Testen von Hypothesen

Einführung:

Wer Entscheidungen zu treffen hat, weiß oft erst im nachhinein ob seine Entscheidung richtig war. Die Unsicherheit eine Entscheidung zu treffen, beinhaltet immer eine gewisse Fehlerwahrscheinlichkeit. Der Hypothesentest gibt uns eine Richtlinie für die Wahl einer Alternativentscheidung. Wir treffen unsere Entscheidung auf der Grundlage dessen, was wir für richtig erachten. Das nennen wir die **Nullhypothese**. Eine Alternativentscheidung nennen wir **alternative Hypothese**.

Das Testen von Hypothesen ist immer ein Vorgang, den man in mehrere Schritte unterteilen kann:

- Formulierung der Null- und der Alternativhypothese
- Spezifikation des Signifikanzniveaus
- Berechnung der Testgröße
- Definition des Ablehnungsbereichs
- Selektion der passenden Hypothese

Beispiel 1:
Die Befragung aller Studenten einer Fachhochschule ergab im letzten Jahr, dass nur 10% der befragten Studenten mit dem Mensaessen unzufrieden waren. Das Mensa-Organisationsteam steht nun vor der Entscheidung, ob Maßnahmen zur Verbesserung der Qualität des Essens ergriffen werden müssen. Um hier richtig zu entscheiden, werden in einer Umfrage 100 Studenten befragt.

a) Sind mindestens 10 Studenten mit dem Essen unzufrieden, so soll die Qualität des Essens verbessert werden. In der ersten Umfrage erklärten 12 Studenten, mit dem Mensaessen unzufrieden zu sein.

b) Das Mensateam ist sich der Zufälligkeit von Stichprobenergebnissen bewusst und lässt in einer 2. Umfrage weitere 100 Studenten befragen. Dabei gibt sich das Team mit einer Sicherheit von 95% mit dem Befragungsergebnis zufrieden. Es erklären nur 8 Studenten, mit dem Essen nicht zufrieden zu sein.

Wie wird das Team in beiden Fällen entscheiden?

Nullhypothese (H_0): Hypothese (Annahme), dass mindestens 10% der Studenten mit dem Essen unzufrieden sind und damit die Qualität des Essens verbessert werden muss.

a) Es wird die H_0- Hypothese getestet. Sie wird angenommen, bzw. beibehalten, wenn die Zahl der Studenten im **Annahmebereich** und sie wird abgelehnt, wenn die Zahl der Studenten im **Ablehnungsgsbereich** liegt.

$H_0 : p \geq 0{,}1$ Annahmebereich $A : \{10, 11, ...; 100\}$
Ablehnungsbereich $\overline{A} : \{0, 1, ...; 9\}$

Beim 1. Test erklärten 12 Studenten, mit dem Essen unzufrieden zu sein.
Die H_0- Hypothese würde damit angenommen, aber dies könnte wegen der Zufälligkeit der Stichprobe falsch sein, wenn der tatsächliche Anteil der

unzufriedenen in der **Grundgesamtheit** (Menge aller Studenten der Fachhochschule) kleiner als 10% ist. Man begeht also bei der Annahme der H_0-Hypothese mit einer gewissen Wahrscheinlichkeit einen Fehler.

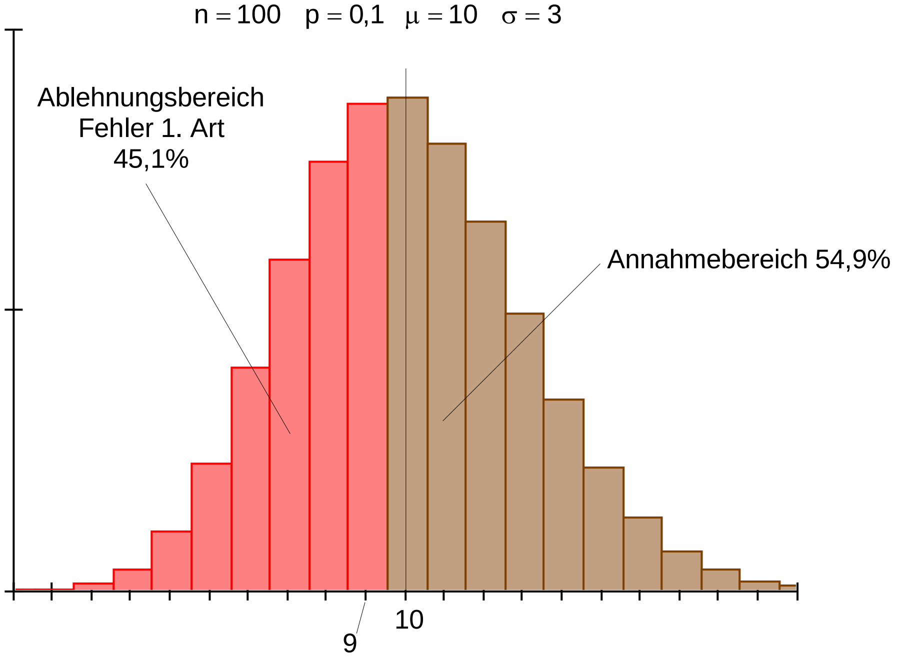
Dieser Fehler, auch als **Irrtumswahrscheinlichkeit** des Tests bezeichnet, berechnet sich aus der Ablehnungswahrscheinlichkeit.

$P(X \le 9) = 0,451$ (siehe Tabelle 1).

Das heißt, höchstens 9 Studenten sind mit dem Essen unzufrieden. Man muss also sagen, „Unter der Annahme, dass tatsächlich 10% aller Studenten unzufrieden sind, kommt es bei der angegebenen Befragung in ca. 45,1% zu einer Fehlentscheidung“.

Eine Verteilungsfunktion soll das verdeutlichen:

$n = 100 \quad p = 0,1 \quad \mu = 10 \quad \sigma = 3$



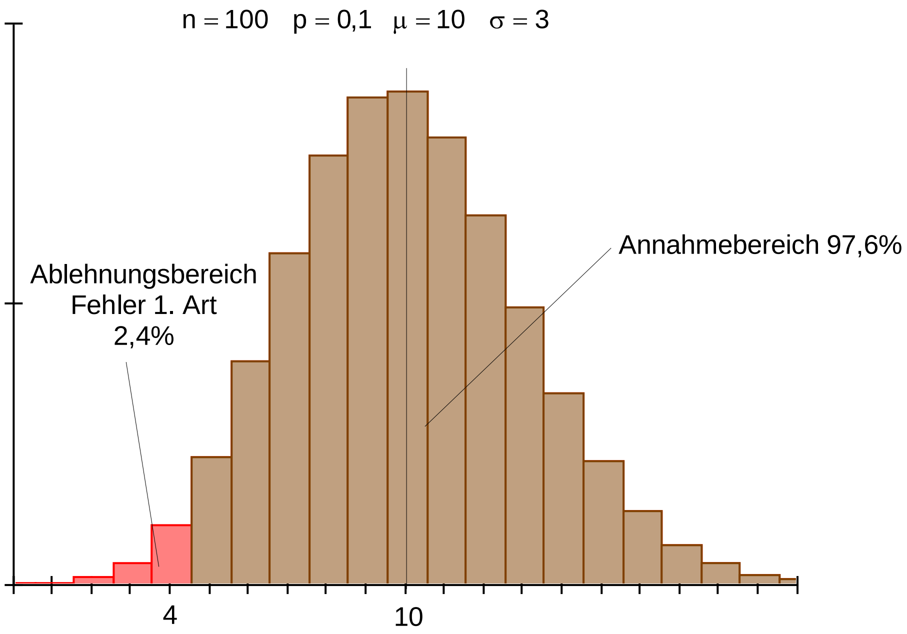
b) Beim 2. Test wird ein Fehler von 5% zugestanden (Sicherheit von 95%). Dadurch ändern sich Annahme- und Ablehnungsbereich.

$P(X \le k) \le 0,05 \Rightarrow k = 4$ (Der Tabelle 1 entnommen)

Ablehnungsbereich $\overline{A} : \{0,1,2,3,4\}$

Annahmebereich $A : \{5,6,7,...,100\}$

Während der Annahmebereich größer wird, wird der Ablehnungsbereich kleiner.
Da das 2. Testergebnis mit 8 unzufriedenen Studenten in den neuen Annahmebereich fällt, entschließt sich das Mensateam, die Qualität des Essens zu verbessern.
Eine Verteilungsfunktion soll das verdeutlichen:

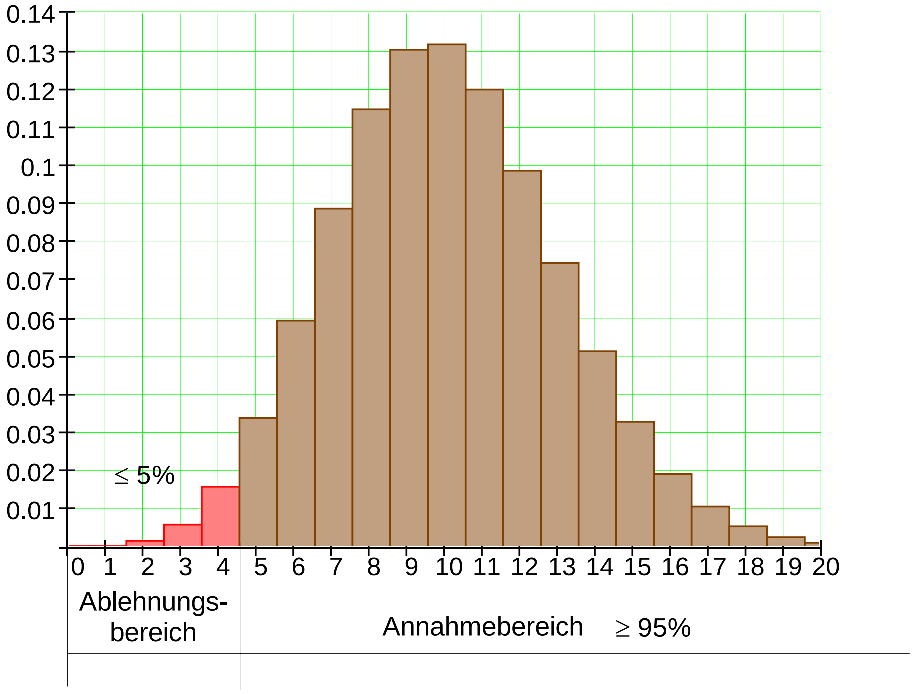


Die Wahrscheinlichkeit beim Testen einen gewissen Fehler zuzulassen, heißt Irrtumswahrscheinlichkeit. Sie wird in der Regel vor Durchführung des Zufallsexperimentes festgelegt. Dabei sind 1% und 5% übliche Werte. Sie ist die größte Wahrscheinlichkeit für eine irrtümliche Ablehnung der H_0- Hypothese. Statt Irrtumswahrscheinlichkeit sagt man auch **Signifikanzniveau**.

Tabelle 1: Kumulierte Binomialverteilung für n = 100 und p = 0,1

k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$
0	0,000	5	0,058	10	0,583	15	0,960	20	0,999
1	0,000	6	0,117	11	0,703	16	0,979	21	1,000
2	0,002	7	0,206	12	0,802	17	0,990	22	1,000
3	0,008	8	0,321	13	0,876	18	0,995	23	1,000
4	0,024	9	0,451	14	0,927	19	0,998	24	1,000

Ein Test, bei dem der Ablehnungsbereich unterhalb, also links vom Erwartungswert liegt, heißt **Linksseitiger Hypothesentest.** Vielfach wird dieses Verfahren dann benutzt , wenn p > a zu testen ist.
Da die Binomialverteilung eine diskrete Verteilung darstellt, gelingt es oft nicht, den Ablehnungsbereich so zu bestimmen, dass die Irrtumswahrscheinlichkeit genau dem vorgegebenen Wert entspricht. Sie sollte aber nicht oberhalb des geforderten Wertes liegen.



Linksseitiger Hypothesentest

Fehler beim Testen von Hypothesen

Die Entscheidung, die aufgrund eines Versuchsergebnisses (Test, Umfrage, ...) getroffen wird kann falsch sein. Die zu testende Hypothese H_0 (mindestens 10% aller Studenten sind mit dem Essen unzufrieden) kann wahr oder falsch sein.
Man unterscheidet zwei Arten von Fehlern:
Fehler 1. Art: Ist der Fehler, eine wahre Hypothese abzulehnen (zu verwerfen).
Fehler 2. Art: Ist der Fehler, eine falsche Hypothese nicht zu verwerfen (sie anzunehmen).

Beispiel 2: Alternativtest
Ein Babybasar verkauft gebrauchte Kinderschuhe. Etwa 60% der Schuhe befinden sich in einem einwandfreien Zustand. Der Rest weist kleine Schäden auf. Ein neuer Lieferant behauptet, er könne gebrauchte Kinderschuhe liefern, von denen sich etwa 80% in einem einwandfreien Zustand befinden. Der Ladeninhaber möchte keine falsche Kaufentscheidung treffen und will die Behauptung des Lieferanten überprüfen.

Dazu testet er 20 Paar Kinderschuhe aus dem Sortiment des Anbieters.

Angenommen, die Behauptung des Lieferanten ist richtig, d.h. die Wahrscheinlichkeit für einwandfreie Schuhe ist p = 0,8.
Dann erwartet man im Versuch mit $n = 20$ Paar Schuhen, ca. $n \cdot p = 20 \cdot 0{,}8 = 16$ einwandfreie Paare.
Wenn mehr als 16 Paar Schuhe einwandfrei sind, dann spricht das eher für die Behauptung des Liefereranten.
Zufällig kann es auch zu weniger als 16 Paar einwandfreien Schuhen kommen, obwohl p = 0,8 ist.

Mit welcher Wahrscheinlichkeit ist dies der Fall?

$P(X \leq 15) = 0{,}370$ (Siehe Tabelle 2).

Mit einer Wahrscheinlichkeit von 37% kann es vorkommen, dass bei dem Test weniger als 16 Paar Schuhe einwandfrei sind, obwohl etwa 80% der Schuhe einwandfrei sind.

Angenommen, die Schuhe des Lieferanten sind auch nur zu 60% einwandfrei, wie der vorhandene Bestand, dann kann man bei dem Test 12 einwandfreie Paar Schuhe erwarten.
Wenn weniger als 12 einwandfreie Paar Schuhe gefunden werden, spricht das eher gegen die Behauptung des Lieferanten (p = 0,8).
Zufällig kann es aber auch zu mehr als 12 einwandfreien Paaren kommen, obwohl p = 0,6 ist.
Mit welcher Wahrscheinlichkeit ist dies der Fall?

$P(X > 12) = P(X \leq 20) - P(X \leq 12) = 1 - 0{,}584 = 0{,}416$ (siehe Tabelle 3)

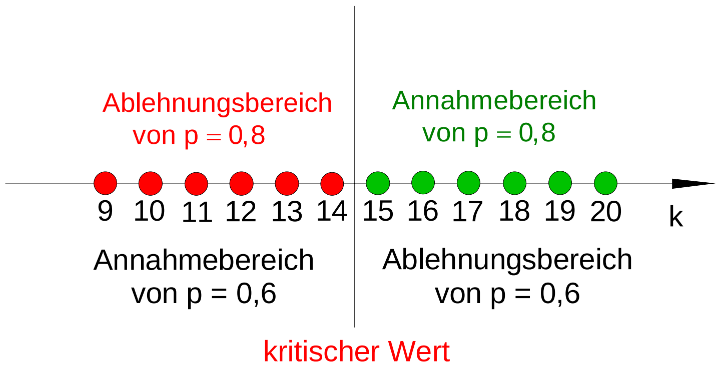
Mit einer Wahrscheinlichkeit von 41,6% kann es vorkommen, dass bei dem Test mehr als 12 Paar Schuhe einwandfrei sind, obwohl nur 60% der Schuhe einwandfrei sind.

Bevor der Test durchgeführt wird, muss man sich entscheiden, bei welcher Anzahl von einwandfreien Schuhen man p = 0,8 oder p = 0,6 für richtig halten will.

Man kann z.B. vorher festlegen, dass man der Behauptung des Lieferanten glaubt, (p = 0,8), falls mindestens 15 Paar Schuhe einwandfrei sind. Andernfalls man davon ausgeht, dass die Schuhe des Lieferanten auch nicht besser sind als der gegenwärtige Bestand (p = 0,6).

Es wird folgende **Entscheidungsregel** aufgestellt:
Falls mindestens 15 Paar einwandfrei sind, wird p = 0,8 als richtig angesehen, sonst soll p = 0,6 gelten. Diese Entscheidung ist willkürlich.

$\text{Annahmebereich}: A = \{15 \ldots 20\} \qquad \text{Ablehnungsbereich}: \overline{A} = \{0 \ldots 14\}$



Fehlermöglichkeiten dieser Entscheidung:

1. p = 0,8 ist richtig, das bedeutet, der neue Lieferant kann wirklich Schuhe höherer Qualität liefern. Zufällig kann es vorkommen, das weniger als 15 Paar Schuhe einwandfrei sind. Dann würde man dem Lieferanten nicht glauben.
   Die Wahrscheinlichkeit einen solchen Fehler zu begehen beträgt
   $P_{80}\left(X \le 14\right) = 0{,}196$ (siehe Tabelle 2).
   Das bedeutet, wenn man einen solchen Zufallsversuch mit 20 Paar Schuhen sehr oft durchführen würde, könnte man in 19,6% der Fälle ein Ergebnis erwarten, das gegen die tatsächliche Qualität der Schuhe spricht.

   Fehler 1. Art: In 19,6% aller Fälle würde die wahre Hypothese, (die Schuhe des neuen Lieferanten sind besser) verworfen werden.

2. p = 0,6 ist richtig, das bedeutet, der neue Lieferant kann auch keine besseren Schuhe liefern, als die, die man bereits hat. Zufällig kann es aber vorkommen, dass trotzdem 15 oder mehr Paar Schuhe einwandfrei sind. Man würde in diesem Fall fälschlicherweise die Schuhe des neuen Lieferanten für besser halten.
   Die Wahrscheinlichkeit einen solchen Fehler zu begehen beträgt

   $P_{60}\left(X \ge 15\right) = P\left(X \le 20\right) - P\left(X \le 14\right) = 1 - 0{,}874 = 0{,}126$ (siehe Tabelle 3).

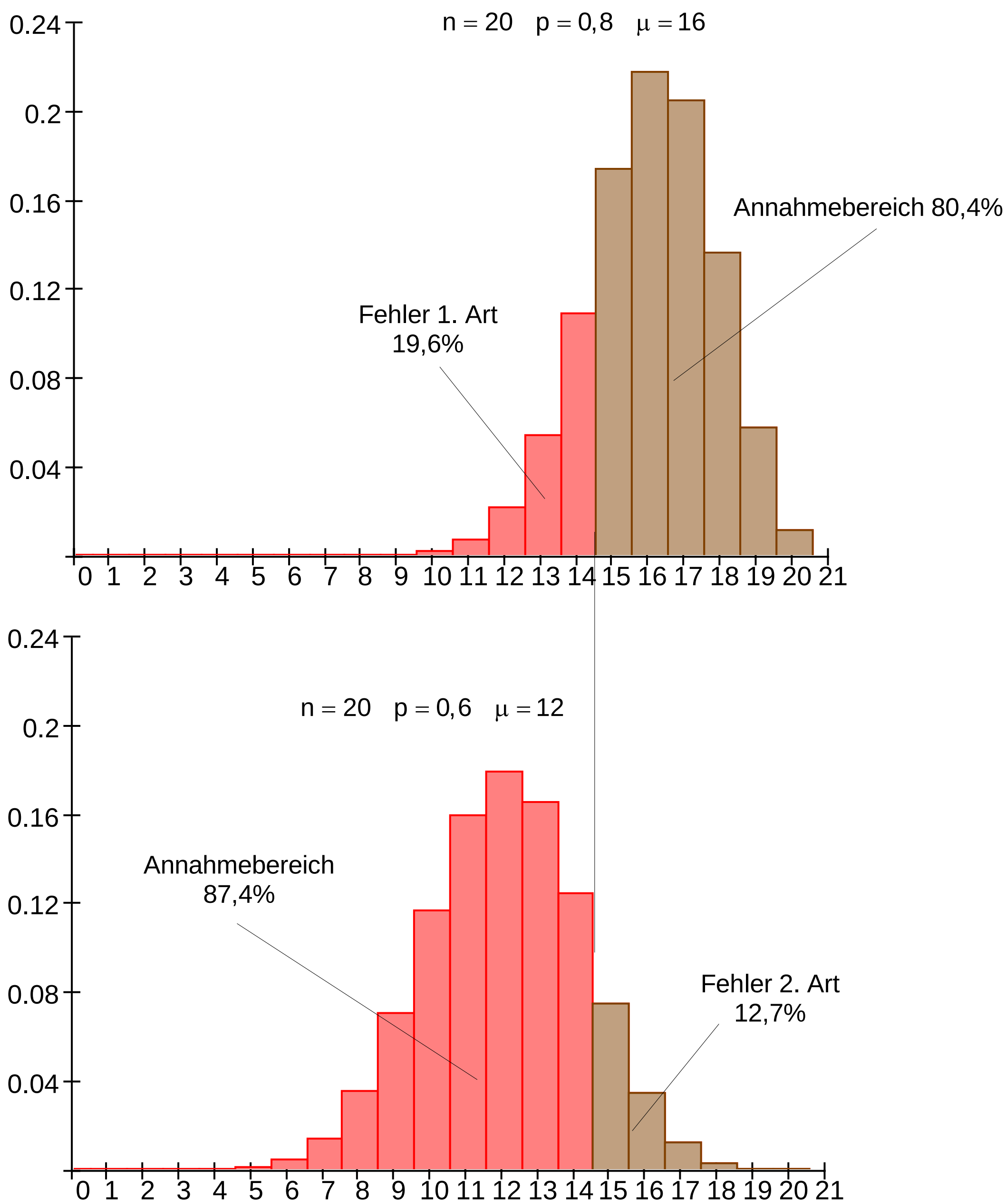
   Das bedeutet, wenn man einen solchen Zufallsversuch mit 20 Paar Schuhen sehr oft durchführen würde, könnte man in 12,6% der Fälle ein Ergebnis erwarten, dass die Qualität der Schuhe höher angesehen wird, als sie tatsächlich ist.

   Fehler 2. Art: In 12,6% aller Fälle würde die falsche Hypothese, (die Schuhe des neuen Lieferanten sind besser) nicht verworfen werden.

Zusammenfassung der Fehlerarten:

	Versuchsergebnis liegt im Annahmebereich von $p = 0{,}8$ $A = \left\{15...20\right\}$	Versuchsergebnis liegt im Ablehnungsbereich von $p = 0{,}8$ $\overline{A} = \left\{0...14\right\}$
Hypothese $p = 0{,}8$ ist wahr	Entscheidung ist richtig, Hypothese $p = 0{,}8$ wird angenommen $P_{80}\left(X \ge 15\right) = 0{,}804 \mathrel{\hat{=}} 80{,}4\%$	Entscheidung ist falsch, Hypothese $p = 0{,}8$ wird abgelehnt $P_{80}\left(X \le 14\right) = 0{,}196 \mathrel{\hat{=}} 19{,}6\%$ (Fehler 1. Art)
Hypothese $p = 0{,}6$ ist wahr	Entscheidung ist falsch, Hypothese $p = 0{,}6$ wird abgelehnt $P_{60}\left(X \ge 15\right) = 0{,}126 \mathrel{\hat{=}} 12{,}6\%$ (Fehler 2. Art)	Entscheidung ist richtig, Hypothese $p = 0{,}6$ wird angenommen $P_{60}\left(X \le 14\right) = 0{,}874 \mathrel{\hat{=}} 87{,}4\%$

Die Wahrscheinlichkeit einen Fehler 1. Art zu machen wird mit α bezeichnet. ($\alpha = 0{,}196$)
Die Wahrscheinlichkeit einen Fehler 2. Art zu machen wird mit β bezeichnet. ($\beta = 0{,}126$)



Irrtumswahrscheinlichkeit wird vorgegeben.

Wird eine Irrtumswahrscheinlichkeit vorgegeben, dann ergibt sich daraus der Annahme und der Ablehnungsbereich.

Vorgabe der Irrtumswahrscheinlichkeit $\alpha \leq 0{,}1$

$P_{0,8}\left(X \leq k\right) \leq 0{,}1 \Rightarrow k = 13$ aus Tabelle 2 abgelesen, denn $P_{0,8}\left(X \leq 13\right) = 0{,}087$

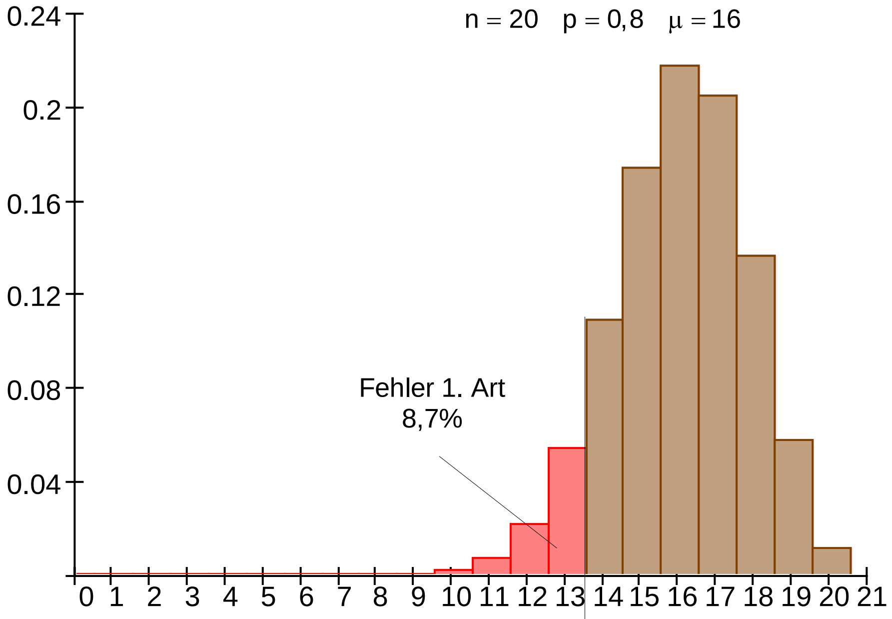
$\Rightarrow \alpha = 0{,}087 < 0{,}1$ ist die Wahrscheinlichkeit für einen Fehler 1. Art.

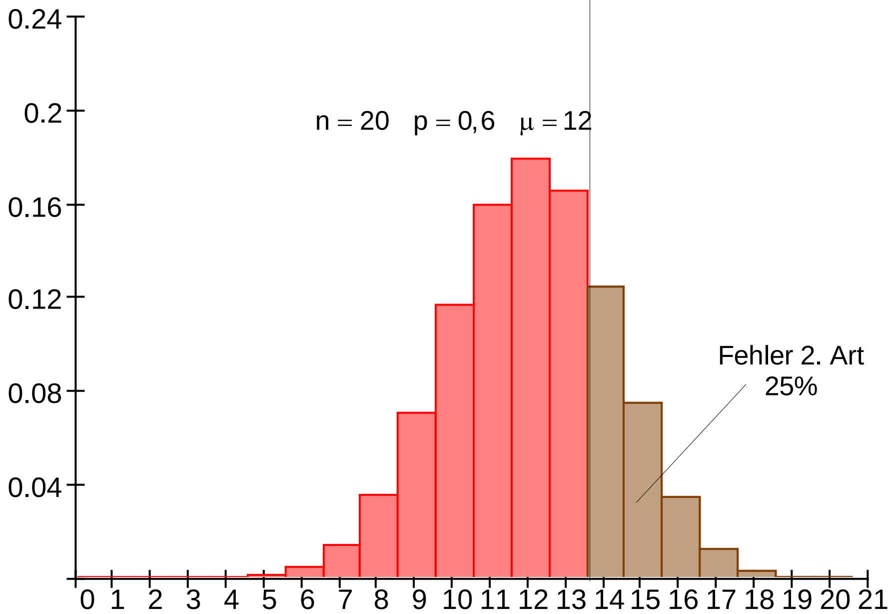
Annahmebereich: $A = \{14, 15, ..., 20\}$ Ablehnungsgsbereich: $\overline{A} = \{0, 1, ..., 13\}$

$P_{0,6}\left(X \geq 14\right) = P_{0,6}\left(X \leq 20\right) - P_{0,6}\left(X \leq 13\right) = 1 - 0{,}750 = 0{,}25$ (Tabelle 3)

$\Rightarrow \beta = 0{,}25$ ist die Wahrscheinlichkeit für einen Fehler 2. Art.

Dadurch, dass die Wahrscheinlichkeit für einen Fehler 1. Art auf mehr als die Hälfte verringert wurde, hat sich der Fehler 2. Art mehr als verdoppelt.





Falls die Hypothese p = 0,8 wahr ist, ist die Wahrscheinlichkeit dafür, dass sie aufgrund eines Testergebnisses fälschlicherweise abgelehnt wird 8,7%.
Denn in 8,7% aller Fälle liegt das Testergebnis im Ablehnungsbereich von p = 0,8.

Falls die Hypothese p = 0,6 wahr ist, ist die Wahrscheinlichkeit dafür, dass sie aufgrund eines Testergebnisses fälschlicherweise abgelehnt wird 25%.
Denn in 25% aller Fälle liegt das Testergebnis im Annahmebereich von p = 0,8.

Tabelle 2: Kumulierte Binomialverteilung für n = 20, p = 0,8

k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$
0	0,000	5	0,000	10	0,003	15	0,370	20	1,000
1	0,000	6	0,000	11	0,010	16	0,589		
2	0,000	7	0,000	12	0,032	17	0,794		
3	0,000	8	0,000	13	0,087	18	0,931		
4	0,000	9	0,001	14	0,196	19	0,988		

Tabelle 3: Kumulierte Binomialverteilung für n = 20, p = 0,6

k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$	k	$P(X \le k)$
0	0,000	5	0,002	10	0,245	15	0,949	20	1,000
1	0,000	6	0,006	11	0,404	16	0,984		
2	0,000	7	0,021	12	0,584	17	0,996		
3	0,000	8	0,057	13	0,750	18	0,999		
4	0,000	9	0,128	14	0,874	19	1,000		

Wann ist was wie zu testen?

Aufstellen der Nullhypothese:

Die Nullhypothese hängt von der jeweiligen Interessenlage ab.

Interessengruppe I behauptet $p \ge a \Rightarrow H_0 : p \ge a$
Interessengruppe II zweifelt $p \ge a$ an $\Rightarrow H_0 : p < a$

Verschiedene Interessengruppen stellen also unterschiedliche Nullhypothesen auf. Das hat zur Folge, dass Annahme- und Ablehnungsbereich für die Interessengruppen unterschiedlich sind. Auch die Art des durchzuführenden Tests hängt von der jeweilig aufgestellten Nullhypothese ab.

Linksseitiger Test falls	$p \ge a$ oder $p > a$ zu testen ist
Rechtsseitiger Test falls	$p \le a$ oder $p < a$ zu testen ist
Beidseitiger Test falls	$p = a$ oder zu testen ist

Grundsätzlich kann man davon ausgehen, dass unterschiedliche Interessengruppen gegensätzliche (konträre) Nullhypothesen aufstellen.