STOCHASTIK

Testen und Schätzen

Hartmut Meyer
https://mathemeyer.com

Inhalt

Binomialverteilung – Umgebungen um den Erwartungswert

Beispiel 1: *Ein LAPLACE-Würfel wird 90-mal geworfen. Es wird nach jedem Wurf notiert, ob eine „Eins“ geworfen wurde oder nicht.*

Die Zufallsgröße X, die für die Anzahl der Einsen in einer 90er-Serie steht, ist also binomialverteilt mit den Parametern $n = 90$ und $p = \frac{1}{6}$.

Für das Zufallsexperiment gibt es insgesamt 91 verschiedene Ergebnisse, denn es gilt $X \in \{0, 1, 2, 3, \dots, 90\}$.

Der Erwartungswert und damit auch das Maximum der Zufallsgröße X ist $\mu = n \cdot p = 15$. Die Wahrscheinlichkeit, dass in einer solchen 90er-Serie genau 15-mal eine Eins geworfen wird, ist allerdings nur $P(X = \mu) = b_{90;\frac{1}{6}}(15) \approx 0{,}1122 \approx 11{,}2\%$.

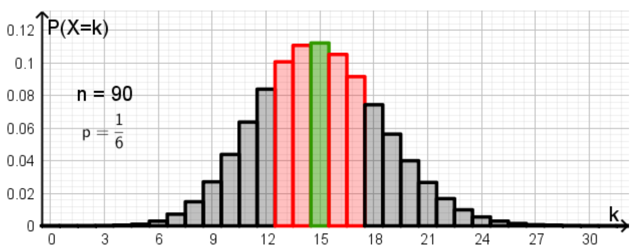
Die Ergebnisse in der Nachbarschaft von $\mu = 15$ haben allerdings eine fast ebenso große Wahrscheinlichkeit:

$P(X = 13) = b_{90;\frac{1}{6}}(13) \approx 0{,}1007 \approx 10{,}1\%$

$P(X = 14) = b_{90;\frac{1}{6}}(14) \approx 0{,}1107 \approx 11{,}1\%$

$P(X = 16) = b_{90;\frac{1}{6}}(16) \approx 0{,}1052 \approx 10{,}5\%$

$P(X = 17) = b_{90;\frac{1}{6}}(17) \approx 0{,}0916 \approx 9{,}2\%$



Fasst man allein nur diese fünf Ergebnisse zusammen, so lohnt es sich darauf zu wetten, dass die Anzahl der geworfenen Einsen bei einer solchen 90er-Serie in dem Bereich $13 \leq X \leq 17$ liegt.

Es gilt: $P(13 \leq X \leq 17) = F_{90;\frac{1}{6}}(17) - F_{90;\frac{1}{6}}(12) \approx 0{,}5203 \approx 52{,}0\%$

In der Umkehrung bedeutet dies, dass alle Ergebnisse mit weniger als 13 bzw. mehr als 17 Einsen zusammengenommen nur mit einer Wahrscheinlichkeit von unter 48% vorkommen.

Erweitert man die Umgebung von $\mu = 15$ auf das Intervall $[6; 24]$ so gilt sogar:

$P(6 \leq X \leq 24) = F_{90;\frac{1}{6}}(24) - F_{90;\frac{1}{6}}(5) \approx 0{,}9930 \approx 99{,}3\%$

Die Ergebnisse der Intervalle $[0; 5]$ und $[25;\ 90]$ kommen somit zusammengenommen nur mit einer Wahrscheinlichkeit von ca. 0,7% vor.

Bei einem großen Stichprobenumfang n konzentrieren sich die Ergebnisse der Bernoulli-Experimente auf eine relativ schmale Umgebung um den Erwartungswert μ.

σ – Umgebungen um den Erwartungswert

In der Statistik wird die Standardabweichung σ als Maß für die Streuung um den Mittel- bzw. Erwartungswert genutzt. Hier gilt: je größer der Wert von σ ist, desto heterogener sind die Merkmale einer Stichprobe.

Bei binomialverteilten Zufallsgrößen mit den Parametern n und p ist die Berechnung der Standardabweichung sehr einfach: $\sigma = \sqrt{n \cdot p \cdot (1-p)}$.

Hier kann der Wert σ genutzt werden, um die Wahrscheinlichkeiten von Umgebungen des Erwartungswertes μ zu berechnen.

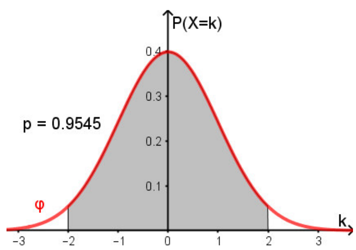
Bei großem Stichprobenumfang n gilt z.B. die Regel: *In dem Intervall $[\mu - \sigma;\ \mu + \sigma]$ liegen ca. 68,3% aller Ergebnisse einer Bernoulli-Kette.*

Diese Regel kann auch auf Vielfache von σ erweitert werden.

Für eine binomialverteilte Zufallsgröße X mit der Standardabweichung σ gelten folgende Regeln, sofern die Laplace-Bedingung – $\sigma > 3$ – erfüllt ist:

- $\boldsymbol{1\sigma - Regel}$: $\boldsymbol{P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0{,}683 \approx 68{,}3\%}$
- $\boldsymbol{2\sigma - Regel}$: $\boldsymbol{P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 0{,}955 \approx 95{,}5\%}$
- $\boldsymbol{3\sigma - Regel}$: $\boldsymbol{P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \approx 0{,}997 \approx 99{,}7\%}$

Die angegebenen Wahrscheinlichkeiten können mit Hilfe der Normalverteilung leicht berechnet werden.



So ist z.B. für $\mu = 0$ und $\sigma = 1$:

$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx \Phi_{0;1}(2) - \Phi_{0;1}(-2) \approx 0{,}9545$

Zur Veranschaulichung:

Beispiel 2: *Eine Münze wird 10000-mal geworfen. Es wird nach jedem Wurf notiert, ob die Seite „Zahl“ oben lag oder nicht.*

Die Zufallsgröße Z, die für die Anzahl der Würfe mit dem Ergebnis „Zahl“ steht, ist binomialverteilt mit den Parametern $n = 10000$ und $p = \frac{1}{2}$.
Es ist: $\mu = 5000$ und $\sigma = 50$. Dann gilt:

$$P(4950 \leq Z \leq 5050) = F_{10000;0,5}(5050) - F_{10000;0,5}(4949) \approx 0{,}6875$$
$$P(4900 \leq Z \leq 5100) = F_{10000;0,5}(5100) - F_{10000;0,5}(4899) \approx 0{,}9556$$
$$P(4850 \leq Z \leq 5150) = F_{10000;0,5}(5150) - F_{10000;0,5}(4849) \approx 0{,}9974$$

(Die leichten Abweichungen zu den σ – Regeln sind durch die Ungenauigkeit der Normalverteilung begründet.)

Mit Hilfe der Umkehrfunktion Φ^{-1} der Normalverteilung können auch für „glatte“ Werte der Umgebungswahrscheinlichkeiten die notwendigen Faktoren für die σ – Umgebungen bestimmt werden:

Für eine binomialverteilte Zufallsgröße X mit der Standardabweichung σ gelten folgende Regeln, sofern die Laplace-Bedingung – $\sigma > 3$ – erfüllt ist:

- $\boldsymbol{90\% - Regel}$: $\boldsymbol{P(\mu - 1{,}64 \cdot \sigma \leq X \leq \mu + 1{,}64 \cdot \sigma) \approx 90{,}0\%}$
- $\boldsymbol{95\% - Regel}$: $\boldsymbol{P(\mu - 1{,}96 \cdot \sigma \leq X \leq \mu + 1{,}96 \cdot \sigma) \approx 95{,}0\%}$
- $\boldsymbol{99\% - Regel}$: $\boldsymbol{P(\mu - 2{,}58 \cdot \sigma \leq X \leq \mu + 2{,}58 \cdot \sigma) \approx 99{,}0\%}$

Für das Beispiel 2 heißt das z.B.:

$[\mu - 1{,}96 \cdot \sigma;\ \mu + 1{,}96 \cdot \sigma]$

$\approx [5000 - 1{,}96 \cdot 50; 5000 + 1{,}96 \cdot 50]$

$\approx [4902; 5098]$

Interpretation:
Mit ca. 95%-iger Wahrscheinlichkeit wird mindesten 4902-mal und höchstens 5098-mal die Seite „Zahl“ bei den 10000 Münzwürfen oben liegen.

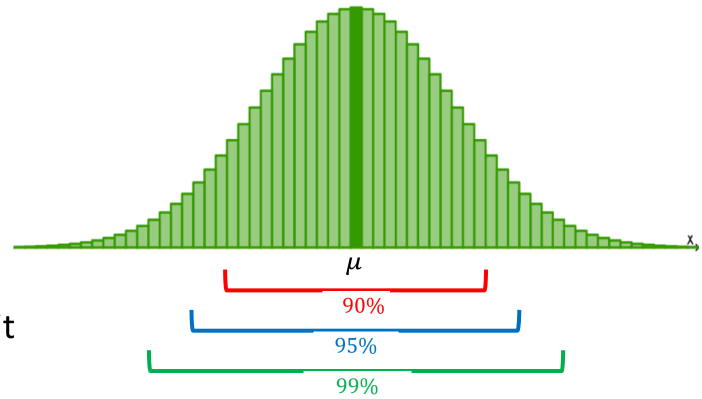
Schluss von der Gesamtheit auf die Stichprobe

Kennt man bei einem Bernoulli-Experiment die Trefferwahrscheinlichkeit (z.B. in Form eines Anteils an der Gesamtheit, oder auf Grund von LAPLACE-Annahmen) dann lässt sich zu einer vorgegebenen Wahrscheinlichkeit (z.B. 90%, 95% oder 99%) ein Intervall angeben, in dem das Ergebnis des Zufallsexperiments liegen wird.

Es stellt sich also die Frage:
„Welches Stichprobenergebnis (welche Anzahl von Treffern) kann erwartet werden?"

Die Antwort wird mit einer **Sicherheitswahrscheinlichkeit** von z.B. 90% bzw. 95% bzw. 99% gegeben.

In der Umkehrung bedeutet dies, dass die Prognose über das Stichprobenergebnis nur mit einer Wahrscheinlichkeit von 10% bzw. 5% bzw. 1% nicht zutrifft.



Bei Prognosen über voraussichtliche Ergebnisse von BERNOULLI-Experimenten sind also Intervalle um den Erwartungswert μ gesucht, in denen das Stichprobenergebnis mit einer hohen Wahrscheinlichkeit liegen wird.

Stichprobenergebnisse, die z.B. außerhalb des 95% - Intervalls liegen sind zwar möglich, sie treten aber sehr selten auf. Solche Ergebnisse nennt man in der Statistik ungewöhnlich, ihre Abweichung vom Erwartungswert **signifikant**.

Ergebnisse außerhalb des 99% - Intervalls treten besonders selten auf, ihre Abweichung vom Erwartungswert wird als **hochsignifikant** bezeichnet.

Stichprobenergebnisse, die innerhalb des Intervalls um den Erwartungswert liegen, nennt man **verträglich** mit der Trefferwahrscheinlichkeit p.

Beispiel 3: *Laut einem Zeitungsartikel wären 70% aller volljährigen Franzosen bereit, sich an dem Wiederaufbau der im April 2019 ausgebrannten Pariser Kirche Notre-Dame finanziell zu beteiligen.*

Wenn diese Angabe stimmt, müsste sich dies auch in einer zufällig zusammengestellten Stichprobe von 600 Franzosen widerspiegeln.
Mit welchem Stichprobenergebnis kann also gerechnet werden?

Die Befragung der 600 ausgewählten Personen kann als BERNOULLI-Experiment aufgefasst werden.

X *ist eine Zufallsgröße, die die Personen zählt, die angeben, sich eventuell finanziell an dem Wiederaufbau zu beteiligen.*
X *ist binomialverteilt mit* $n = 600$ *und* $p = 0{,}7$.

Gesucht ist das kleinste symmetrische Intervall $[k_1;\ k_2]$ *um den Erwartungswert* $\mu = 420$ *für das gilt:* $P(k_1 \leq X \leq k_2) \geq 95\%$ *bzw.* $P(k_1 \leq X \leq k_2) \geq 99\%$.

Mit einer Sicherheit von mindestens 95% wird man in der Stichprobe mindestens 398 und höchstens 442 spendenwillige Personen antreffen, denn:

$$P(398 \leq X \leq 442) = F_{600;0,7}(442) - F_{600;0,7}(397) \approx 95{,}5\%$$

Mit einer Sicherheit von mindestens 99% wird man in der Stichprobe mindestens 391 und höchstens 449 spendenwillige Personen antreffen, denn:

$$P(391 \leq X \leq 449) = F_{600;0,7}(449) - F_{600;0,7}(390) \approx 99{,}1\%$$

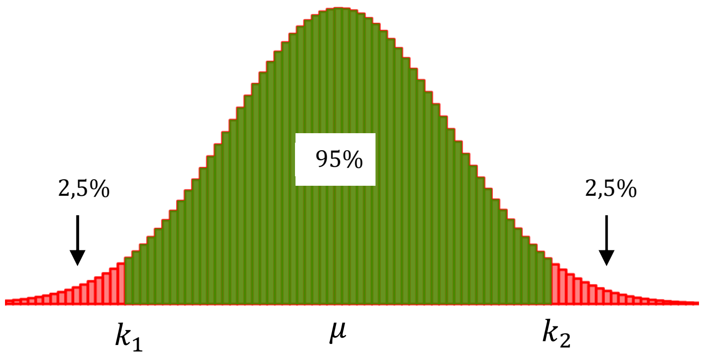
Würde das Ergebnis der Befragung <u>außerhalb</u> des Intervalls $[398; 442]$ *bzw.* $[391; 449]$ *liegen, so wäre die Abweichung vom Erwartungswert* $\mu = 420$ *signifikant bzw. sogar hochsignifikant.*

Solche Ergebnisse sind so ungewöhnlich, dass ein Zweifel an der Angabe des Zeitungsartikels berechtigt wäre.

In Beispiel 3 wurde gezeigt, dass die Summe der Wahrscheinlichkeiten für Stichprobenergebnisse aus dem Intervall $[398; 442]$ ca. 95,5% beträgt.
Aufgrund der Symmetrie des Intervalls um den Erwartungswert μ bedeutet dies in der Umkehrung, dass Ergebnisse, die im Intervall $[0; 397]$ liegen in der Summe mit einer Wahrscheinlichkeit von höchstens 2,5% vorkommen.
Gleiches gilt auch für Ergebnisse, die im Intervall $[443; 600]$ liegen.

Allgemein gilt für das Intervall $[k_1;\ k_2]$ mit der Wahrscheinlichkeit von mindestens 95%:

- $P(k_1 \leq X \leq k_2) \geq 0{,}95$
- $P(X \leq k_1 - 1) \leq 0{,}025$
- $P(X \geq k_2 + 1) \leq 0{,}025$



Für die Berechnung der beiden Intervallgrenzen k_1 und k_2 gibt es verschiedene Methoden. Mit steigender Genauigkeit sind dies:
(In Beispiel 3 gilt: $\mu = 420$ und $\sigma = \sqrt{126} \approx 11{,}2$. Es gilt: $\sigma > 3$, daher ist die Normalverteilung anwendbar.)

Methode 1: Berechnung mit Hilfe der σ-Regeln

Die 95%-Regel lautet: $P(\mu - 1{,}96 \cdot \sigma \leq X \leq \mu + 1{,}96 \cdot \sigma) \approx 95{,}0\%$.

$$[\mu - 1{,}96 \cdot \sigma;\ \mu + 1{,}96 \cdot \sigma]$$
$$\approx \left[420 - 1{,}96 \cdot \sqrt{126}; 420 + 1{,}96 \cdot \sqrt{126}\right] \approx [398; 442]$$

Es gilt also: $k_1 = 398$ und $k_2 = 442$.

Methode 2: Berechnung mit Hilfe der Funktion Φ^{-1}

Berechnung von k_1: $P(X \leq k_1 - 1) \leq 0{,}025$
„MOIVRE-LAPLACE“: $\Phi_{420;\,\sqrt{126}}(k_1 - 1 + 0{,}5) \leq 0{,}025$
Umkehrfunktion: $k_1 - 1 + 0{,}5 \leq {\Phi^{-1}}_{420;\,\sqrt{126}}(0{,}025)$
$k_1 - 0{,}5 \leq 397{,}999$
$k_1 \leq 398{,}499 \qquad \Rightarrow k_1 = 398$

Berechnung von k_2: $P(X \geq k_2 + 1) \leq 0{,}025 \;\Leftrightarrow\; P(X \leq k_2) \geq 0{,}975$

„Moivre-Laplace“: $\Phi_{420;\sqrt{126}}(k_2 + 0{,}5) \geq 0{,}975$

Umkehrfunktion: $k_2 + 0{,}5 \geq \Phi^{-1}{}_{420;\sqrt{126}}(0{,}975)$

$k_2 + 0{,}5 \geq 442{,}001$

$k_2 \geq 441{,}501 \qquad \Rightarrow k_2 = 442$

Methode 3: Rechnung mit Hilfe der Binomialverteilung – Probieren

Berechnung von k_1: $P(X \leq k_1 - 1) \leq 0{,}025$

Probieren mit dem Taschenrechner ergibt:
$P(X \leq 397) = F_{600;0,7}(397) \approx 0{,}0234 \leq 0{,}025$
$P(X \leq 398) = F_{600;0,7}(398) \approx 0{,}0287 > 0{,}025$

Also ist $k_1 - 1 = 397 \;\Rightarrow\; k_1 = 398$

$X = k$	$P(X = k)$
395	0,0153
396	0,0189
397	0,0234
398	0,0287
399	0,0349

Berechnung von k_2: $P(X \geq k_2 + 1) \leq 0{,}025 \;\Leftrightarrow\; P(X \leq k_2) \geq 0{,}975$

Probieren mit dem Taschenrechner ergibt:
$P(X \leq 441) = F_{600;0,7}(441) \approx 0{,}9734 < 0{,}975$
$P(X \leq 442) = F_{600;0,7}(442) \approx 0{,}9785 \geq 0{,}975$

Also ist $k_2 = 442$

$X = k$	$P(X = k)$
440	0,9672
441	0,9734
442	0,9785
443	0,9828
444	0,9863

Alle drei Methoden führen hier zu dem gleichen Ergebnis:
$k_1 = 398$ und $k_2 = 442$. Es ist:

- $P(k_1 \leq X \leq k_2) = F_{600;0,7}(442) - F_{600;0,7}(397) \approx 0{,}9551 \approx 95{,}5\%$
- $P(X \leq k_1 - 1) = F_{600;0,7}(397) \approx 0{,}0234 \approx 2{,}3\%$
- $P(X \geq k_2 + 1) = 1 - F_{600;0,7}(442) \approx 0{,}0215 \approx 2{,}2\%$

Die Methoden 1 und 2 basieren allerdings auf der Formel von Moivre-Laplace, die ja nur eine Näherung der Binomialverteilung darstellt. Zuverlässige Ergebnisse werden hier nur erzielt, wenn die Laplace-Bedingung $\sigma > 3$ deutlich erfüllt ist.

Hypothesentests

Bei einem Hypothesentest nimmt man an, die dem Zufallsexperiment zugrunde liegende Erfolgswahrscheinlichkeit p zu kennen oder man hat zumindest eine Vermutung über die Größe von p, d.h. man hat eine **Hypothese** über die Wahrscheinlichkeit p.

Ob diese Hypothese haltbar ist, kann mit Hilfe einer Stichprobe oder mehrmaligen Wiederholungen des Experiments **getestet** werden.
Wird z.B. eine Stichprobe erhoben, so kann das Ergebnis dieser Stichprobe daraufhin untersucht werden, ob es verträglich mit dieser Hypothese ist oder ob es signifikant von dem zu erwartenden Ergebnis abweicht.

Bei jedem Test gibt es also zwei mögliche Ergebnisse:
die angenommene Hypothese wird bestätigt oder verworfen.
Die ursprüngliche Annahme nennt man auch **Nullhypothese** $\boldsymbol{H_0}$ das Gegenteil dieser Nullhypothese nennt man die **Gegenhypothese** $\boldsymbol{H_1}$.

Hierfür muss zunächst das **Signifikanzniveau** festgelegt werden:

Möchte man, dass die **Sicherheitswahrscheinlichkeit**, also die Wahrscheinlichkeit mit der ein mögliches Stichprobenergebnis <u>innerhalb</u> des Intervalls um den Erwartungswert liegt, z.B. 95% bzw. 99% beträgt, so nennt man die Wahrscheinlichkeit, in diesen Fällen 5% bzw. 1% , mit der das Stichprobenergebnis <u>außerhalb</u> dieses Intervalls liegt, das **Signifikanzniveau** dieses Hypothesentests.
Die Festlegungen auf ein Signifikanzniveau von 1% und 5% sind willkürlich.
Für einzelne Tests sind auch andere Werte denkbar. In der Praxis nutzt man in der Statistik aber häufig einen dieser beiden Werte.

Das Intervall um den Erwartungswert, in dem die Ergebnisse liegen, die mit der Hypothese, also mit der angenommenen Erfolgswahrscheinlichkeit p verträglich sind, nennt man den **Annahmebereich** der Hypothese. Den Bereich, der außerhalb dieses Intervalls liegt, nennt man den **Verwerfungsbereich** der Hypothese.

Hypothesentests werden oftmals durchgeführt, wenn man:

- *vermutet*, dass eine aufgestellte Behauptung nicht stimmt
- *eine Beobachtung macht*, die nicht zu der bisher geltenden Lehrmeinung passt
- *hofft* oder *befürchtet*, dass eine Maßnahme zu einer Veränderung des bislang Geltenden führt

Wenn eine (neue) Vermutung, also eine Hypothese, aufgestellt wird, so ist dies die Gegenhypothese zur bis dahin geltenden Hypothese, der Nullhypothese.

Eine neue, bislang unbewiesene Hypothese kann nie direkt bestätigt werden, wohl aber können extreme (signifikant abweichende) Stichprobenergebnisse dazu führen, dass eine bis dahin geltende Hypothese/Tatsache verworfen werden kann und damit von der Gültigkeit des Gegenteils, also der „neuen“ Hypothese ausgegangen werden kann.

Der Begriff „Annahmebereich“ darf hierbei nicht missverstanden werden: Liegt ein Stichprobenergebnis im Annahmebereich einer Hypothese, so bedeutet dies nicht, dass damit die Hypothese bestätigt wird. Man kann in solch einem Fall nur sagen, dass es aufgrund des Stichprobenergebnisses keinen Anlass gibt, die Hypothese zu verwerfen.

Ein Hypothesentest ist also immer eine indirekte Methode.

Untersucht wird stets, ob die Ergebnisse des Tests zu der Nullhypothese H_0 „passen“ oder nicht.

Der zweiseitige Hypothesentest

Zur Erinnerung an Beispiel 3:
Laut einem Zeitungsartikel wären 70% aller volljährigen Franzosen bereit, sich an dem Wiederaufbau der im April 2019 ausgebrannten Pariser Kirche Notre-Dame finanziell zu beteiligen.

Angenommen, jemand hätte Zweifel an dieser Angabe. Dann besteht die Möglichkeit, die Richtigkeit der Zahl „70%“ mit Hilfe einer Stichprobe zu testen.

Zunächst einige Festlegungen für die Durchführung des Tests:
- Der Stichprobenumfang soll $n = 600$ sein.
- Das Signifikanzniveau soll 5% betragen.
- Die Nullhypothese H_0 lautet:
  „der Anteil der spendenwilligen Franzosen beträgt 70%“.
- Die Gegenhypothese H_1 lautet:
  „der Anteil der spendenwilligen Franzosen beträgt nicht 70%“.

Ist p also die Wahrscheinlichkeit für eine spendenwillige Person, so gilt kurz:

$$\boldsymbol{H_0: p = 0,7} \text{ **und** } \boldsymbol{H_1: p \neq 0,7}.$$

Bevor die Stichprobe ausgewertet wird, muss eine **Entscheidungsregel** erstellt werden. Hierfür kann man die Rechnungen aus Beispiel 3 nutzen:

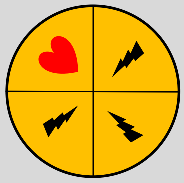
Für den Annahmebereich der Hypothese H_0 gilt: $\{398, \dots, 442\}$.
Somit ist der Verwerfungsbereich von H_0: $\{0, \dots, 397\} \cup \{443, \dots, 600\}$.

Die Entscheidungsregel lautet also:

- *Trifft man in der Stichprobe mindestens 398 aber höchstens 442 spendenwillige Franzosen an, so kann man nicht davon ausgehen, dass die Angabe des Zeitungsartikels falsch ist.*
- *Finden sich in der Stichprobe jedoch weniger als 398 oder mehr als 442 spendenwillige Franzosen, so kann man davon ausgehen, dass die Angabe des Zeitungsartikels nicht korrekt ist.*

Geben bei dieser Befragung also z.B. nur 390 oder sogar 450 Personen an, sich finanziell an dem Wiederaufbau beteiligen zu wollen, so wird dem „Zweifler“ Recht gegeben, die Zeitung kann nicht mehr länger von einem Anteil von 70% sprechen.

Beispiel 4: *Bei einem Gewinnspiel vor einem Supermarkt wird das nebenstehende Glücksrad verwendet.*



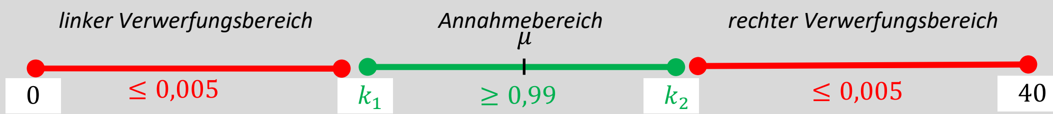
Ein Teilnehmer bezweifelt, dass das Glücksrad wirklich mit der zu erwartenden Wahrscheinlichkeit von $p = \frac{1}{4}$ *das Symbol ♥ anzeigt.*

Es wird vereinbart, das Glücksrad 40-mal zu drehen und dabei die ♥ zu zählen. Gesucht ist eine Entscheidungsregel, die die Annahme des Teilnehmers stützt bzw. verwirft. (Sicherheitswahrscheinlichkeit: 99%).

H_0: *„das Symbol ♥ hat die Wahrscheinlichkeit* $p = \frac{1}{4}$*"*
H_1: *„das Symbol ♥ hat die Wahrscheinlichkeit* $p \neq \frac{1}{4}$*"*

X *ist eine Zufallsgröße, die die Drehungen mit dem Symbol ♥ zählt.*
X *ist binomialverteilt mit* $n = 40$ *und* $p = 0{,}25 \quad \Rightarrow \quad \mu = 10;\ \sigma \approx 2{,}7$.

Gegen H_0 *würden sowohl sehr wenige, als auch sehr viele Drehungen mit dem Ergebnis ♥ sprechen:*



Es handelt sich um einen zweiseitigen Test mit dem Annahmereich $[k_1;\ k_2]$.
Da die Sicherheitswahrscheinlichkeit 99% betragen soll gilt:

(1) $P(X \leq k_1 - 1) \leq 0{,}005$
Probieren ergibt: $F_{40;0,25}(3) \approx 0{,}0047$ *und* $F_{40;0,25}(4) \approx 0{,}0160$
$\Rightarrow\ k_1 - 1 = 3 \ \Leftrightarrow\ \boldsymbol{k_1 = 4}$

(2) $P(X \leq k_2) \geq 0{,}995$
Probieren ergibt: $F_{40;0,25}(16) \approx 0{,}9884$ *und* $F_{40;0,25}(17) \approx 0{,}9953$
$\Rightarrow\ \boldsymbol{k_2 = 17}$

Entscheidungsregel:
Dreht man bei diesen 40 Durchführungen mindestens 4-mal aber höchstens 17-mal das Symbol ♥, so sind die Zweifel unberechtigt, erscheint aber weniger als 4-mal oder mehr als 17-mal das Symbol ♥, so sind die Zweifel des Teilnehmers berechtigt.

Fehler beim Testen von Hypothesen

Wie bei jedem Hypothesentest verbergen sich in der Entscheidungsregel jedoch zwei mögliche Fehlerquellen:

(1) Zum einen könnte man in dieser 40er-Serie nur zufällig weniger als 4 bzw. mehr als 17-mal das Symbol „Herz" gedreht haben, obwohl das Glücksrad völlig in Ordnung ist, also mit der Wahrscheinlichkeit $p = \frac{1}{4}$ das „Herz" zeigt. Die Wahrscheinlichkeit hierfür ist zwar sehr klein, aber nicht unmöglich: $P(X < 4) + P(X > 17) \approx 0{,}0093 \approx 0{,}9\%$.

Man hätte in diesem Fall fälschlicherweise die Hypothese H_0 verworfen, obwohl sie in Wahrheit richtig ist.

Diesen Fehler nennt man **Fehler erster Art** $(\boldsymbol{\alpha})$.
Seine Größe entspricht dem vorgegebenen Signifikanzniveau.

(Die Differenz zwischen dem vorgegebenen Wert 1% und dem berechneten Wert ca. 0,9% erklärt sich daraus, dass nur ganze Zahlen für X betrachtet werden.)

(2) Zum anderen könnte man in dieser 40er-Serie – ebenfalls zufällig – zwischen 4 und 17 Drehungen mit dem „Herz" erzielen, obwohl für die Wahrscheinlichkeit von „Herz" tatsächlich nicht $p = \frac{1}{4}$, also $p \neq \frac{1}{4}$, gilt.

Man hätte in diesem Fall fälschlicherweise die Hypothese H_0 bestätigt, obwohl sie in Wahrheit falsch ist.

Diesen Fehler nennt man **Fehler zweiter Art** $(\boldsymbol{\beta})$**.**

Für die Berechnung der Größe des Fehlers zweiter Art müsste man z.B. die tatsächliche Wahrscheinlichkeit für das Symbol „Herz" kennen. Dies ist in den seltensten Fällen der Fall.

	H_0 ist wahr	H_0 ist falsch
Das Ergebnis liegt im Annahmebereich von H_0	✓	Fehler 2.Art
Das Ergebnis liegt im Verwerfungsbereich von H_0	Fehler 1.Art	✓

Der einseitige Hypothesentest

In den Beispielen 3 und 4 wurde geprüft, ob dem betrachteten Zufallsexperiment tatsächlich eine behauptete bzw. erwartete Erfolgswahrscheinlichkeit zugrunde liegt oder nicht. Signifikante Abweichungen konnten dabei sowohl nach unten als auch nach oben auftreten.

In der Realität beinhalten „neue“ Hypothesen häufig gleichzeitig eine Tendenz:

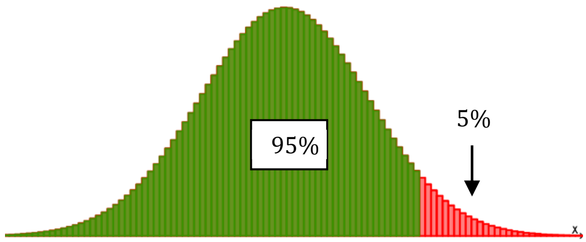
(1) Der „Zweifler“ an dem Wahrheitsgehalt des Zeitungsartikels in Beispiel 3 könnte z.B. vermuten, dass die Zahl 70% zu klein ist, der Anteil der spendenwilligen Personen deutlich **größer** ist.

(2) Der Teilnehmer an dem Glückspiel vor dem Supermarkt wird eventuell das Gefühl haben, dass das Symbol ❤ deutlich **seltener** erscheint als es (theoretisch) vorkommen sollte.

In diesen beiden Fällen könnten signifikante Abweichungen vom Erwartungswert also nur zu einer Seite hin auftreten.

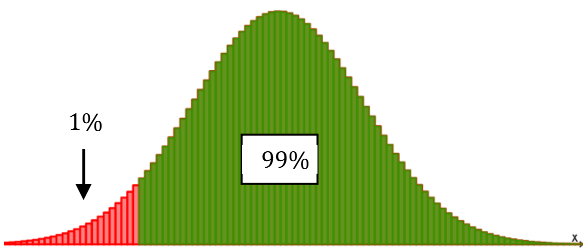
zu (1): H_0: „der Anteil der spendenwilligen Personen beträgt **höchstens** 70%“
H_1: „der Anteil der spendenwilligen Personen beträgt **mehr als** 70%“

Gegen die Hypothese H_0 würden nur sehr viele spendenwilligen Personen in der Stichprobe sprechen. Also würden nur **signifikante Abweichungen nach oben** zum Verwerfen der Hypothese H_0 führen und damit die Vermutung des „Zweiflers“ stützen.



zu (2): H_0: „das Symbol ❤ hat eine Wahrscheinlichkeit **mindestens** 25%“
H_1: „das Symbol ❤ hat eine Wahrscheinlichkeit von **weniger als** 25%“

Gegen die Hypothese H_0 würden nur sehr wenige Drehungen mit dem Ergebnis ❤ sprechen. Also würden nur **signifikante Abweichungen nach unten** zum Verwerfen der Hypothese H_0 führen und damit die Vermutung des Teilnehmers stützen.



Einen einseitigen Hypothesentest, bei dem nur signifikante Abweichungen nach <u>oben</u> zu einem Verwerfen der Hypothese H_0 führen, nennt man einen **rechtsseitigen Test**. Führen nur signifikante Abweichungen nach <u>unten</u> zum Verwerfen der Hypothese H_0, so ist dies ein **linksseitiger Test**.

Beispiel 5: Eine Maschine stellt Linsen für den Bau eines Lasergerätes her. Die Dicke der Linsen darf dabei um maximal einen Mikrometer von dem Idealmaß (Norm) abweichen.

Um den Ausschussanteil der Linsen möglichst gering zu halten, wird die Maschine regelmäßig alle drei Tage neu justiert. Der Firma gelingt es hiermit den Anteil der Linsen, die der Norm nicht entsprechen bei maximal 2% zu halten.

Aus Kostengründen sollen die Intervalle für die Neujustierung auf 7 Tage erhöht werden. Ein Mitarbeiter vermutet, dass der Ausschussanteil dadurch deutlich erhöht wird. Er möchte dies überprüfen und entnimmt der Tagesproduktion eine Stichprobe von 150 Linsen und misst deren Dicke.

Gesucht ist eine Entscheidungsregel für diesen Test. Das Signifikanzniveau soll hierbei 5% betragen.

Der Hypothese H_0: „der Ausschussanteil beträgt maximal 2%“ steht also die Hypothese H_1: „der Ausschussanteil beträgt nun mehr als 2%“ entgegen.
Ist p der Anteil der nicht der Norm entsprechenden Linsen, so gilt also:

$$H_0:\ p \leq 0{,}02 \ \text{bzw.}\ H_1:\ p > 0{,}02.$$

Die zufällige Entnahme einer Linse aus der Tagesproduktion entspricht – streng genommen – nicht einem BERNOULLI-Experiment. Da man aber von einer großen Gesamtheit ausgehen kann, kann die Binomialverteilung, als gute Näherung, genutzt werden.

Die Zufallsgröße X zähle die „Ausschusslinsen“ in der Stichprobe. X ist (annähernd) binomialverteilt mit $n = 150$ und $p = 0{,}02$ $(\Rightarrow\ \mu = 3;\ \ \sigma \approx 1{,}7)$

Da nur ein relativ großer Anteil von unbrauchbaren Linsen gegen H_0 spräche, handelt es sich hier also um einen **rechtsseitigen Hypothesentest** mit dem Annahmebereich $[0; k]$ und dem Verwerfungsbereich $[k+1;\ 150]$.

Das Signifikanzniveau ist auf $\alpha = 5\%$ festgelegt.
Also soll für den Test gelten: $P(X \leq k) \geq 1 - \alpha = 0{,}95$

Probieren ergibt:
$F_{150;\,0{,}02}(5) \approx 0{,}9181$ *und* $F_{150;\,0{,}02}(6) \approx 0{,}9680 \;\; \Rightarrow \;\; \boldsymbol{k = 6}$

Also ist der Annahmebereich $[0; 6]$ *und der Verwerfungsbereich* $[7; 150]$.

Entscheidungsregel:
Befinden sich in der Stichprobe maximal 6 fehlerhafte Linsen, so kann man davon ausgehen, dass die Verlängerung der Wartungsintervalle keinen Einfluss auf den Ausschussanteil hat.
Findet der Mitarbeiter allerdings 7 oder mehr fehlerhafte Linsen, so hat sich die Wahrscheinlichkeit für die Produktion fehlerhafter Linsen erhöht.

Der Fehler 1.Art in dem Beispiel 5 lautet:

Der Mitarbeiter findet in der Stichprobe – zufällig – mehr als 6 fehlerhafte Linsen und verwirft daraufhin die Hypothese H_0, macht also die Umstellung für einen Qualitätsverlust verantwortlich. In Wahrheit hat sich aber der Anteil der Ausschussteile durch die Umstellung nicht geändert.

Wie durch das vorgegebene Signifikanzniveau gefordert, liegt die Wahrscheinlichkeit für den Fehler 1.Art unter 5%.
(Genau: $\alpha = P(X > 7) = 1 - F_{150;\,0{,}02}(6) \approx 0{,}0320 \approx 3{,}2\%$)

Der Fehler 2.Art in dem Beispiel 5 lautet:

Der Mitarbeiter bestätigt durch seinen Test die Hypothese H_0, da er – zufällig – nicht mehr als 6 fehlerhafte Linsen gefunden hat, d.h. er geht davon aus, dass die Umstellung keinen Qualitätsverlust verursacht. In Wahrheit sorgt die Verlängerung der Wartungsintervalle aber für mehr Ausschuss.

Die Größe dieses Fehlers kann nicht angegeben werden, da die tatsächliche Wahrscheinlichkeit p_2 für eine fehlerhafte Linse nicht bekannt ist.

Es gilt: $\beta = F_{150;\,\boldsymbol{p_2}}(6)$, also die Wahrscheinlichkeit dafür, mit der neuen Erfolgswahrscheinlichkeit p_2 im Annahmebereich von H_0 zu „landen“.

Der Alternativtest

Ein Spezialfall des einseitigen Tests ist der Alternativtest. Hier weiß man im Vorhinein, dass zu dem zugrundeliegenden BERNOULLI-Experiment entweder die Erfolgswahrscheinlichkeit p_1 oder die Erfolgswahrscheinlichkeit p_2 gehört.

Um eine Entscheidung über die Erfolgswahrscheinlichkeit zu treffen, geht man zunächst von einer der beiden Wahrscheinlichkeiten, z.B. $H_0: p = p_1$ aus.
Liegt das Ergebnis einer Stichprobe im Verwerfungsbereich von H_0, so geht man im Folgenden davon aus, dass die Erfolgswahrscheinlichkeit p_2 richtig ist.
Liegt das Ergebnis hingegen im Annahmebereich von H_0, so geht man weiterhin von der Richtigkeit von p_1 aus.

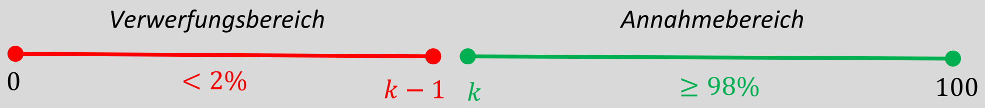
Beispiel 6: *Saatgut für Erbsen wird in zwei Güteklassen mit unterschiedlicher Keimgarantie angeboten: Von den Erbsen der Güteklasse A keimen 90%, von denen der Güteklasse B nur 75%.*

Ein Großhändler erhält eine große Lieferung Erbsen-Saatgut, von dem er allerdings nicht weiß, ob es sich um Saatgut der Güteklasse A oder B handelt.
Er will dies mit Hilfe einer Stichprobe von 100 zufällig entnommenen Saatkörnern testen.
Gesucht ist eine Entscheidungsregel bei der das Risiko, das Saatgut irrtümlich als Güteklasse B einzustufen, kleiner als 2% ist.

Lösung:
Die Formulierung gibt vor, dass der Großhändler zunächst einmal von der Güteklasse A ausgeht. Falls er sich damit irrt, handelt es sich um die Güteklasse B.

Wenn p *die Keimwahrscheinlichkeit ist, lauten die Hypothesen also:*
H_0*: „es handelt sich um Saatgut der Güteklasse A“* $p = p_A = 0{,}9$
H_1*: „es handelt sich um Saatgut der Güteklasse B“* $p = p_B = 0{,}75$

H_0 *soll verworfen werden, wenn in der Stichprobe relativ wenig keimende Erbsen gefunden werden. Es ist also ein* ***linksseitiger Hypothesentest*** *mit dem Annahmebereich* $[k; 100]$ *und dem Verwerfungsbereich* $[0;\ k-1]$.



Die Zufallsgröße X_A steht für die keimenden Erbsen in der Stichprobe.
X ist binomialverteilt mit $n = 100$ und $p = 0{,}9$ $(\Rightarrow\ \mu = 90;\ \ \sigma = 3)$

Das Signifikanzniveau/die Irrtumswahrscheinlichkeit ist auf $\alpha = 2\%$ festgelegt. Also soll für den Test gelten: $P(X \leq k-1) \leq 0{,}02$

Probieren ergibt: $F_{100;\,0,9}(82) \approx 0{,}0100$ *und* $F_{100;\,0,9}(83) \approx 0{,}0206$

$$\Rightarrow k - 1 = 82 \ \Leftrightarrow\ \boldsymbol{k = 83}$$

Also ist der Verwerfungsbereich $[0; 82]$, der Annahmebereich $[83; 100]$.

Entscheidungsregel:
Sollten von den 100 ausgesäten Erbsen mindestens 83 keimen, so kann der Großhändler weiterhin davon ausgehen, dass es sich um Saatgut der Güteklasse A handelt.
Keimen jedoch nur 82 oder weniger Erbsen, kann der Großhändler davon ausgehen, dass es sich um Saatgut der Güteklasse B handelt.

Der Fehler 1.Art lautet hier:

Der Großhändler verwirft die Hypothese „es handelt sich um Saatgut der Güteklasse A“, da – zufällig – weniger als 83 Erbsen keimen, obwohl es sich in Wahrheit um Saatgut der Güteklasse A handelt.

Der Fehler 2.Art lautet hier:

*Der Großhändler verwirft die Hypothese „es handelt sich um Saatgut der Güteklasse A“ **nicht**, da – zufällig – mindestens 83 Erbsen keimen, obwohl es sich in Wahrheit um Saatgut der Güteklasse B handelt.*

Die Größe des Fehlers 2.Art kann hier konkret berechnet werden, da die Alternativ-Wahrscheinlichkeit ja bekannt ist:

$$\beta = P(X_B > 82) = 1 - F_{100;\mathbf{0,75}}(82) \approx 0{,}0376 \approx 3{,}8\%.$$

Schluss von der Stichprobe auf die Gesamtheit

Die Vorgehensweise bei einem Hypothesentest ist also immer gleich:

(1) Man nimmt an, die Hypothese H_0, die gerade getestet werden soll, gilt.
(2) Man berechnet unter dieser Annahme das 95% - bzw. 99% - Intervall (oder ein beliebig anderes) einer Zufallsgröße.
(3) Man verwirft die Hypothese H_0, wenn das in einer Stichprobe ermittelte Ergebnis außerhalb dieses Intervalls liegt.

In der Praxis der Statistik ist die Erfolgswahrscheinlichkeit p des zugrundeliegenden BERNOULLI-Experiments aber häufig nicht bekannt.
Dann muss das Verfahren umgedreht werden:

Man ermittelt zunächst das Stichprobenergebnis und bestimmt dann alle Erfolgswahrscheinlichkeiten, d.h. alle Hypothesen, die zu diesem Ergebnis passen, also nicht durch das Testergebnis verworfen werden können.
In diesem Fall schließt man also von der Stichprobe auf die Gesamtheit.

Schätzen von Erfolgswahrscheinlichkeiten

Betrachtet man das Beispiel 3, mit den spendenwilligen Franzosen, so stellt sich die Frage: Wie ist die Zeitung auf den Wert „70%“ gekommen?
Angenommen die Zeitung hat redlich recherchiert, dann haben die Redakteure sicherlich nicht alle erwachsenen Franzosen befragt, sondern ihrerseits eine (repräsentative) Stichprobe erhoben.
Wenn sie es sich einfach gemacht haben, ist z.B. folgendes Szenario denkbar:

Die Zeitung hat eine Umfrage unter 1000 erwachsenen Franzosen in Auftrag gegeben, von denen sich 700 als potenzielle Spender zu erkennen gaben.
Ein, der Prozentrechnung kundiger, Redakteur hat daraufhin den Anteil „70%“ ermittelt.

Nimmt man, wie in Beispiel 3 geschehen, diesen Wert als Erfolgswahrscheinlichkeit des BERNOULLI-Experiments „Befragung von Personen in einer Stichprobe“, so ist das Stichprobenergebnis $X = 700$ sicherlich mit der Hypothese : $H_0\colon p = 0{,}7$ verträglich.
Allerdings werden auch andere Hypothesen, also andere Erfolgswahrscheinlichkeiten, nicht durch das Ergebnis $X = 700$ verworfen.
Anders formuliert: Zu dem Ergebnis $X = 700$ „passen“ unendlich viele Hypothesen : $H_0\colon p = p_0$.

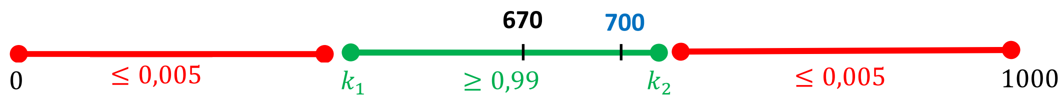
Die Frage ist nun:

Mit welchen Wahrscheinlichkeiten p_0 ist das Ergebnis $X = 700$ verträglich?

Die Befragung entspricht einem zweiseitigen Hypothesentest mit den Hypothesen: $H_0: p = p_0$ und $H_1: p \neq p_0$.

Testen wir z.B. auf dem 1%-Signifikanzniveau, so gilt: $\alpha = 0{,}01 \Rightarrow \frac{\alpha}{2} = 0{,}005$.

Für z.B. $p_0 = 0{,}67$ gilt dann mit $n = 1000$, $\mu = 670$ und $\sigma = \sqrt{201}$.



Berechnung des Annahmebereiches $[k_1;\ k_2]$ von H_0:
(Hier könnte auch wieder mit dem Taschenrechner probiert werden.)

$P(X_{0,67} \leq k_1 - 1) \leq 0{,}005$

$\Phi_{670;\,\sqrt{201}}(k_1 - 1 + 0{,}5) \leq 0{,}005$

$k_1 - 1 + 0{,}5 \leq \Phi^{-1}{}_{670;\,\sqrt{201}}(0{,}005)$

$k_1 - 0{,}5 \leq 633{,}48$

$k_1 \leq 633{,}98$

$\Rightarrow k_1 = 633$

$P(X_{0,67} \leq k_2) \geq 0{,}995$

$\Phi_{670;\,\sqrt{201}}(k_2 + 0{,}5) \geq 0{,}995$

$k_2 + 0{,}5 \geq \Phi^{-1}{}_{670;\,\sqrt{201}}(0{,}995)$

$k_2 + 0{,}5 \geq 706{,}52$

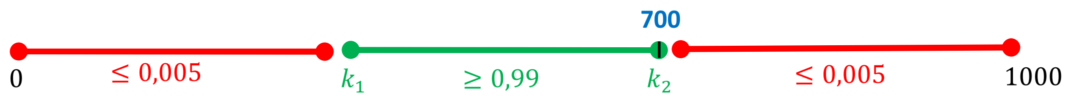
$k_2 \geq 706{,}02$

$\Rightarrow k_2 = 707$

Da $X = 700$ im Annahmebereich $[633; 707]$ liegt, ist das Ergebnis auch mit der Wahrscheinlichkeit $p_0 = 0{,}67$ verträglich.

Für die Berechnung der kleinsten und der größten Wahrscheinlichkeit p_{min} bzw. p_{max} mit der das Stichprobenergebnis $X = 700$ verträglich ist, nimmt man an, dass das Ergebnis $X = 700$ jeweils die obere bzw. die untere Grenze des Annahmebereiches ist, also gerade noch nicht zur Verwerfung der Hypothese führt. Also: $k_2 = 700$ bzw. $k_1 = 700$

Berechnung der kleinsten Wahrscheinlichkeit p_{min} mit der das Ergebnis $X = 700$ (gerade noch) verträglich ist: $k_2 = 700$:



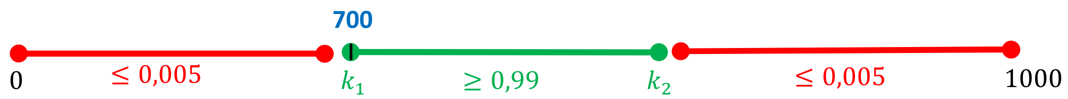
$P\left(X_{p_{min}} \leq 700\right) \approx 0{,}995$

p	$P(X_p \leq 700)$
0,67	0,9805
0,66	0,9968
0,663	0,9943
0,662	**0,9953**

Probieren mit dem Taschenrechner ergibt:
(Ergebnis hier auf 3 Stellen gerundet.)

Für die Wahrscheinlichkeit $p_{min} \approx 0{,}662$ liegt das Ergebnis $X = 700$ noch im Annahmebereich der Hypothese H_0.

Berechnung der größten Wahrscheinlichkeit p_{max} mit der das Ergebnis $X = 700$ (gerade noch) verträglich ist: $k_1 = 700$:



$P(X_{p_{max}} \leq 700 - 1) \approx 0{,}005$

p	$P(X_p \leq 699)$
0,73	0,0156
0,74	0,0020
0,735	0,0059
0,736	**0,0048**

Probieren mit dem Taschenrechner ergibt:
(Ergebnis ebenfalls auf 3 Stellen gerundet.)

Für die Wahrscheinlichkeit $p_{max} \approx 0{,}736$ liegt das Ergebnis $X = 700$ noch im Annahmebereich der Hypothese H_0.

Das Ergebnis $X = 700$ ist somit verträglich mit allen Erfolgswahrscheinlichkeiten $66{,}2\% \leq p \leq 73{,}6\%$.

Das sogenannte **Konfidenzintervall** (Vertrauensintervall) für das Stichprobenergebnis $X = 700$ lautet somit: $[p_{min};\ p_{max}] = [0{,}662; 0{,}736]$.

Die Zeitung hätte also, wenn sie statistisch korrekt hätte sein wollen, schreiben müssen:

Einer Umfrage zufolge liegt der Anteil der spendenwilligen erwachsenen Franzosen mit einer Sicherheitswahrscheinlichkeit von 99% zwischen ca. 66,2% und ca. 73,6%. ☺