Konfidenzintervalle

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

© *mdaake/Adobe Stock*

Wie viel Prozent der Wähler werden sich bei der nächsten Bundestagswahl für die FDP entscheiden? Wie hoch wird die Wahlbeteiligung ausfallen? Solchen und ähnlichen Fragen können Ihre Schülerinnen und Schüler in diesem Beitrag statistisch auf den Grund gehen. Mit bekannten relativen Häufigkeiten aus Stichproben (etwa Umfragen) berechnen die Lernenden Konfidenzintervalle oder bestimmen bei bekannten Wahrscheinlichkeiten die zugehörigen Prognoseintervalle. Neben den genauen Formeln bietet dieser Beitrag auch Näherungsformeln und Abschätzungen an, die auch ohne einen CAS-Rechner bestimmt werden können.

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: mdaake/Adobe Stock
Lektorat: Maria Hitzenauer, Regensburg
Korrektorat: Johanna Stotz, Wyhl a. K.

Konfidenzintervalle

Oberstufe (erhöhtes Niveau)

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

Die Schüler lernen:

den Begriff der Prognoseintervalle anzuwenden, um Konfidenzintervalle zu bestimmen, Konfidenzintervalle mit Hilfe eines CAS-Rechners zu berechnen und zu interpretieren, eine Näherungsformel für Konfidenzintervalle kennen, Stichprobenumfänge abzuschätzen, eine „Faustformel“ kennen, mit deren Hilfe Prognose- oder Konfidenzintervalle sowie Stichprobenumfänge abgeschätzt werden können.

Überblick:

Legende der Abkürzungen:
Ab = Arbeitsblatt, **LEK** = Lernerfolgskontrolle, **Wh** = Wiederholung

Thema	Material	Methode
Sigma-Regeln der Binomialverteilung – frischen Sie Ihr Wissen auf!	M 1	Ab, Wh
Prognoseintervalle – Wiederholung	M 2	Ab, Wh
Binomialverteilung näherungsweise anwenden	M 3	Ab
Konfidenzintervalle – eine Einführung	M 4	Ab
Eine Näherungsformel für Konfidenzintervalle	M 5	Ab
Stichprobenumfang abschätzen	M 6	Ab
Eine „Faustformel“	M 7	Ab
Sind Sie fit? – Testen Sie Ihr Wissen!	M 8	Ab, LEK

Erklärung zu Differenzierungssymbolen

einfaches Niveau	mittleres Niveau	schwieriges Niveau
	Dieses Symbol markiert Zusatzaufgaben.	

Kompetenzprofil:

Inhalt: Binomialverteilung, Sigma-Regeln, Prognose- und Konfidenzintervalle, Konfidenzellipse, Näherungsformel und Abschätzung für Konfidenzintervalle

Medien: GTR/CAS

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

Konfidenzintervalle – Hinweise

Bei der Bundestagswahl sind alle Wahlberechtigten in Deutschland aufgerufen, sich für eine der zugelassenen Parteien zu entscheiden. Zwischen den Wahlen hingegen ermitteln Meinungsforschungsinstitute mit der sogenannten „Sonntagsfrage" die aktuelle politische Stimmung in Deutschland, indem sie einige tausend repräsentativ ausgewählte Wahlberechtigte stichprobenartig befragen. Repräsentativ bedeutet in diesem Zusammenhang, dass die Gruppe der Befragten möglichst gut die Bevölkerung in Deutschland widerspiegelt. Die befragten Personen einer solchen Umfrage bilden eine zufällige Stichprobe aus der Gesamtheit aller Wahlberechtigten.

Ähnliche Ergebnisse von stichprobenartigen Umfragen gibt es natürlich auch zu anderen Themen. Fast täglich kann man sie den Medien entnehmen.

Allgemein lässt sich die Ausgangssituation so beschreiben: Man möchte wissen, mit welcher Wahrscheinlichkeit p ein bestimmtes Merkmal einer Zufallsgröße in einer Gesamtheit vorkommt. Der Umfang der Gesamtheit ist aber zu groß, um diesen Anteil direkt zu messen. Deshalb wird eine Stichprobe erhoben, in der sich die relative Häufigkeit des bestimmten Merkmals leicht berechnen lässt. Diese relative Häufigkeit ist aber nur eine Punktschätzung für die unbekannte Wahrscheinlichkeit. Die Ergebnisse einer Stichprobe tragen Zufallscharakter, daher kann man die unbekannte Wahrscheinlichkeit besser durch ein Intervall schätzen. Dieses bestimmt man mithilfe einer Punktschätzung.

Die Ermittlung und Interpretation solcher „Konfidenzintervalle" ist Gegenstand dieses Beitrages. Er beschränkt sich auf exakt oder näherungsweise binomialverteilte Zufallsgrößen. Die Kenntnisse über Prognoseintervalle aus einem früheren Beitrag[1] dieser Reihe sind eine große Hilfe beim Verständnis der Konfidenzintervalle.

1 vgl. R0491 201005 Prognoseintervalle mit CAS-Rechner (U.6.8; EL 67; Okt. 2020)

M 1 Sigma-Regeln der Binomialverteilung – frischen Sie Ihr Wissen auf!

Im Folgenden sollten Sie einige Sachverhalte wiederholen, die für das Verständnis der Ermittlung und Interpretation von Konfidenzintervallen wichtig sind. Eine ausführlichere Darstellung dazu finden Sie in dem Beitrag „Prognoseintervalle" der Raabe Unterrichtsmaterialien.

Die Sigma-Regeln bringen zum Ausdruck, wie viel Prozent der Werte einer binomialverteilten Zufallsgröße X näherungsweise in einem vorgegebenen Vielfachen der Sigma-Umgebung des Erwartungswertes liegen.

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 n 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 nachkommenden Untersuchungen zu Prognose- und Konfidenzintervallen verwenden Sie vorwiegend die folgende Näherung der Zwei-Sigma-Regel:

Für genügend große Werte von n liegen ca. 95 % der Werte einer Binomialverteilung in der Zwei-Sigma-Umgebung ihres Erwartungswertes.[2]

$P(\mu - 2\cdot\sigma \le X \le \mu + 2\cdot\sigma) \approx 0{,}95$

Hinweis: Alles, was Sie in diesem Zusammenhang für k = 2 lernen, können Sie problemlos auf andere Werte von k übertragen.

2 Häufig liest man auch von der Bedingung, dass $\sigma > 3$ erfüllt sein muss.

M 2 Prognoseintervalle – Wiederholung

Aus den Doppelungleichungen der Sigma-Regeln können Sie ebenso Doppelungleichungen für Prognoseintervalle absoluter Häufigkeiten H und Prognoseintervalle relativer Häufigkeiten h herleiten. Die zum jeweiligen k-Wert gehörende Wahrscheinlichkeit gibt die Sicherheitswahrscheinlichkeit für das Prognoseintervall an.

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

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

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

$$p - k \cdot \frac{\sqrt{p \cdot (1-p)}}{\sqrt{n}} \leq h \leq p + k \cdot \frac{\sqrt{p \cdot (1-p)}}{\sqrt{n}} \qquad (*)$$

Mit diesen Doppelungleichungen können Sie prognostizieren, in welchem Intervall bei einer Stichprobe vom Umfang n mit der vorgegebenen Sicherheitswahrscheinlichkeit die absolute Häufigkeit H bzw. die relative Häufigkeit h eines bestimmten Merkmals liegen wird, wenn p die **bekannte oder als wahr angenommene Wahrscheinlichkeit** dieses Merkmals in der Gesamtheit ist.

Liegt die beobachtete absolute bzw. relative Häufigkeit bei einer Stichprobe vom Umfang n nicht im 95 %-Prognoseintervall, dann gibt das Anlass, an p zu zweifeln. Man sagt: Die beobachtete absolute/relative Häufigkeit weicht signifikant von der Wahrscheinlichkeit p ab. Liegt die absolute/relative Häufigkeit außerhalb des 99,7 %-Prognoseintervalls, dann spricht man sogar von einer hochsignifikanten Abweichung. Liegt die absolute/relative Häufigkeit allerdings im 95 %-Prognoseintervall, dann nennt man das Ergebnis der Stichprobe statistisch verträglich mit p.

M 3 Binomialverteilung näherungsweise anwenden

Wenn die Anzahl N der Elemente einer **hypergeometrisch verteilten Gesamtheit** sehr viel größer ist als der Stichprobenumfang n ($N \geq 20n$), können Sie bei Zufallsstichproben, die Sie eigentlich durch das Urnenmodell „Ziehen ohne Zurücklegen" modellieren müssten, näherungsweise das mathematische Modell der Binomialverteilung verwenden.

Aufgabe

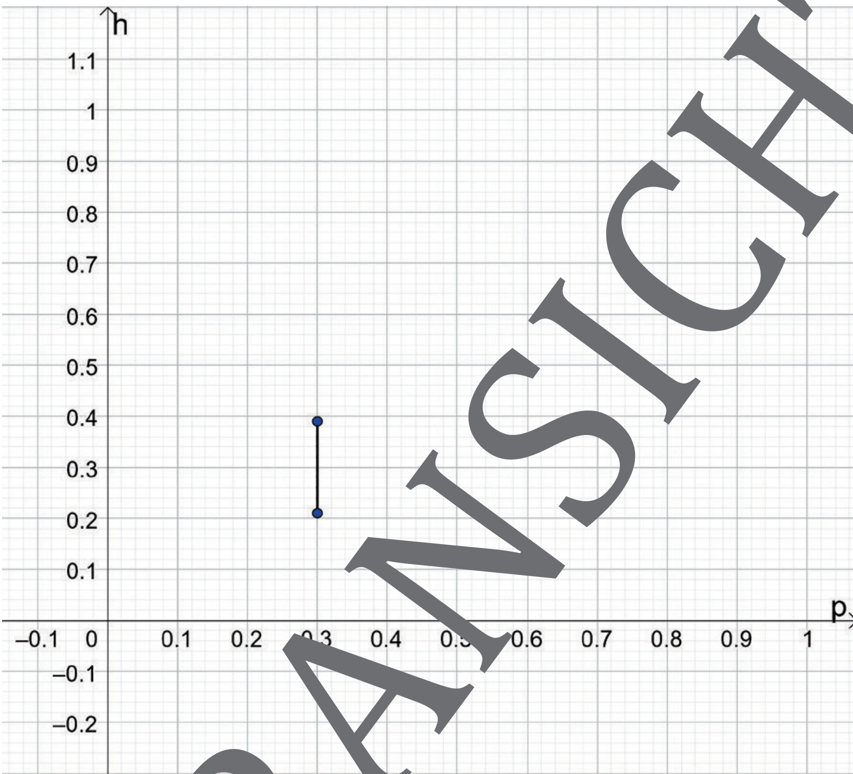
1. Im nachstehenden Diagramm (siehe Seite 5) ist das 95 %-Prognoseintervall für p = 0,3 und n = 100 grafisch dargestellt.
   a) Prüfen Sie, ob die Darstellung korrekt ist.
   b) Ergänzen Sie die fehlenden Werte in der Tabelle sowie die Darstellungen für 95 %-Prognoseintervalle mit n = 100 und p = 0; 0,1;…; 1.

Wahrscheinlichkeit p	0,0	0,1	0,2	0,3	0,4
obere Grenze für h				0,39	
untere Grenze für h				0,21	
p bezweifeln (h = 0,6)?				ja	

Wahrscheinlichkeit p	0,5	0,6	0,7	0,8	0,9	1,0
obere Grenze für h						
untere Grenze für h						
p bezweifeln (h = 0,6)?						

   c) Ermitteln Sie, welche dieser Prognoseintervalle mit der relativen Häufigkeit h = 0,6 statistisch verträglich sind.
   d) Bestimmen Sie grafisch ein Intervall für alle diejenigen Wahrscheinlichkeiten, die man nicht bezweifeln muss, wenn man in der Stichprobe die relative Häufigkeit h = 0,6 beobachtet.

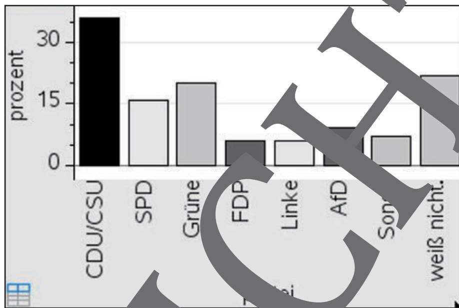
Diagramm zu Aufgabe 1



© Dr. W. Zappe

M 4 Konfidenzintervalle – eine Einführung

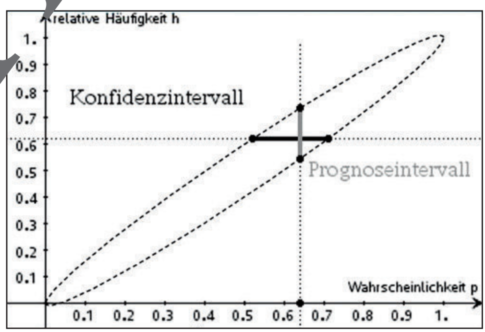
Mit der Sonntagsfrage ermitteln Meinungsforschungsinstitute zwischen den Wahlen die aktuelle politische Stimmung in Deutschland[3]. Man möchte wissen, mit welcher Wahrscheinlichkeit p der Stimmenanteil einer Partei in der Gesamtheit aller Wahlberechtigten in Deutschland vorkommt. Der Umfang dieser Gesamtheit erlaubt es aber aus Kosten- und Zeitgründen nicht, diesen Anteil durch Befragung aller Wählerinnen und Wähler jeden Sonntag zu ermitteln. Deshalb erhebt man eine Stichprobe von einigen Tausend dieses Personenkreises und schätzt die Zustimmung zu einer Partei in der Gesamtheit durch die relative Häufigkeit der Stichprobe (Punktschätzung). Für die relative Häufigkeit zählen nur die Antworten auf die Frage „Welche Partei würden Sie wählen, wenn am kommenden Sonntag Bundestagswahl wäre?“ aus.



© Dr. W. Zappe

Mithilfe der bereits bekannten Doppelungleichung von Prognoseintervallen für relative Häufigkeiten $p - k\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{n}} \le h \le p + k\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{n}}$ kann man dann auf ein Intervall schließen, das die **unbekannte Wahrscheinlichkeit** p in der Gesamtheit mit einer durch k festgelegten Sicherheitswahrscheinlichkeit überdeckt.

Anstelle des Begriffs „Sicherheitswahrscheinlichkeit“ wird hier auch gern vom „Konfidenzniveau“ gesprochen.



© Dr. W. Zappe

3 Quelle: www.wahlrecht.de/umfragen/forsa.htm (zuletzt aufgerufen am 30.09.2020) es wurden 2501 Personen befragt im Zeitraum vom 11.08 bis 14.08.2020

Die Berechnung des Intervalls für p wird mithilfe eines CAS-Rechners sehr einfach. Ohne CAS-Rechner können Sie p auch mitunter durch geeignete Näherungsverfahren bestimmen. Zu einem Konfidenzintervall zu einer beobachteten relativen Häufigkeit h gehören alle diejenigen Wahrscheinlichkeiten p, die nach der Beobachtung von h nicht in Zweifel gezogen werden müssen. Das sind also genau diejenigen Wahrscheinlichkeiten p, in deren Prognoseintervall die relative Häufigkeit h liegt.

Beispiel – Berechnung eines Konfidenzintervalls

Bei der Sonntagsfrage gaben 900 der 2501 befragten Wahlberechtigten an, dass sie die CDU/CSU wählen würden. Ermitteln Sie ein Konfidenzintervall für den Anteil der CDU/CSU-Wähler in Deutschland mit einem Konfidenzniveau von 95 %.

Vorgehensweise:

Entscheiden, ob das Modell der Binomialverteilung zugrunde gelegt werden kann.	Nur näherungsweise, denn die Befragung entspricht „Ziehen ohne Zurücklegen“, aber es ist N >> n (N: ca. 60 Mio. Wahlberechtigte, n = 2501)
Umfang n der Stichprobe, die relative Häufigkeit h und den Wert für k bezüglich der Sicherheitswahrscheinlichkeit [illegible]	$n = 2501;\quad h = \frac{900}{2501} \approx 0{,}36;\quad k = 2$
Grenzen des 95 %-Konfidenzintervalls berechnen (am besten im CAS-Rechner eine Routine erstellen und abspeichern).	$\mathbf{n}:=2501 \blacktriangleright 2501$ $\mathbf{h}:=\frac{900}{2501.} \blacktriangleright 0.359856$ $\mathbf{k}:=2 \blacktriangleright 2$ $\text{solve}\left(p-\mathbf{k}\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{\mathbf{n}}}\le\mathbf{h}\le p+\mathbf{k}\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{\mathbf{n}}},p\right)$ $\blacktriangleright 0.340899\le p\le 0.37926$ *© Dr. W. Zappe*
Konfidenzintervall sinnvoll gerundet unter Beachtung der Richtung der Ungleichheitszeichen und unter Bezug auf den vorliegenden Sachverhalt angeben.	Das Intervall [0,35; 0,37] überdeckt den wahren Anteil der CDU/CSU-Wähler unter den Wahlberechtigten in Deutschland auf einem Konfidenzniveau von ca. 95 %.

Beispiel – Konfidenzellipse

Wie kommt man eigentlich zu nebenstehender Aussage[4], die sich auch so lesen lässt, als hätten 62 % aller Amerikaner diese Meinung geäußert?

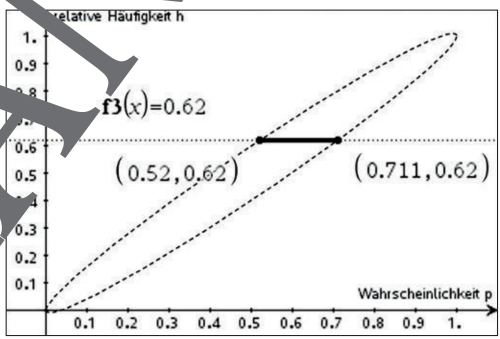
> FACTANK (05.01.2018): Ungefähr sechs von zehn Amerikanern (62 %) gaben an, dass der Konsum von Marihuana legalisiert werden sollte.

Wir dürfen annehmen, dass die Aussage auf dem Ergebnis einer Umfrage unter einer Auswahl von US-Bürgern beruht, denn es gibt mehr als 330 Millionen Einwohner in den USA, die bestimmt nicht alle befragt wurden. Wegen N >> n können wir das Modell der Binomialverteilung näherungsweise anwenden, obwohl eigentlich das Modell „Ziehen ohne Zurücklegen“ vorliegt.

Angenommen, die Stichprobe hatte einen Umfang von n = 100. Wir können eine Wahrscheinlichkeit p, die die Doppelungleichung (*) für n = 100, k = 2 und h = 0,62 mit einer Vertrauenswahrscheinlichkeit von 95 % erfüllt, sowohl durch Rechnung wie im vorigen Beispiel als auch auf grafischem Wege mithilfe eines Ellipsendiagramms („Konfidenzellipse“), ermitteln:

Die Ellipsenbögen werden durch die beiden Funktionen $f(p) = p \pm 2 \cdot \frac{\sqrt{p \cdot (1-p)}}{\sqrt{100}}$ gebildet.

Die Funktionswerte dieser Funktionen geben die untere bzw. obere Grenze zu den Prognoseintervallen an, die zu einer gegebenen Wahrscheinlichkeit gehören. Das 95 %-Konfidenzintervall entspricht dem Teil der Geraden h = 0,62, den die **Konfidenzellipse aus** dieser Geraden „herausschneidet“.



© Dr. W. Zappe

---

[4] https://translate.google.com/translate?hl=de&sl=en&u=https://www.pewresearch.org/fact-tank/2018/10/08/americans-support-marijuana-legalization/&prev=search (zuletzt aufgerufen am [illegible].2020)

Dieser Teil enthält alle Prognoseintervalle, deren Wahrscheinlichkeiten mit h = 0,6[illegible] statistisch verträglich sind. Es gilt näherungsweise $0{,}52 \leq p \leq 0{,}71$.

Ergebnis: Das Intervall [0,52; 0,71] überdeckt die unbekannte Wahrscheinlichkeit p für den Anteil der Befürworter einer Legalisierung von Marihuana in der amerikanischen Bevölkerung auf einem Konfidenzniveau von 95 %.

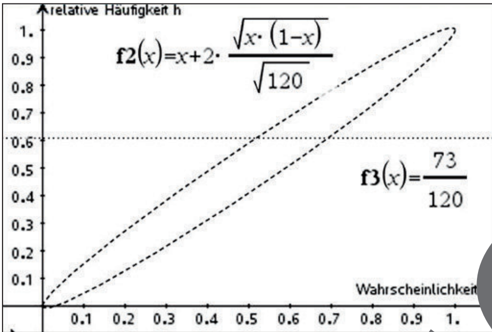
Hinweis: Eine Formulierung wie „Das Intervall [0,52; 0,71] enthält die unbekannte Wahrscheinlichkeit p mit einer Wahrscheinlichkeit von ca. 95 %" ist inhaltlich falsch. Die verwendete Vertrauenswahrscheinlichkeit von 0,95 ist so zu deuten, dass im Durchschnitt in etwa einer von 20 unabhängigen Wiederholungen des gesamten Experiments (also in ca. 5 % der Fälle) ein Konfidenzbereich konstruiert wird, der den wahren Wert für p nicht enthält. In der Praxis kommt es aber i. A. nur zu genau einer Durchführung des Experiments, der dazu berechnete Konfidenzbereich enthält entweder p oder eben nicht.

Aufgaben

2. Beim Würfeln unter gleichbleibenden Bedingungen mit einem 2x2-Lego-Baustein wird das Ereignis A „Eine der vier gleichgroßen Seitenflächen liegt oben" betrachtet. Eine theoretische Wahrscheinlichkeit p für dieses Ereignis sei unbekannt. In einer Stichprobe von 360 Würfen ergab sich 109-mal das Ereignis A. Gesucht ist das zugehörige 95 %-Konfidenzintervall für die unbekannte Wahrscheinlichkeit p.

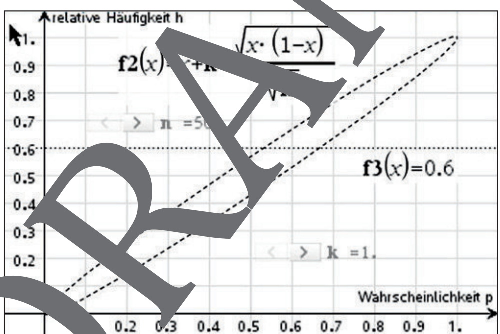
© Gearstd / iStock/Getty Images Plus

3. Es soll durch eine Stichprobe abgeschätzt werden, wie beliebt eine neue Fernsehserie ist. Zu diesem Zweck werden 120 Zuschauer befragt, von denen [illegible] angeben, die Serie regelmäßig einzuschalten.
   a) Erläutern Sie in diesem Sachzusammenhang die Abbildung auf der nächsten Seite.
   b) Beschreiben Sie, wie Sie das zugehörige Konfidenzintervall mithilfe der grafischen Darstellung ermitteln können.
   c) Überprüfen Sie Ihr Ergebnis durch eine Rechnung.



© Dr. W. Zappe

4. Erzeugen Sie auf Ihrem CAS-Rechner eine Konfidenzellipse, in der Sie die Parameter n und k mit Schiebereglern verändern können.
   a) Geben Sie die Länge des Konfidenzintervalls und die Werte für h, n und k an, die Sie der unten stehenden Abbildung entnehmen können.
   b) Experimentieren Sie mit den Schiebereglern und beschreiben Sie die Auswirkungen einer Vergrößerung von n bzw. k auf die Länge des Konfidenzintervalls für eine relative Häufigkeit h = 0,6.
   c) Erklären Sie Ihre Beobachtungen auch anhand der Doppelungleichung (*).



© Dr. W. Zappe

5. Für die Umfrage zur Legalisierung von Marihuana, die im Beispiel thematisiert ist, wurden sogar 1754 Personen befragt. Ermitteln Sie für diesen Wert ein Konfidenzintervall auf einem Vertrauensniveau von 95 % und vergleichen Sie das Ergebnis mit dem Resultat im Beispiel.

6. Das Muster enthält Kreuze und Punkte. Zählen Sie in einer Stichprobe, die mindestens vier Karos des Gesamtfeldes umfasst, die Anzahlen der Kreuze und Punkte und ermitteln Sie anhand der relativen Häufigkeit der Kreuze in dieser Stichprobe ein Konfidenzintervall für den Anteil der Kreuze im gesamten Muster mit einer Vertrauenswahrscheinlichkeit von 95 %.

© D[illegible] appe

7. Zwei Wochen vor der Landtagswahl 2019 in Thüringen wurden Ergebnisse zweier Umfragen veröffentlicht, die von ARD und ZDF in Auftrag gegeben wurden (Ergebnisse siehe nebenstehende Abbildung[5]).

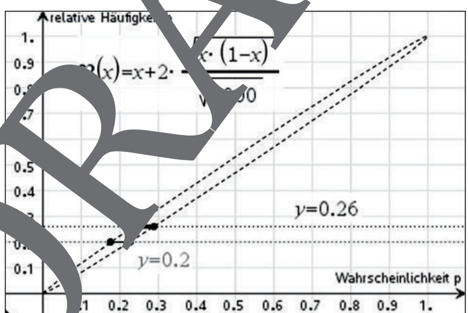
Wen würden Sie wählen, wenn Sonntag Landtagswahl wäre?		
Partei	**ZDF**	**ARD**
CDU	26 %	[illegible] %
Die Linke	27 [illegible]	[illegible] %
SPD	9 %	8 %
AfD	[illegible] %	[illegible] %
Grüne	8 %	7 [illegible]
FDP	5 %	4 %
Sonstige	[illegible] %	4 %

Quelle: Forschungsgruppe Wahlen (ZDF), ARD (Infratest Dimap), je 1000 Befragte in Thüringen zwischen 14. und 16.10.2019.

a) Erläutern Sie, weshalb zwei Umfragen zum gleichen Thema und zur gleichen Zeit zu verschiedenen Ergebnissen führen können.

b) Ermitteln Sie rechnerisch für ein Vertrauensniveau von 95 % die Konfidenzintervalle zu den Umfrageergebnissen von ZDF und ARD bezüglich CDU, Linke und FDP im Vergleich. Interpretieren Sie die folgenden Aussagen:
- Die Partei „Die Linke" kann sich aufgrund beider Umfragen sicher sein, bei der Landtagswahl ganz vorn zu liegen.
- Die FDP kann aus der ZDF-Umfrage schlussfolgern, die 5 %-Hürde bei der Landtagswahl zu erreichen, jedoch zeigt die ARD-Umfrage, dass das nicht möglich ist.

c) Interpretieren Sie die unten stehende Abbildung im gegebenen Sachzusammenhang.



© [illegible] W. Zappe

d) Begründen Sie, warum die Schwankungsbreiten der Konfidenzintervalle bei den „kleineren Parteien" geringer ausfallen als bei den „großen Parteien".

5 „Thüringer Landeszeitung" vom 18.10.2019

8. Nachstehend wird in einem Lückentext an einem Beispiel demon…trie… wie man ein Konfidenzintervall ohne CAS-Rechner, nur mit einem ü…lichen wissenschaftlichen Taschenrechner lösen kann. Ergänzen Sie d…n Lückentext und vergleichen Sie die so gefundene Lösung mit dem Resulta… das der CAS-Rechner liefert.
Eine Stichprobe vom Umfang n = 100 liefert eine relative H…figkei… des beobachteten Merkmals von h = 0,1. Berechnen Sie das Konfiden…tervall für die unbekannte Wahrscheinlichkeit p dieses Merkm…s in der Gesam…it auf einem Vertrauensniveau von 95 %, ohne ein CAS z… verwenden.

Ansatz mit Doppelungleichung:

$$p-2\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{100}}\leq 0{,}1\leq p+2\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{100}}$$

Für die Randwerte der Ungleichung gilt: $p\pm 2\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{100}}=0{,}1$.

Subtrahieren von p liefert: $\pm 2\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{100}}=0{,}1-p$.

Quadrieren beider Seiten:

____

Ausmultiplizieren: $0{,}04p-0{,}04p^2=0{,}01-0{,}2p+p^2$

Normalf… …rstellen:

____

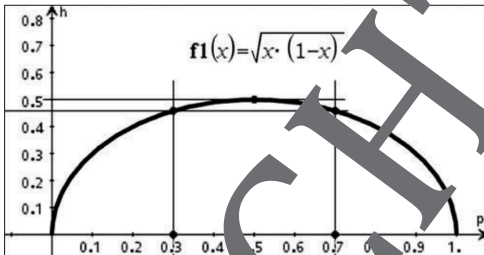
…rsformel a… …nden:

____

Er…bnis: $p_1\approx 0{,}054$ und $p_2\approx$

____

M 5 Eine Näherungsformel für Konfidenzintervalle



© Dr. W. Zappe

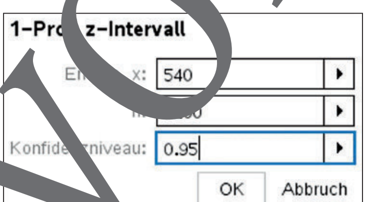
Die rechnerische Bestimmung des Konfidenzintervalls ohne CAS ist ziemlich aufwendig. Wenn wir uns mit einer Näherungslösung zufriedengeben, wird der Rechenaufwand wesentlich geringer. Zudem werden wir sehen, dass diese Näherungsformel auch von Anwendungen verwendet wird. Für die Näherung gehen wir davon aus, dass sich die unbekannte Wahrscheinlichkeit des Merkmals in der Gesamtheit bei genügend großem Stichprobenumfang von der ermittelten relativen Häufigkeit im Intervall $0,3 \le p \le 0,7$ nur wenig unterscheidet.

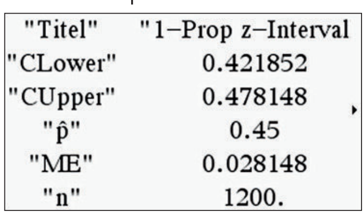
Der Graph von $f(p)=\sqrt{p\cdot(1-p)}$ verläuft in diesem Intervall ziemlich flach (siehe Abbildung oben). Wenn wir also p durch einen Näherungswert h mit $0,3 \le h \le 0,7$ ersetzen, führt das nur zu einem geringen Unterschied zwischen den Funktionswerten.

Wird in der Doppelungleichung $p-k\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{n}} \le h \le p+k\cdot\frac{\sqrt{p\cdot(1-p)}}{\sqrt{n}}$ unter der Wurzel p durch h ersetzt, so ergibt sich als Näherungsformel für die Berechnung eines Konfidenzintervalls: $p-k\cdot\frac{\sqrt{h\cdot(1-h)}}{\sqrt{n}} \le h \le p+k\cdot\frac{\sqrt{h\cdot(1-h)}}{\sqrt{n}}$.

Der Term $d(k,h,n)=2\cdot k\cdot\frac{\sqrt{h\cdot(1-h)}}{\sqrt{n}}$ ist damit ein Näherungswert für die Länge des Konfidenzintervalls. Auf dem TI-Nspire gibt es eine vorinstallierte Applikation zum Ermitteln des Konfidenzintervalls, bei der diese Näherungsformel genutzt wird.

Menü → Statistik → Konfidenzintervalle → 1 – Prop z – Intervall





© Dr. W. Zappe