintervalle im Zusammenhang mit den Sigma-Regeln der Binomialverteilung quantitativ berechnen und inhaltlich interpretieren. Sie bieten einen sehr guten Zugang zur Betrachtung von Konfidenzintervallen.

RAABE
KLASSE SCHULE

Impressum

RAABE UNTERRICHTS-MATERIALIEN Wahrscheinlichkeitsrechnung und Statistik

Dr. Josef Raabe Verlags-GmbH
Ein Unternehmen der Klett Gruppe
Rotebühlstraße 77
70178 Stuttgart
Telefon +49 711 62900-0
Fax +49 711 62900-60
meinRAABE@raabe.de
www.raabe.de

Redaktion: Anna-Greta Wittnebel
Satz: Röser Media GmbH & Co. KG, Karlsruhe
Bildnachweis Titel: © johnwoodcock/DigitalVision Vectors/Getty Images Plus
Lektorat: Maria Hitzenauer, Regensburg; Rebecca Saalfeld, Köln
Korrektorat: Johanna Stotz, Wyhl a. K.

Prognoseintervalle mit CAS-Rechner

Oberstufe (erhöhtes Niveau)

Dr. Wilfried Zappe, Ilmenau
Illustrationen von Dr. Wilfried Zappe

Die Schüler lernen:

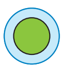
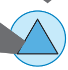
- die Sigma-Regeln für binomialverteilte Zufallsgrößen anzuwenden,
- Prognoseintervalle für absolute und relative Häufigkeiten bei Stichproben binomialverteilter Zufallsgrößen zu berechnen und zu interpretieren,
- die Begriffe „signifikante Abweichung" und „statistische Verträglichkeit" sachgerecht zu verwenden,
- die Möglichkeit der Näherung hypergeometrisch verteilter Zufallsgrößen durch binomialverteilte Zufallsgrößen kennen und anzuwenden,
- die Möglichkeiten des CAS-Rechners zur Berechnung und Visualisierung von Prognoseintervallen auszunutzen.

Überblick:

Legende der Abkürzungen:
Ab = Arbeitsblatt

Thema	Material	Methode
Die Sigma-Regeln der Binomialverteilung	M 1	Ab
Hypergeometrische Verteilung/Binomialverteilung	M 2	Ab
Prognoseintervalle für absolute Häufigkeiten	M 3	Ab
Prognoseintervalle für relative Häufigkeiten	M 4	Ab

Erklärung zu Differenzierungssymbolen

		
einfaches Niveau	mittleres Niveau	schwieriges Niveau
	Dieses Symbol markiert Zusatzaufgaben.	

Kompetenzprofil

Inhalt: Binomialverteilung, Sigma-Regeln, Hypergeometrische Verteilung, absolute Häufigkeit, relative Häufigkeit, Prognoseintervall, signifikante Abweichung, statistische Verträglichkeit, Näherung, Zufallsgröße, Ergebnis, Ereignis

Medien: GTR/CAS

Kompetenzen: Probleme mathematisch lösen (K2), mit symbolischen, formalen und technischen Elementen der Mathematik umgehen (K5), kommunizieren (K6)

Prognoseintervalle mit CAS-Rechner – Hinweise

Prognoseintervalle geben an, mit welcher absoluten bzw. relativen Häufigkeit ein Ereignis in einer Stichprobe vom Umfang n mit einer vorgegebenen Sicherheitswahrscheinlichkeit auftritt, wenn die Wahrscheinlichkeit für das Auftreten des Ereignisses in der **Grundgesamtheit bekannt** ist. Im Mittelpunkt steht die Beziehung $P(\mu - k \cdot \sigma \leq X \leq \mu + k \cdot \sigma)$ einer binomialverteilten Zufallsgröße. Dabei bedeuten:

μ: Erwartungswert der binomialverteilten Zufallsgröße X

σ: Standardabweichung der binomialverteilten Zufallsgröße X

k: ist eine positive reelle Zahl

Die Beziehung $P(\mu - k \cdot \sigma \leq X \leq \mu + k \cdot \sigma)$ beschreibt die Wahrscheinlichkeit, mit der Werte einer binomialverteilten Zufallsgröße X in der k-fachen Sigma-Umgebung des Erwartungswertes von X liegen. Es lässt sich z. B. zeigen, dass in der zweifachen Sigma-Umgebung des Erwartungswertes von X ca. 95,4 % der Werte von X liegen: $P(\mu - 2 \cdot \sigma \leq X \leq \mu + 2 \cdot \sigma) \approx 0{,}954$.

Für einen exakten Nachweis werden Kenntnisse der Normalverteilung gebraucht, jedoch kann man gerade bei Verfügbarkeit von CAS solche Zusammenhänge auch für binomialverteilte Zufallsgrößen plausibel machen.

Wir gehen deshalb hier davon aus, dass die Schülerinnen und Schüler[1] vor der Einführung von Prognoseintervallen binomialverteilte Zufallsgrößen im Unterricht kennengelernt haben und notwendige Berechnungen und Veranschaulichungen mit einem CAS-Rechner vornehmen können.

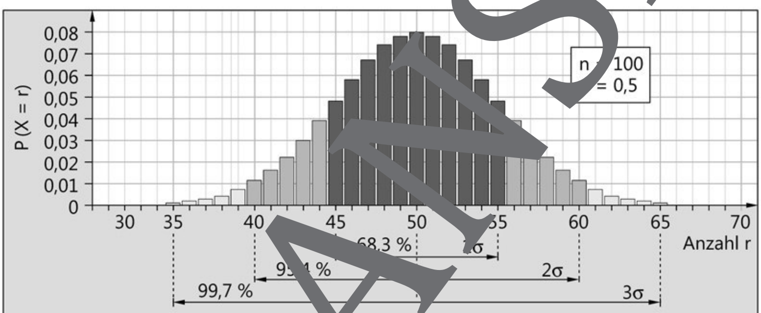
Für den Fall, dass die Schüler nicht wissen, dass man für $N >> n$ hypergeometrisch verteilte Zufallsgrößen durch binomialverteilte Zufallsgrößen annähern kann, ist ein Abschnitt eingefügt, der eine beispielgebundene Argumentation dazu liefert. In den einzelnen Abschnitten erfolgt eine Darstellung von wesentlichen Theorieaspekten. Typische Problemstellungen werden an Beispielen mit Lösung erläutert und durch weitere Aufgaben gefestigt. Der Einsatz von CAS-Rechnern wird mit dem TI-Nspire CX CAS gezeigt.

---

[1] Aus Gründen der besseren Lesbarkeit wird im weiteren Verlauf nur noch „Schüler“ verwendet.

M 1 Die Sigma-Regeln der Binomialverteilung

Theorie: Die Sigma-Regeln der Binomialverteilung beschreiben, wie viel Prozent der Werte einer binomialverteilten Zufallsgröße X näherungsweise in einem zum Erwartungswert μ symmetrischen Intervall liegen. Die Grenzen dieses Intervalls gibt man mithilfe von Vielfachen der Standardabweichung σ an. Sie heißen deshalb auch „Sigma-Umgebungen“.
Die Sigma-Regeln können den Schülern gut als Erfahrungstatsache nahegebracht werden. Die Verfügbarkeit von CAS-Rechnern ist dabei eine große Hilfe, weil sich geeignete Berechnungen und Simulationen rasch und effizient erstellen lassen und somit ausreichend empirisches Material für eine glaubhafte Verallgemeinerung erzeugen lässt.
Aus der Vielfalt der möglichen Sigma-Regeln wird eine besonders hervorgehoben und für die folgenden Untersuchungen in didaktischer Vereinfachung verwendet.



Grafik: Dr. Wolfgang Zeitlmaier

Sigma-Regeln: Ist die Zufallsgröße X mit den Parametern n und p binomialverteilt, dann gelten für ihren Erwartungswert $\mu = n \cdot p$ und ihre Standardabweichung $\sigma = \sqrt{n \cdot p \cdot (1-p)}$ für genügend große Werte von n (Laplace-Bedingung: $\sigma > 3$) näherungsweise folgende Intervallwahrscheinlichkeiten:

k	$P(\mu - k \cdot \sigma \le X \le \mu + k \cdot \sigma) \approx$	k	$P(\mu - k \cdot \sigma \le X \le \mu + k \cdot \sigma) \approx$
1	0,683	1,64	0,90
2	0,955	1,96	0,95
3	0,997	2,58	0,99

Für alle folgenden Untersuchungen zu Prognoseintervallen wird vorwiegend die nachstehende Näherung der Zwei-Sigma-Regel verwendet:

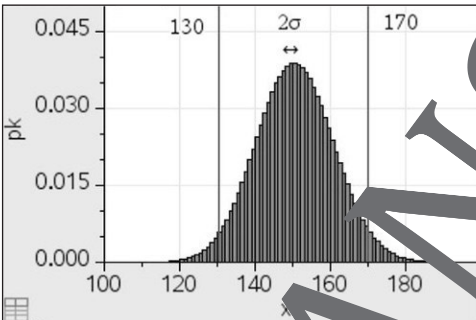
> **Für genügend große Werte von n** liegen ca. **95 %** der Werte einer **binomialverteilten** Zufallsgröße X in der **Zwei-Sigma-Umgebung** ihres Erwartungswertes:
> $P(\mu - 2\cdot\sigma \le X \le \mu + 2\cdot\sigma) \approx 0{,}95$

Beispiel 1: Ganzzahlige Grenzen eines Zwei-Sigma-Intervalls

Die Zufallsgröße X sei binomialverteilt mit den Parametern n = 500 und p = 0,3.

$\mu = n\cdot p = 500\cdot 0{,}3 = 150;\quad \sigma = \sqrt{n\cdot p\cdot(1-p)} = \sqrt{500\cdot 0{,}3\cdot 0{,}7} \approx 10{,}2 > 3$

(Die Laplace-Bedingung ist erfüllt.)



© Dr. Wilfried Zappe

Zwei-Sigma-Umgebung von μ:

$\mu - 2\cdot\sigma \le X \le \mu + 2\cdot\sigma$
$\Rightarrow 150 - 2\cdot 10{,}2 \le X \le 150 + 2\cdot 10{,}2$
$\Rightarrow 129{,}6 \le X \le 170{,}4$

Wegen der Ganzzahligkeit der Werte einer binomialverteilten Zufallsgröße und unter Beachtung der Ungleichheitszeichen in der Doppelungleichung ist das Zwei-Sigma-Intervall gegeben durch: $130 \le X \le 170$.
Für die zugehörige Wahrscheinlichkeit ergibt sich: $P(130 \le X \le 170) \approx 0{,}95$

binomCdf(500,0.3,130,170)	0.954713

Beispiel 2: Simulation in Lists & Spreadsheet

Es sei $X \sim B_{100;0,5}$.

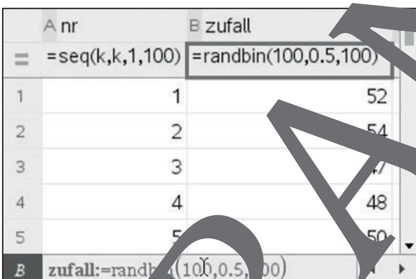
Für X gilt

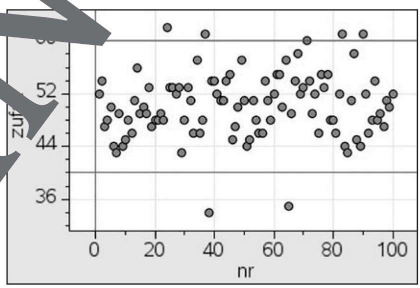
$\mu = n \cdot p = 100 \cdot 0{,}5 = 50$ und

$\sigma = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{100 \cdot 0{,}5 \cdot 0{,}5} = 5.$

Nach der Zwei-Sigma-Regel müssten ca. 95 % der zugehörigen Werte von X innerhalb der Zwei-Sigma-Umgebung [40; 60] des Erwartungswertes $\mu = 50$ liegen. Das lässt sich durch eine Simulation sehr schön veranschaulichen. In **Lists & Spreadsheet** wird in der Spalte B mit der Anweisung „randbin(100,0.5,100)" eine Liste „zufall" von 100 binomialverteilten Zufallszahlen erzeugt und in Spalte A eine Liste „nr" mit „seq(k,k,1,100)" der Zahlen von 1 bis 100. In **Data & Statistics** werden die Wertepaare der Listen „nr" und „zufall" grafisch dargestellt.

Die Grenzen des Zwei-Sigma-Intervalls werden unter **Menü – Analysieren – Funktion zeichnen** als $f_1(x) = 40$ und $f_2(x) = 60$ eingezeichnet.





© Dr. Wilfried Zappe

Die außerhalb des Zwei-Sigma-Intervalls liegenden Werte lassen sich abzählen und mit der Prognose vergleichen. Geht man zurück in die Tabellenkalkulation, wird mit <Ctrl><R> eine Neuberechnung der Simulation veranlasst. Ohne großen Aufwand können genügend viele solcher Simulationen realisiert und z. B. durch Mittelbildung ausgewertet werden.

Aufgaben

1. Berechnen Sie für die angegebenen Parameter jeweils die fehlenden Werte und die zugehörigen Intervallwahrscheinlichkeiten. Arbeiten Sie in Gruppen, und wählen Sie für jede Gruppe genau einen der Werte von $k \in \{1;\ 2;\ 3;\ 1{,}64;\ 1{,}96;\ 2{,}58\}$.

n	p	μ	σ	k	$P(\mu - k \cdot \sigma \le X \le \mu + k \cdot \sigma)$
100	0,5	50	5		
1000	0,2				
10 000	0,8				
10	0,5				
500		40			
	0,5		100		

Hinweis: Zeitsparend kann bei Verwendung des TI-Nspire die Anlage und Speicherung eines Rechenblattes in der Anwendung „Notes" nach folgendem Muster sein:

$n:=100 \blacktriangleright 100 \quad p:=0.5 \blacktriangleright 0.5 \quad k:=2 \blacktriangleright 2$

$e:=n\cdot p \blacktriangleright 50. \quad s:=\sqrt{n\cdot p\cdot(1-p)} \blacktriangleright 5.$

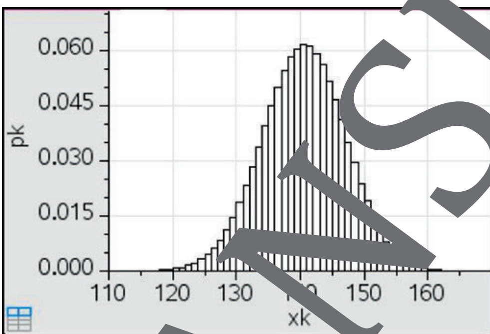
$\mathrm{binomCdf}(n,p,e-k\cdot s,e+k\cdot s) \blacktriangleright 0.9648$

Alternativ: Verwenden Sie ein Tabellenkalkulationsprogramm.

2. Ergänzen Sie in der Tabelle die fehlenden Werte für die Sigma-Regeln:

k	$P(\mu - k\cdot\sigma \le X \le \mu + k\cdot\sigma) \approx$	k	$P(\mu - k\cdot\sigma \le X \le \mu + k\cdot\sigma) \approx$
1			0,90
2	0,955	1,96	0,95
3	0,997		0,99

3. Markieren Sie farbig im folgenden Histogramm einer $B_{200;0,7}$-verteilten Zufallsgröße den Ein-Sigma-Bereich.



© Dr. Wilfried Zappe

4. Die Zufallsgröße X sei binomialverteilt mit n = 150 und p = 0,4. Es wird 30-mal zufällig ein Wert von X erzeugt. Berechnen Sie die Wahrscheinlichkeit des Ereignisses A, dass mindestens einer dieser Werte außerhalb des Zwei-Sigma-Intervalls liegt.
Beurteilen Sie, wie sich das Ergebnis verändert, wenn n vergrößert wird.[2]

---

[2] Eine ähnliche Aufgabe gibt es in Klett: „Lambacher Schweizer Mathematik Stochastik" Seite 74 Nr. 3 (ISBN 978-3-12-735710-3)

M 2 Hypergeometrische Verteilung/Binomialverteilung

© John Rensten/The Image Bank/Getty Images Plus

Theorie: Eine häufig anzutreffende Anwendung bei Prognoseintervallen (später auch bei Konfidenzintervallen) ist die mathematische Auswertung und Beurteilung von Meinungsumfragen. Diese sind in der Regel Stichproben durch „Ziehen ohne Zurücklegen", denn eine einmal befragte Person wird nicht in derselben Umfrage ein zweites Mal zum gleichen Thema befragt. Die Grundlage für eine mathematische Modellierung für das Modell „Ziehen ohne Zurücklegen" ist die hypergeometrische Verteilung. Ihr liegt folgende Vorstellung zugrunde: In einer Urne befinden sich N Kugeln, davon M rote. Die Wahrscheinlichkeit, beim Ziehen von n Kugeln ohne Zurücklegen k rote Kugeln zu ziehen, wird durch den nebenstehenden Term (oft auch als „Lottoformel" bezeichnet) berechnet.

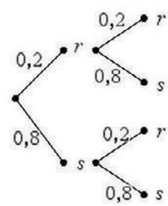
$$P(X=k)=\frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}}$$

Wenn der Umfang N der Gesamtheit sehr viel größer ist als der Stichprobenumfang n (man liest oft von $N \geq 20\,n$), unterscheiden sich die hypergeometrische Verteilung und die Binomialverteilung kaum, und man nutzt dann meistens die leichter zu handhabende Binomialverteilung. (ohne Beweis)

Beispiel 3: Ziehen mit und ohne Zurücklegen

Eine Urne enthalte 400 schwarze und 100 rote Kugeln.

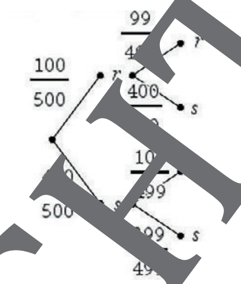
Die Zufallsgröße X beschreibt die Anzahl der roten Kugeln, wenn aus der Urne zweimal gezogen wird und die zuerst gezogene Kugel wieder zurückgelegt wird.



© Dr. W. Zappe

X = k	0	1	2
Ergebnisse	(s; s)	(r; s) (s; r)	(r; r)
P(X = k)	0,64	0,32	0,04

Die Zufallsgröße Y beschreibe die Anzahl der roten Kugeln, wenn aus der Urne zweimal gezogen wird und die zuerst gezogene Kugel nicht wieder zurückgelegt wird.
Beim zweimaligen Ziehen ohne Zurücklegen ändern sich in der zweiten Stufe die Wahrscheinlichkeiten (siehe Baumdiagramm).



© Dr. W. Zappe

Y = k	0	1	2
Ergebnisse	(s; s)	(r; s) (s; r)	(r; r)
P(Y = k)	$\approx 0{,}6397$	$\approx 0{,}3206$	$\approx 0{,}0397$

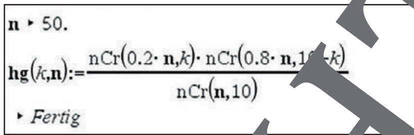
Die Wahrscheinlichkeitsverteilungen beim Ziehen mit und beim Ziehen ohne Zurücklegen unterscheiden sich kaum.
Woran liegt das?

Wenn sehr viele Kugeln in der Urne liegen und nur wenige davon gezogen werden, dann ändert sich der Anteil einer Sorte im Laufe der Ziehungen nur wenig. Betrachtet man z. B. den oberen Pfad im Baumdiagramm zur Zufallsgröße Y, dann ist $\frac{99}{499} \approx 0{,}198 \approx 0{,}2$ fast gleich dem Wert $\frac{100}{500} = 0{,}2$.

Beispiel 4: Simulation von Binomialverteilung und hypergeometrischer Verteilung
Die Zufallsgröße X beschreibt die Anzahl k roter Kugeln beim zehnmaligen Ziehen einer Kugel **mit Zurücklegen** aus einer Urne mit $n \geq 50$ Kugeln, von denen ein Fünftel rot sind. X ist also binomialverteilt mit n = 10 und p = 0,2.
Die Zufallsgröße Y beschreibt die Anzahl k roter Kugeln beim zehnmaligen Ziehen einer Kugel **ohne Zurücklegen** aus einer Urne mit $N \geq 50$ Kugeln, von denen vor dem ersten Ziehen ein Fünftel rot sind. Y ist also hypergeometrisch verteilt.
Jede Zufallsgröße kann die Werte k mit $0 \leq k \leq 10$ annehmen.
Die Gesamtzahl der Kugeln soll variiert werden können.

Zufallsgröße Y:

In der Applikation „Notes" wird dem Parameter n ein Wert, z. B. N = 50, zugewiesen. Der Term für die Berechnung der hypergeometrischen Wahrscheinlichkeit wird als Funktion $h_g(k, N)$ definiert:



Grundgesamtheit: N = 50, Anzahl rote Kugeln: $M = 0{,}2 \cdot N$, Stichprobenumfang: n = 10

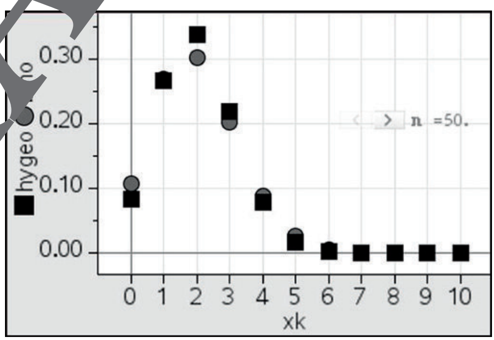
$$h_g(k,N) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} = \frac{\binom{0{,}2 \cdot N}{k} \cdot \binom{0{,}8 \cdot N}{10-k}}{\binom{N}{10}}$$

In der Anwendung **Lists & Spreadsheet** werden die Werte für k, für die Binomialverteilung mit n = 10 und p = 0,2 sowie für die hypergeometrische Verteilung $h_g(k, N)$ tabelliert.

	A xk	B bino	C hygeo
=	=seq(k,k,0,10)	=binompdf(10,0.2)	=seq(hg(k,n),k,0,10)
1	0	0.1073[illegible]	[illegible]519
2	1	0.268435	0.266191
3	2	0.30199	0.336898

In der Anwendung **Data & Statistics** werden die Wertepaare grafisch dargestellt.

Für die Grundgesamtheit n (bzw. N) wird ein Schieberegler eingerichtet. Nun lässt sich zeigen, dass bei immer größer werdendem n die Funktionswerte beider Verteilungen immer besser in Übereinstimmung kommen.



© Dr. Wilfried Zappe

Aufgaben

1. Es sei X eine hypergeometrisch verteilte Zufallsgröße mit $N = 1000$, $M = 200$ und $n = 50$. Berechnen Sie $P(X = 10)$ mithilfe der Lottoformel und vergleichen Sie diesen Wert mit $P(Y = 10)$ der binomialverteilten Zufallsgröße $Y_{50;0,2}$.

$$P(X=k)=\frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}}$$

2. Der Hausmeister eines Gymnasiums mit 500 Schülern kennt jeden fünften Schüler persönlich. Auf dem Schulflur begegnen ihm acht Schüler. Mit welcher Wahrscheinlichkeit trifft er keinen ihm bekannten Schüler? Rechnen Sie mit der „Lottoformel“ und näherungsweise mit der Binomialverteilung.

M 3 Prognoseintervalle für absolute Häufigkeiten

Theorie: Mithilfe der Zwei-Sigma-Regel lässt sich z. B. vor der Durchführung einer Kette von n gleichartigen Bernoulli-Versuchen prognostizieren, dass in dieser Stichprobe vom Umfang n mit einer Wahrscheinlichkeit von ca. 95 % die **absolute Häufigkeit H** eines Merkmals im Intervall $n\cdot p-2\cdot\sqrt{n\cdot p\cdot(1-p)}\le H\le n\cdot p+2\cdot\sqrt{n\cdot p\cdot(1-p)}$ liegen wird, wenn p die **bekannte** oder als **wahr angenommene** Wahrscheinlichkeit dieses Merkmals in der Gesamtheit ist. Man führt also einen Schluss von der Gesamtheit auf die Stichprobe (Intervall für H) durch. Liegt das Stichprobenergebnis im Zwei-Sigma-Prognoseintervall, dann heißt das Ergebnis „statistisch verträglich" mit der in der Grundgesamtheit geltenden Wahrscheinlichkeit p. Liegt das Stichprobenergebnis außerhalb des 95 %-Prognoseintervalls, dann spricht man von einer „signifikanten Abweichung" von p. Liegt das Stichprobenergebnis sogar außerhalb des Drei-Sigma-Intervalls, dann heißt die Abweichung „hochsignifikant".

Signifikanzniveau 95 %	
Statistisch verträglich	Liegt im Zwei-Sigma-Prognoseintervall.
Signifikante Abweichung	Liegt außerhalb des Zwei-Sigma-Prognoseintervalls, aber innerhalb des Drei-Sigma-Intervalls.
Hochsignifikante Abweichung	Liegt außerhalb des Drei-Sigma-Intervalls.

Beispiel 5: Prognoseintervall für „Ziehen mit Zurücklegen"

Es wird angenommen, dass ein Spielwürfel die „Sechs" mit der Wahrscheinlichkeit $p=\frac{1}{6}$ liefert.

Gesucht ist ein Prognoseintervall, dass bei 200 Würfen mit diesem Spielwürfel die Anzahl der „Sechsen" mit einer Sicherheitswahrscheinlichkeit von ca. 95 % enthält. Wie ist das Ergebnis zu interpretieren, wenn bei diesen 200 Würfen 15-mal eine „Sechs" geworfen wurde?

Vorgehensweise:

Entscheiden, ob dem Sachverhalt das Modell der Binomialverteilung zugrunde gelegt werden kann.	Ja, es sind zwei unabhängige Ergebnisse möglich: „Sechs“ oder „keine Sechs“.
Kenngrößen der Binomialverteilung und den Wert für k bezüglich der Sicherheitswahrscheinlichkeit erfassen.	$n = 200;\ p = \frac{1}{6}$ $k = 2$
Erwartungswert und Standardabweichung berechnen.	$\mu = 200 \cdot \frac{1}{6} \approx 33{,}33$ $\sigma = \sqrt{200 \cdot \frac{1}{6} \cdot \frac{5}{6}} \approx 5{,}27$
Grenzen des 95 %-Prognoseintervalls angeben.	$g_u \approx 33{,}33 - 2 \cdot 5{,}27 = 22{,}79$ $g_o \approx 33{,}33 + 2 \cdot 5{,}27 = 43{,}87$
Prognoseintervall für H unter Beachtung der Ganzzahligkeit der Grenzen und der Richtung der Ungleichheitszeichen in der Sigma-Regel angeben.	$22{,}79 \le H \le 43{,}87$ also: $23 \le H \le 43$

Die Anzahl der „Sechsen“ wird bei 200 Würfen mit einer Sicherheitswahrscheinlichkeit von ca. 95 % im Intervall [23; 43] liegen.

Hinweis: Eine Probe sollte durchgeführt werden. Es gilt für eine binomialverteilte Zufallsgröße X mit n = 200 und $p = \frac{1}{6}$:
$P(23 \le X \le 43) \approx 0{,}954$.

$\text{binomCdf}\left(200, \frac{1}{6}, 23, 43\right)$ 0.953643

Die Sicherheitswahrscheinlichkeit von ca. 95 % wird also erreicht.

Interpretation: Wenn bei 200 Würfen nur 15-mal eine „Sechs“ erscheint, gibt das Anlass dazu, an der Behauptung $p = \frac{1}{6}$ zu zweifeln, weil 15 Treffer außerhalb des Zwei-Sigma-Intervalls liegen.

Beispiel 6: Prognoseintervall für „Ziehen ohne Zurücklegen“

Nach Auskunft des KBA betrug der Anteil der Pkw Mazda am 1. Januar 2019 in Deutschland 1,8 %. Wie viele Pkw vom Typ Mazda sind bei einer Verkehrszählung unter 500 erfassten Pkw bei einer Sicherheitswahrscheinlichkeit von 95 % zu erwarten?

Das Zählen von Fahrzeugen und Feststellen des Herstellers entspricht dem Urnenmodell „Ziehen ohne Zurücklegen“. Da aber der Umfang der Stichprobe mit n = 500 verschwindend gering ist gegenüber der Gesamtzahl von Pkw in Deutschland von über 47 Millionen, kann näherungsweise von einer Binomialverteilung ausgegangen werden.

Vorgehensweise:

Entscheiden, ob dem Sachverhalt das Modell der Binomialverteilung zugrunde gelegt werden kann.	Die Binomialverteilung kann näherungsweise angewendet werden.
Kenngrößen der Binomialverteilung und Wert für k bezüglich der Sicherheitswahrscheinlichkeit erfassen.	$n = 500;\ p = 0{,}018$ $k = 2$
Prognoseintervall berechnen.	$\mu = 500 \cdot 0{,}018 = 9\ ;\ \sigma \approx 2{,}97$ $9 - 2 \cdot 2{,}97 \le H \le 9 + 2 \cdot 2{,}97$ $3{,}06 \le H \le 14{,}94$
Prognoseintervall für H unter Beachtung der Ganzzahligkeit der Grenzen und Richtung der Ungleichheitszeichen in der Sigma-Regel angeben.	$4 \le H \le 14$

Probe: Es gilt für eine binomialverteilte Zufallsgröße X mit $n = 500$ und $p = 0{,}018$: $P(4 \le X \le 14) \approx 0{,}94 < 0{,}95$.

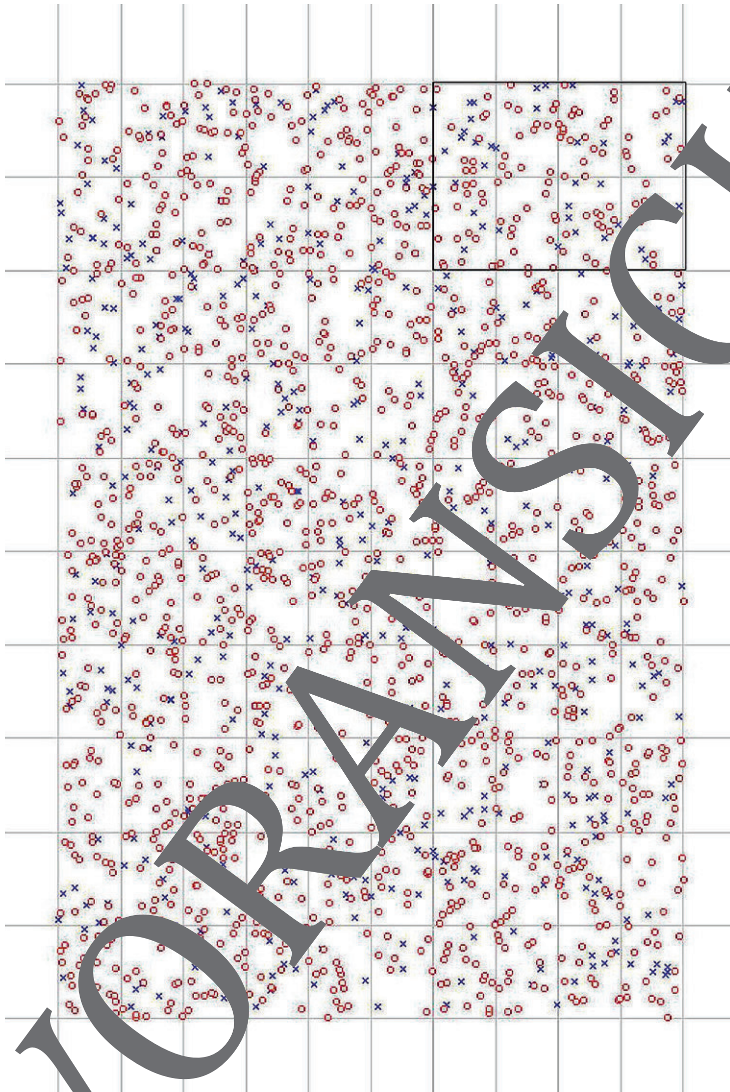
binomCdf(500,0.018,4,14)	0.939573
binomCdf(500,0.018,3,15)	0.973089

Das ist weniger als 95 %, man könnte deshalb das Prognoseintervall durch „Runden nach außen“ vergrößern, um die geforderte Sicherheitswahrscheinlichkeit von ca. 95 % zu erreichen: $P(3 \le X \le 15) \approx 0{,}97 > 0{,}95$. Die Anzahl der Pkw Mazda wird mit einer Sicherheitswahrscheinlichkeit von ca. 95 % im Intervall [3; 15] liegen.

Aufgaben

1. Für Gerstensamen wird eine Keimfähigkeit von 98 % versprochen. Ein Landwirt will dieses Versprechen durch eine Stichprobe überprüfen. Er entnimmt einem großen Sack Gerstensaatgut 100 Körner und sät sie aus.
   In welchem Intervall müsste die Anzahl der gekeimten Gerstenkörner liegen, ohne dass der Landwirt Zweifel an dem Versprechen zur Keimfähigkeit hegen muss, wenn die Sicherheitswahrscheinlichkeit ca. 95 % betragen soll?

2. Das Punktemuster auf der nächsten Seite enthält insgesamt eintausend Punkte, von denen 25 % als Kreuze markiert sind.
   a) Begründen Sie, dass das Modell der Binomialverteilung für die folgenden Untersuchungen näherungsweise verwendet werden kann.
   b) Wählen Sie eine Stichprobe von vier Rasterfeldern aus und ermitteln Sie für diese Stichprobe die Anzahl aller Punkte sowie die Anzahl der als Kreuze markierten Punkte.
   c) Bestimmen Sie für den beobachteten Wert von n und p = 0,25 das zugehörige 95 %-Prognoseintervall.
   d) Untersuchen Sie, ob das Stichprobenergebnis statistisch verträglich ist mit p = 0,25.



© Dr. Wilfried Zappe

M 4 Prognoseintervalle für relative Häufigkeiten

Theorie: Aus dem bisherigen Unterricht wissen die Schüler:

– Relative Häufigkeiten h schwanken zufallsbedingt um die theoretische Wahrscheinlichkeit p.
– Relative Häufigkeiten h lassen sich bei großem Stichprobenumfang für eine Punktschätzung einer Wahrscheinlichkeit p nutzen.

Diese Zusammenhänge sollen nun vertieft und quantifiziert werden. Aus der Doppelungleichung für die Prognoseintervalle für absolute Häufigkeiten H einer binomialverteilten Zufallsgröße erhält man nach Division durch den Stichprobenumfang n eine Doppelungleichung, die Prognoseintervalle für relative Häufigkeiten h beschreibt. Die zum jeweiligen k-Wert gehörende Wahrscheinlichkeit gibt die Sicherheitswahrscheinlichkeit für das Prognoseintervall an.

$$\mu - k \cdot \sigma \le H \le \mu + k \cdot \sigma$$

$$n \cdot p - k \cdot \sqrt{n \cdot p \cdot (1-p)} \le H \le n \cdot p + k \cdot \sqrt{n \cdot p \cdot (1-p)} \quad \Big| : n$$

$$p - k \cdot \sqrt{\frac{n \cdot p \cdot (1-p)}{n^2}} \le \frac{H}{n} \le p + k \cdot \sqrt{\frac{n \cdot p \cdot (1-p)}{n^2}}$$

Wegen $\frac{H}{n} = h$ ergibt sich daraus:

$$p - k \cdot \frac{\sqrt{p \cdot (1-p)}}{\sqrt{n}} \le h \le p + k \cdot \frac{\sqrt{p \cdot (1-p)}}{\sqrt{n}} \quad (*).$$

Damit lässt sich prognostizieren, in welchem Intervall bei einer Stichprobe vom Umfang n mit der gegebenen Sicherheitswahrscheinlichkeit die relative Häufigkeit h eines Merkmals liegen wird, wenn p die bekannte oder als wahr angenommene Wahrscheinlichkeit dieses Merkmals in der Gesamtheit ist.
Auch hier gilt: Liegt die beobachtete relative Häufigkeit h bei einer Stichprobe vom Umfang n nicht im 95 %-Prognoseintervall, dann gibt das Anlass, an p zu zweifeln. Man sagt: Die beobachtete relative Häufigkeit h weicht signifikant von der Wahrscheinlichkeit p ab. Liegt h außerhalb des 99,7 %-Prognoseintervalls, dann spricht man sogar von einer hochsignifikanten Abweichung. Liegt h im 95 %-Prognoseintervall, dann nennt man das Ergebnis der Stichprobe statistisch verträglich mit p.

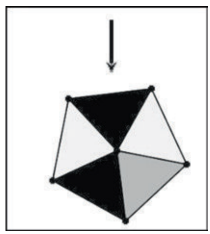
Die Länge $d(n) = p + 2 \cdot \frac{\sqrt{p \cdot (1-p)}}{\sqrt{n}} - \left(p - 2 \cdot \frac{\sqrt{p \cdot (1-p)}}{\sqrt{n}}\right) = 4 \cdot \frac{\sqrt{p \cdot (1-p)}}{\sqrt{n}}$

des 95 %-Prognoseintervalls ist die Differenz von oberer und unterer Intervallgrenze. Weil d(n) proportional zum Faktor $\frac{1}{\sqrt{n}}$ ist, spricht man auch vom „**Eins-durch-Wurzel-n-Gesetz**“.

Das lässt sich z. B. wie folgt deuten: Bei einer Vervierfachung des Stichprobenumfangs halbieren sich die Zufallsschwankungen. Das „Eins-durch-Wurzel-n-Gesetz“, also die Abhängigkeit von Prognoseintervallen vom Stichprobenumfang n bei festem Wert für p, lässt sich durch „**Wurzel-Trichterdiagramme**“ veranschaulichen (Aufgabe 4). Stellt man hingegen die Abhängigkeit der Prognoseintervalle von der zugrunde gelegten Wahrscheinlichkeit p bei festem n grafisch dar, erhält man „Ellipsendiagramme“ (Aufgabe 5). Solche Ellipsendiagramme bieten neben interessanten Aufgabenstellungen einen guten Zugang zum inhaltlichen Verständnis der Konfidenzintervalle.

Beispiel 7: Prognoseintervall für relative Häufigkeiten beim Glücksrad

Es wird behauptet, dass das Glücksrad die Farbe Schwarz mit der Wahrscheinlichkeit p = 0,4 liefert. Gesucht ist ein Prognoseintervall, das bei 80 Drehungen des Glücksrades die relative Häufigkeit für die Anzeige eines schwarzen Sektors mit einer Sicherheitswahrscheinlichkeit von 95 % enthält. Wie ist das Ergebnis zu interpretieren, wenn bei diesen 80 Drehungen 15-mal ein schwarzer Sektor angezeigt wurde?



© Dr. Wilfried Zappe

Vorgehensweise:

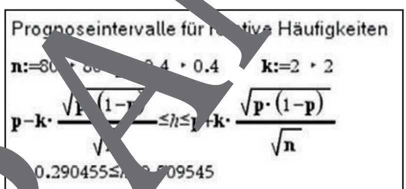
Entscheiden, ob dem Sachverhalt das Modell der Binomialverteilung zugrunde gelegt werden kann	Ja. Es sind zwei Ergebnisse möglich: „Schwarz“ oder „nicht Schwarz“ Die Ergebnisse sind unabhängig voneinander.
Kenngrößen der Binomialverteilung und den Wert k bezüglich der Sicherheitswahrscheinlichkeit erfassen	n = 80; $p = 0{,}4$ $k = 2$

Grenzen des 95 %-Prognoseintervalls berechnen.	$p-k\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{n}}\leq h\leq p+k\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{n}}$ $0{,}4-2\cdot\frac{\sqrt{0{,}4\cdot 0{,}6}}{\sqrt{80}}\leq h\leq 0{,}4+2\cdot\frac{\sqrt{0{,}4\cdot 0{,}6}}{\sqrt{80}}$ $0{,}2905\leq h\leq 0{,}5095$

Der Anteil aller Drehungen, bei denen ein schwarzer Sektor gezeigt wird, wird bei 80 Drehungen mit einer Sicherheitswahrscheinlichkeit von ca. 95 % im Intervall [0,2905; 0,5095] liegen.

Wurde bei 80 Drehungen 15-mal ein roter Sektor angezeigt, dann ist die relative Häufigkeit dafür $\frac{15}{80}\approx 0{,}19$. Dieser Wert liegt nicht im Prognoseintervall und gäbe Anlass, an der Behauptung p = 0,4 zu zweifeln.

Hinweis: Für Nutzer des TI-Nspire ist die Erstellung und Abspeicherung eines „Notes"-Arbeitsblattes zu empfehlen, das bei Bedarf nur aufgerufen und angepasst werden muss.



Beispiel 8: Prognoseintervall für „Ziehen ohne Zurücklegen"

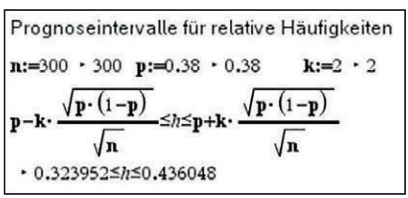
Nach Auskunft des Statistischen Jahrbuchs 2018 verfügten im Jahr 2017 38 % der ca. 37 Millionen privaten Haushalte in Deutschland über genau zwei Mobiltelefone. Zwei Jahre später wurde bei einer Umfrage unter 300 repräsentativ ausgewählten Haushalten festgestellt, dass 160 von ihnen über genau zwei Mobiltelefone verfügten. Würde man auf der Grundlage eines 95 %-Prognoseintervalls an eine Veränderung des Prozentsatzes von 38 % glauben wollen oder eher an eine zufallsbedingte Schwankung?

Vorgehensweise:

Entscheiden, ob dem Sachverhalt das Modell der Binomialverteilung zugrunde gelegt werden kann.	Nur näherungsweise, denn eine „Umfrage" entspricht dem Modell „Ziehen ohne Zurücklegen", aber die Grundgesamtheit N ist wesentlich größer als die Stichprobe n mit $N \approx 37 \cdot 10^6$ und $n = 300$. Ergebnisse: „Haushalt mit genau zwei Mobiltelefonen" oder „Haushalt mit weniger oder mehr als zwei Mobiltelefonen".
Kenngrößen der Binomialverteilung und den Wert für k bezüglich der Sicherheitswahrscheinlichkeit erfassen.	$n = 300$; $p = 0{,}38$; $k = 2$
Grenzen des 95 %-Prognoseintervalls berechnen.	Untere Grenze: $0{,}38 - 2 \cdot \frac{\sqrt{0{,}38 \cdot 0{,}62}}{\sqrt{300}} \approx 0{,}3240$ Obere Grenze: $0{,}38 + 2 \cdot \frac{\sqrt{0{,}38 \cdot 0{,}62}}{\sqrt{300}} \approx 0{,}4360$ $0{,}33 \le h \le 0{,}43$

Der Anteil der Privathaushalte mit genau zwei Mobiltelefonen wird mit einer Sicherheitswahrscheinlichkeit von ca. 95 % im Intervall [0,324; 0,436] liegen. Weil $\frac{160}{300} \approx 0{,}53$ außerhalb dieses Prognoseintervalls liegt, ist es berechtigt anzunehmen, dass sich der Anteil der Haushalte mit genau zwei Mobiltelefonen verändert hat.

Der Rechenaufwand wird verringert, wenn man das im vorigen Beispiel erwähnte Notes-Rechenblatt aufruft und nur die Angaben für n und p ändert. Die Eingaben müssen jeweils mit Enter abgeschlossen werden.



Aufgaben

1. Es wird behauptet, dass in einer Mischung Studentenfutter der Anteil der Rosinen bei 30 % liegt. Aus einer Tüte mit 204 g Inhalt wurde mit einem Löffel eine Stichprobe entnommen, die 21 g wog. Die Stichprobe enthielt acht Rosinen und 37 Nüsse. Überprüfen Sie, ob der Anteil der Rosinen in dieser Stichprobe rechnerisch mit $p = 0{,}3$ mit einer Sicherheitswahrscheinlichkeit von 95 % statistisch verträglich ist. Diskutieren Sie, ob hier das Modell der Binomialverteilung sinnvoll angewendet werden kann.

© Lauri Patterson/E+/Getty Images Plus

2. Bei einer Umfrage unter 5046 Personen kurz vor einer Wahl hatte eine Partei einen Stimmenanteil von 4,6 %. Kann die Partei damit rechnen, dass der Stimmenanteil unter der wahlberechtigten Bevölkerung die 5 %-Marke mit einer Sicherheitswahrscheinlichkeit von 68,3 % erreicht? Untersuchen Sie die gleiche Fragestellung, wenn eine Sicherheitswahrscheinlichkeit von 99,7 % gewählt wird.

3. Der Buchstabe e kommt in deutschsprachigen Texten mit $p = 0{,}174$ vor. (Die Umlaute ä, ö und ü werden wie ae, oe und ue gezählt, ß als eigenständiges Zeichen.[3]) Ermitteln Sie die relative Häufigkeit des Buchstabens e in den folgenden Texten, die im Internet den Autoren George Orwell, Wilhelm Busch und Arne Arotnow zugeordnet werden. Prüfen Sie, ob die Ergebnisse signifikant von $p = 0{,}174$ abweichen.

(1) Je weiter sich eine Gesellschaft von der Wahrheit entfernt, desto mehr wird sie jene hassen, die die Wahrheit aussprechen.

Zusatz: Welcher Text gehört zu welchem Schriftsteller?

(2) Auch Wasser wird zum edlen Tropfen, mischt man es mit Malz und Hopfen.

(3) Erde Ernten erbringt.
Erde Eichen erhebt.
Erde Elbe entspringt.
Erde Erdgas entschwebt.

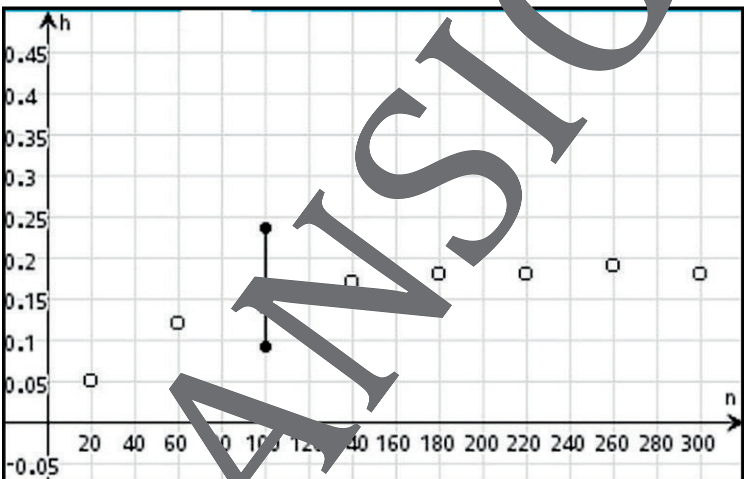
3 https://de.wikipedia.org/wiki/Buchstabenhäufigkeit

4. Ein Bernsteinwürfel hat Einschlüsse, sodass man nicht weiß, ob alle Seiten gleichwahrscheinlich mit $p = \frac{1}{6}$ sind.

Ein solcher Würfel wurde n-mal geworfen. Dabei wurden die relativen Häufigkeiten h der „Sechsen" ermittelt und in ein Diagramm eingetragen.

© Dr. W. Zappe

n	20	60	100	140	180	220	260	300
h	0,05	0,12	0,14	0,17	0,18	0,18	0,19	0,18



© Dr. Wilfried Zappe

a) Ermitteln Sie für jeden Wert von n das 95 %-Prognoseintervall der relativen Häufigkeiten zur angenommenen Wahrscheinlichkeit $p = \frac{1}{6}$.

b) Zeichnen Sie die Prognoseintervalle als vertikale Strecken in das Diagramm ein, analog zum abgebildeten Beispiel für n = 100.

c) Begründen Sie, dass die Grenzen der Prognoseintervalle auf den Graphen der Funktionen

$$g(x) = \frac{1}{6} + 2 \cdot \frac{\sqrt{\frac{5}{36}}}{\sqrt{x}} \text{ bzw. } f(x) = \frac{1}{6} - 2 \cdot \frac{\sqrt{\frac{5}{36}}}{\sqrt{x}} \text{ liegen.}$$

Skizzieren Sie die Graphen dieser Funktionen in das Diagramm ein. Warum nennt man wohl ein solches Diagramm auch ein „Wurzel-Trichter-Diagramm"?

d) Beurteilen Sie, welche der folgenden Aussagen in Bezug auf die im Würfelversuch ermittelten Ergebnisse wahr sind. Geben Sie eine kurze Begründung an.

(1) Die Wahrscheinlichkeit, mit dem Bernsteinwürfel eine „Sechs" zu werfen, liegt bei p = 0,18.

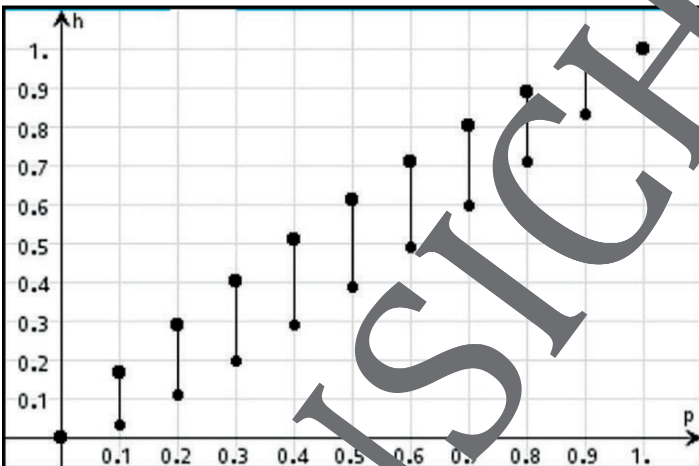
(2) Die Wahrscheinlichkeit, dass mit dem Bernsteinwürfel mit Wahrscheinlichkeit $p = \frac{1}{6}$ eine „Sechs" geworfen wird, kann man mit einer Sicherheitswahrscheinlichkeit von ca. 95 % nicht in Zweifel ziehen.

(3) Die Länge d(n) des 95 % Prognoseintervalls für relative Häufigkeiten für eine feste Wahrscheinlichkeit p wird berechnet durch $d(n) = 4 \cdot \frac{\sqrt{p \cdot (1-p)}}{\sqrt{n}}$.

(4) Will man die Länge des Prognoseintervalls für relative Häufigkeiten bei fester Wahrscheinlichkeit p halbieren, so muss man den Stichprobenumfang verdoppeln.

(5) Will man die Länge des Prognoseintervalls für relative Häufigkeiten bei fester Wahrscheinlichkeit p halbieren, so muss man den Stichprobenumfang vervierfachen.

5. In dem abgebildeten „Ellipsen-Diagramm“ sind für n = 80, k = 2 und die Wahrscheinlichkeiten p von 0; 0,1; 0,2; ...; 0,9; 1 die zugehörigen 95 %-Prognoseintervalle der relativen Häufigkeiten grafisch dargestellt.



© Dr. Wilfried Zappe

a) Lesen Sie aus dem Diagramm näherungsweise das Prognoseintervall zu p = 0,3 ab. Überprüfen Sie durch eine Berechnung Ihr Ableseergebnis.
b) Entscheiden Sie anhand der grafischen Darstellung, ob die relative Häufigkeit h = 0,35 signifikant von p = 0,2 abweicht.
c) Geben Sie anhand der grafischen Darstellung mindestens drei relative Häufigkeiten an, die mit p = 0,8 statistisch verträglich sind.
d) Lesen Sie aus dem Diagramm ein Intervall für alle Wahrscheinlichkeiten p ab, die mit der relativen Häufigkeit h = 0,3 statistisch verträglich sind.
e) Welche Veränderungen an der grafischen Darstellung vermuten Sie, wenn diese statt für n = 80 für n = 320 erstellt wird?