

Grundlagen zum Hypothesentest

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 Nullhypothese H_0 und der Alternativhypothese H_1
- Festlegung des Signifikanzniveaus
- Bestimmung des Annahme- und Ablehnungsbereichs der Nullhypothese
- Ziehung der Stichprobe
- Treffen der Testentscheidung und Interpretation:
Liegt das Ergebnis der Stichprobe innerhalb des Annahmebereichs, wird H_0 angenommen, anderenfalls abgelehnt

Für das richtige Aufstellen der Hypothesen gibt es ein paar Regeln:

- Was ich zeigen oder beweisen will, gehört in die Alternativhypothese
- Das Gleichheitszeichen gehört immer in die Nullhypothese
- Beim aufstellen der Nullhypothese geht man davon aus, "Alles bleibt beim alten, nichts hat sich geändert"

Die Annahme der Nullhypothese führt immer zur Ablehnung der Alternativhypothese, ist aber kein Beweis dafür, dass die Nullhypothese stimmt.

Die Ablehnung der Nullhypothese führt zur Annahme der Alternativhypothese.

Beispiel 1:

Die Befragung aller Studenten einer Fachhochschule ergab im letzten Jahr, dass 10% der befragten Studenten mit dem Mensaessen unzufrieden waren. Es wird vermutet, dass die Unzufriedenheit im laufenden Semester sogar noch zugenommen hat. Das Mensa- Organisationsteam steht nun vor der Entscheidung, ob Maßnahmen zur Verbesserung der Qualität des Essens ergriffen werden müssen. Um eine Entscheidung treffen zu können, werden in einer Umfrage 100 Studenten befragt.

- a) Sind mehr als 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 wieder 100 Studenten befragen. Dabei gibt sich das Team mit einer Sicherheit von 95% mit dem Befragungsergebnis zufrieden. Es erklären 13 Studenten, mit dem Essen nicht zufrieden zu sein.

Wie wird das Team in beiden Fällen entscheiden?

Die Entscheidung soll über einen Hypothesentest gefunden werden.

Zu zeigen ist $p > 0,1$

Das heißt: Mehr als 10% aller Studenten sind mit dem Essen unzufrieden.

Damit liegen die Null- und die Alternativhypothese fest.

$H_0 : p \leq 0,1$ und $H_1 : p > 0,1$

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 **Ablehnungsbereich** liegt.

$$H_0 : p \leq 0,1 \quad \begin{array}{l} \text{Annahmebereich} \quad A : \{0,1,\dots,10\} \\ \text{Ablehnungsbereich} \quad \bar{A} : \{11,12,\dots,100\} \end{array}$$

Beim 1. Test erklärten 12 Studenten, mit dem Essen unzufrieden zu sein.

Die H_0 - Hypothese würde damit abgelehnt, aber dies könnte wegen der Zufälligkeit der Stichprobe falsch sein, wenn der tatsächliche Anteil der unzufriedenen in der **Grundgesamtheit** (Menge aller Studenten der Fachhochschule) tatsächlich 10% ist. Man begeht also bei der Ablehnung der H_0 - Hypothese mit einer gewissen Wahrscheinlichkeit einen Fehler.

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

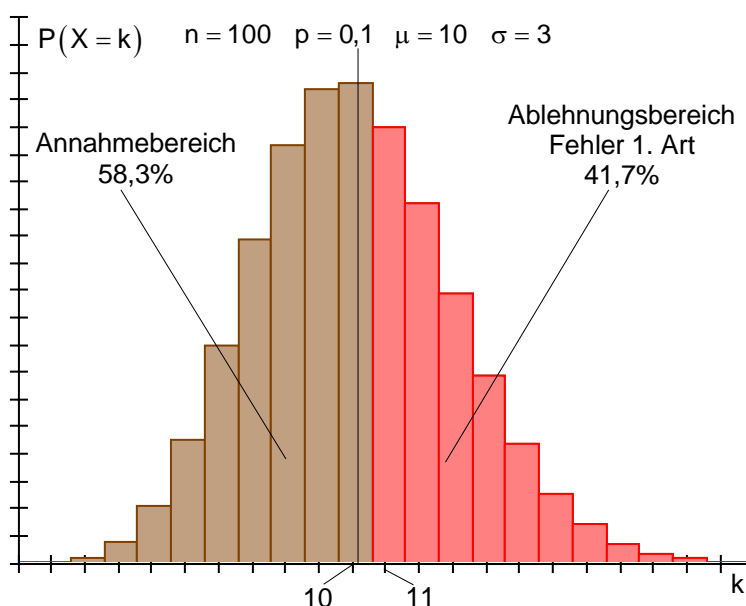
$$P(X \geq 11) = 1 - P(X \leq 10) \approx 1 - 0,583 = 0,417 \quad (\text{siehe Tabelle 1}).$$

Das heißt, mindestens 11 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 mit einer Wahrscheinlichkeit von ca. 41,7% zu einem solchen Ergebnis und damit zu fälschlichen Ablehnung der Nullhypothese“.

Eine Verteilungsfunktion soll das verdeutlichen:



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

$$P(X \geq k) \leq 0,05 \Leftrightarrow 1 - P(X \leq k - 1) \leq 0,05 \quad | -1$$

$$\Leftrightarrow -P(X \leq k - 1) \leq 0,05 - 1 \quad | \cdot (-1)$$

$$\Leftrightarrow P(X \leq k - 1) \geq 1 - 0,05 = 0,95$$

$$\Leftrightarrow P(X \leq 15) \approx 0,96 > 0,95$$

$$\Leftrightarrow k - 1 = 15 \Leftrightarrow k = 16 \quad (\text{Der Tabelle 1 entnommen})$$

$$\text{Annahmehereich} \quad A : \{0, 1, \dots, 15\}$$

$$\text{Ablehnungsbereich} \quad \bar{A} : \{16, 17, \dots, 100\}$$

Während der Annahmehereich größer wird, wird der Ablehnungsbereich kleiner.

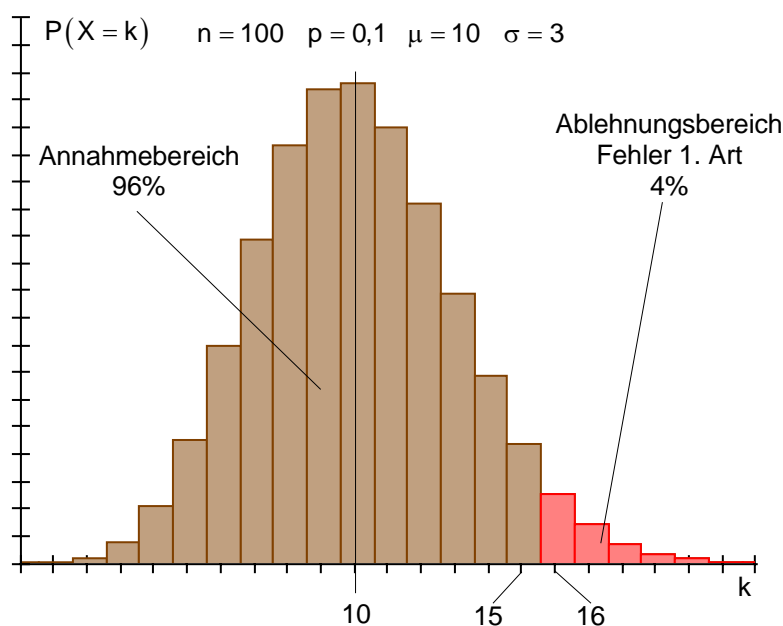
Da das 2. Testergebnis mit 13 unzufriedenen Studenten in den neuen

Annahmehereich fällt, wird die Nullhypothese angenommen.

Das Mensateam sieht keine Veranlassung, das Essen zu verbessern.

Erst wenn mehr als 15 Studenten mit dem Essen nicht zufrieden wären, würde die Nullhypothese abgelehnt, die Alternativhypothese angenommen und das Essen verbessert werden.

Eine Verteilungsfunktion soll das verdeutlichen:

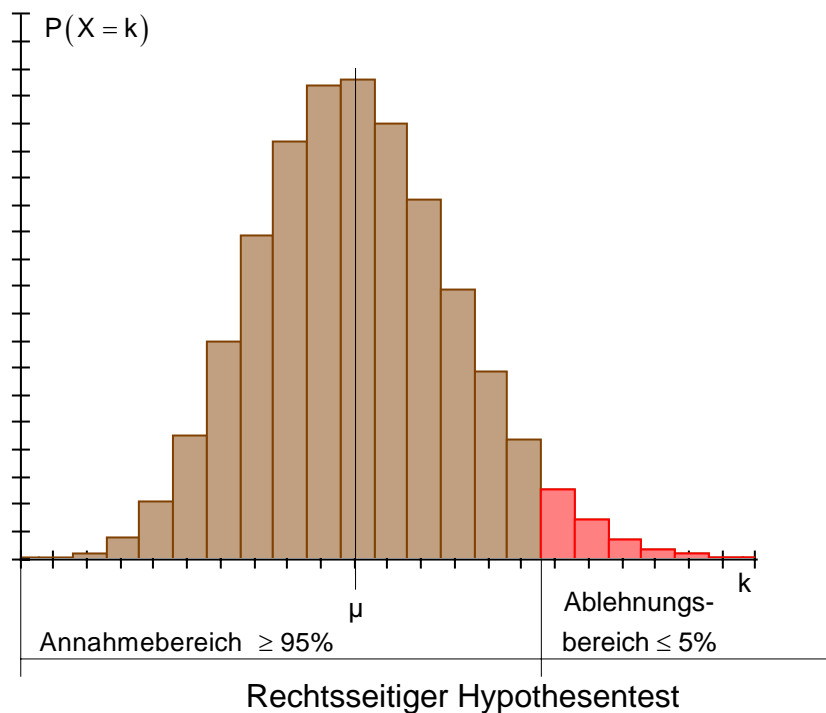


Die Wahrscheinlichkeit beim Testen einen gewissen Fehler zuzulassen, heißt Irrtumswahrscheinlichkeit. Sie wird in der Regel vor Durchführung des Zufallsexperimentes festgelegt. Dabei sind 1% und 5% übliche Werte. Sie ist die größte Wahrscheinlichkeit für eine irrtümliche Ablehnung der H_0 -Hypothese.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

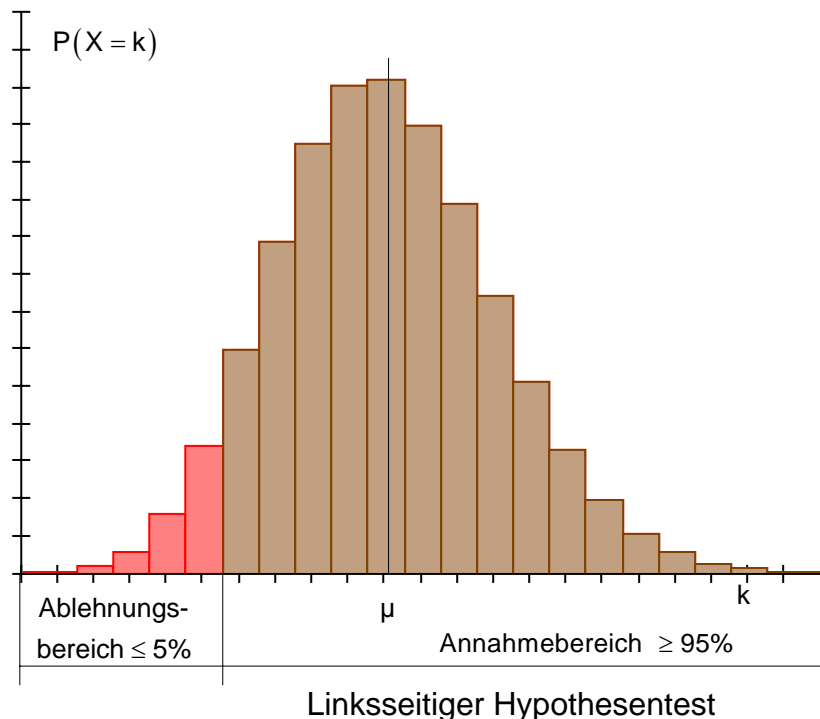
Tabelle 1:Kumulierte Binomialverteilung für $n = 100$ und $p = 0,1$

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

Ein Test, bei dem der Ablehnungsbereich oberhalb, also rechts vom Erwartungswert liegt, heißt **Rechtsseitiger Hypothesentest**. Vielfach wird dieses Verfahren dann benutzt, wenn die Alternativhypothese $H_1: p > a$ lautet, und $H_0: p \leq a$ 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.



Ein Test, bei dem der Ablehnungsbereich unterhalb, also links vom Erwartungswert liegt, heißt **Linksseitiger Hypothesentest**. Vielfach wird dieses Verfahren dann benutzt, wenn die Alternativhypothese $H_1: p < a$ lautet, und $H_0: p \geq a$ zu testen ist.



Fehler beim Testen von Hypothesen

Die Entscheidung, die aufgrund eines Versuchsergebnisses (Test, Umfrage, ...) getroffen wird kann falsch sein. Die zu testende Hypothese H_0 (höchstens 10% aller Studenten sind mit dem Essen unzufrieden) kann wahr oder falsch sein.

Man unterscheidet zwei Arten von Fehlern:

Fehler 1. Art: Die Nullhypothese wird verworfen, obwohl sie richtig ist.

Fehler 2. Art: Die Nullhypothese wird angenommen, obwohl sie falsch ist.

Der Fehler 2. Art lässt sich nur berechnen, wenn man für die Alternativhypothese eine andere Wahrscheinlichkeit, als für H_0 annimmt.

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.

Fall I:

Angenommen, die Behauptung des Lieferanten ist richtig, d.h. die Wahrscheinlichkeit für einwandfreie Schuhe ist $p = 0,8$.

Der Ladeninhaber bezweifelt die Aussage des Lieferanten, er geht von $p < 0,8$ aus.

Er stellt folgende Hypothesen auf:

Nullhypothese $H_0: p \geq 0,8$ und die Alternativhypothese $H_1: p < 0,8$

Im Versuch mit $n = 20$ Paar Schuhen erwartet man

$$\mu = n \cdot p = 20 \cdot 0,8 = 16$$

einwandfreie Paare.

Wenn mindestens 16 Paar Schuhe einwandfrei sind, dann spricht das für die Behauptung des Lieferanten, dann soll H_0 angenommen werden.

Zufällig kann es auch zu weniger als 16 Paar einwandfreien Schuhen kommen, obwohl $p = 0,8$ ist.

Die Nullhypothese soll abgelehnt werden, wenn weniger als 16 Paar Schuhe einwandfrei sind .

Mit welcher Wahrscheinlichkeit ist dies der Fall?

$$P(X \leq 15) = 0,370 \quad (\text{Siehe Tabelle 3}).$$

Mit einer Wahrscheinlichkeit von 37% kann es vorkommen, dass bei dem Test weniger als 16 Paar Schuhe einwandfrei sind, obwohl 80% der Schuhe einwandfrei sind. Man würde also mit einer Wahrscheinlichkeit von 37% irrtümlicher Weise die Nullhypothese ablehnen.

Fall II:

Angenommen, die Schuhe des Lieferanten sind auch nur zu 60% einwandfrei.

Mit dieser Annahme stellt der Ladenbesitzer folgende Hypothesen auf:

Nullhypothese $H_0: p \leq 0,6$ und die Alternativhypothese $H_1: p > 0,6$

Im Versuch mit $n = 20$ Paar Schuhen erwartet man

$$\mu = n \cdot p = 20 \cdot 0,6 = 12$$

einwandfreie Paare.

Wenn mehr als 12 einwandfreie Paar Schuhe gefunden werden, spricht das eher gegen die Vermutung des Ladenbesitzers ($p = 0,6$).

Zufällig kann es aber auch zu mehr als 12 einwandfreien Paaren kommen, obwohl $p = 0,6$ ist.

Die Nullhypothese soll abgelehnt werden, wenn mehr als 12 Paar Schuhe einwandfrei sind .

Mit welcher Wahrscheinlichkeit ist dies der Fall?

$$P(X > 12) = P(X \leq 20) - P(X \leq 12) = 1 - 0,584 = 0,416 \quad (\text{siehe Tabelle 2})$$

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. Man würde also mit einer Wahrscheinlichkeit von 41,6% irrtümlicher Weise die Nullhypothese ablehnen.

In beiden Fällen ist die Wahrscheinlichkeit dafür, eine Fehlentscheidung zu treffen ziemlich groß (Fall I 37%, Fall II 41,6%).

Bevor der Test durchgeführt wird, ist es sinnvoll sich dafür zu entscheiden, bei welcher Anzahl von einwandfreien Schuhen man $p = 0,8$ oder $p = 0,6$ für richtig halten will. Eine solche Entscheidung ist willkürlich. Dabei sollte man nicht zu nah am Erwartungswert liegen, damit die Wahrscheinlichkeit für eine Fehlentscheidung nicht zu groß wird.

Es wird folgende **Entscheidungsregel** aufgestellt:

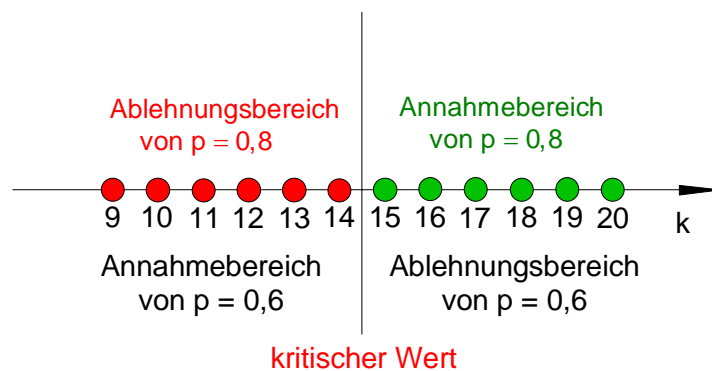
Falls mindestens 15 Paar einwandfrei sind, wird $p = 0,8$ als richtig angesehen, sonst soll $p = 0,6$ gelten.

Die Hypothesen lauten: $H_0: p \geq 0,8$ und $H_1: p < 0,8$.

Aus der Vorgabe folgen Annahme- und Ablehnungsbereich für H_0 .

Annahmebereich: $A = \{ 15 \dots 20 \}$ Ablehnungsbereich: $\bar{A} = \{ 0 \dots 14 \}$

Falls H_0 abgelehnt werden muss, soll $H_1: p < 0,8 = 0,6$ gelten.



Fehlermöglichkeiten dieser Entscheidung:

1. $p = 0,8$ ist richtig, das bedeutet, der neue Lieferant kann wirklich Schuhe höherer Qualität liefern. Zufällig kann es vorkommen, dass weniger als 15 Paar Schuhe einwandfrei sind. Dann würde man dem Lieferanten nicht glauben.

Die Wahrscheinlichkeit einen solchen Fehler zu begehen beträgt

$$P_{80}(X \leq 14) = 0,196 \quad (\text{siehe Tabelle 3}).$$

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

$$P_{60}(X \geq 15) = P(X \leq 20) - P(X \leq 14) = 1 - 0,874 = 0,126 \quad (\text{siehe Tabelle 2}).$$

Das bedeutet, wenn man einen solchen Zufallsversuch mit 20 Paar Schuhe 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.

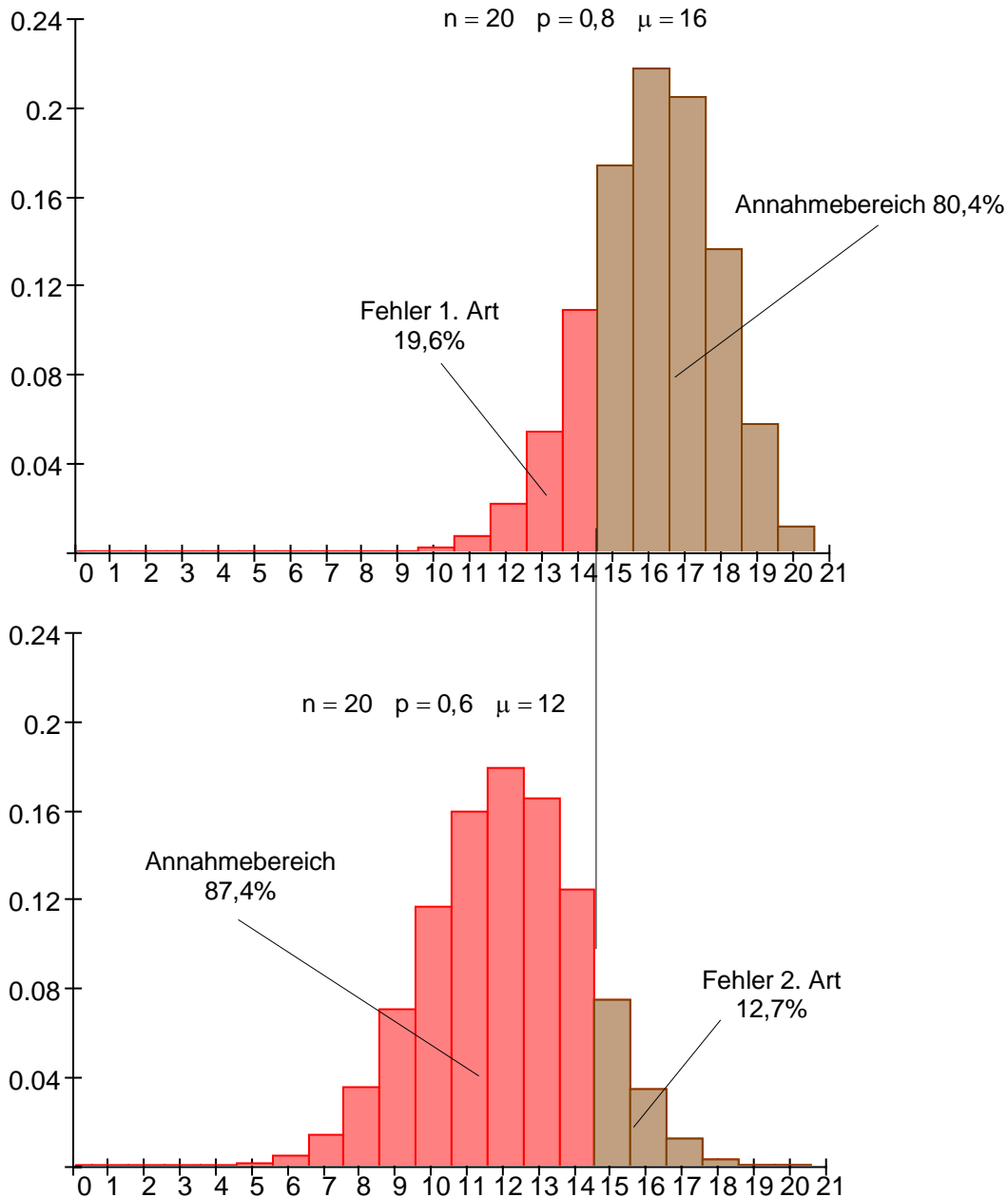
Zusammenfassung der Fehlerarten:

	Versuchsergebnis liegt im Annahmehereich von $H_0 : p \geq 0,8 \quad A = \{ 15 \dots 20 \}$	Versuchsergebnis liegt im Ablehnungsbereich von $H_0 : p \geq 0,8 \quad \bar{A} = \{ 0 \dots 14 \}$
Hypothese $H_0 : p \geq 0,8$ ist wahr	Entscheidung ist richtig, Hypothese $H_0 : p \geq 0,8$ wird angenommen $P_{80}(X \geq 15) = 0,804 \hat{=} 80,4\%$	Entscheidung ist falsch, Hypothese $H_0 : p \geq 0,8$ wird zu unrecht abgelehnt $P_{80}(X \leq 14) = 0,196 \hat{=} 19,6\%$ (Fehler 1. Art)
Hypothese $H_1 : p < 0,8$ $= 0,6$ ist wahr	Entscheidung ist falsch, Hypothese $H_1 : p < 0,8 = 0,6$ wird abgelehnt $P_{60}(X \geq 15) = 0,126 \hat{=} 12,6\%$ (Fehler 2. Art)	Entscheidung ist richtig, Hypothese $H_1 : p < 0,8 = 0,6$ wird angenommen $P_{60}(X \leq 14) = 0,874 \hat{=} 87,4\%$

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. ($\beta = 0,126$)

Um den Fehler 2. Art zu berechnen, betrachtet man den Annahmehereich der Nullhypothese unter der Voraussetzung das die Alternativhypothese gilt. Der Fehler 2. Art ist die Wahrscheinlichkeit dafür, dass ein Testergebnis in den Annahmehereich der Nullhypothese fällt, obwohl die Alternativhypothese gilt.



Irrtumswahrscheinlichkeit wird vorgegeben.

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

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

$P_{0,8}(X \leq k) \leq 0,1 \Rightarrow k = 13$ aus Tabelle 3 abgelesen, denn $P_{0,8}(X \leq 13) = 0,087$

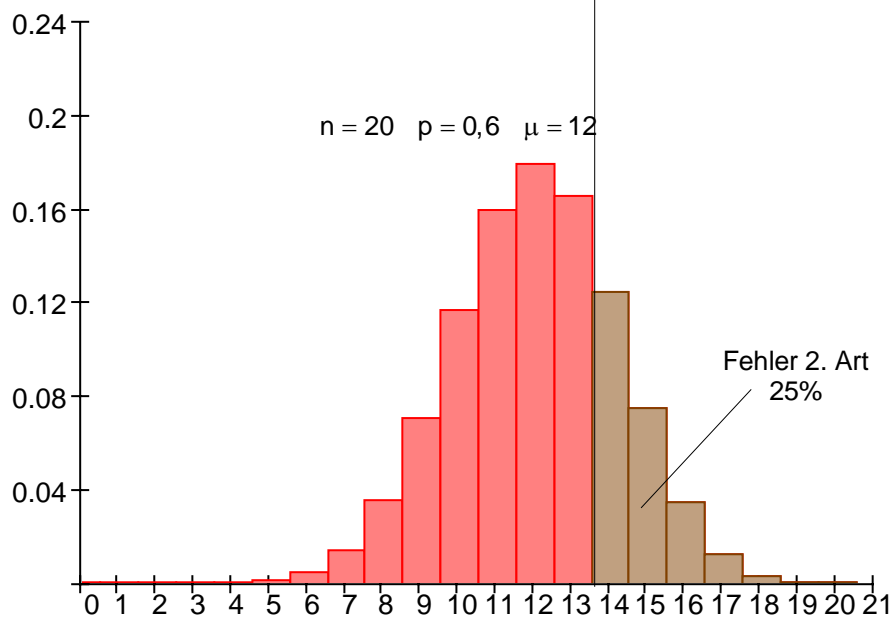
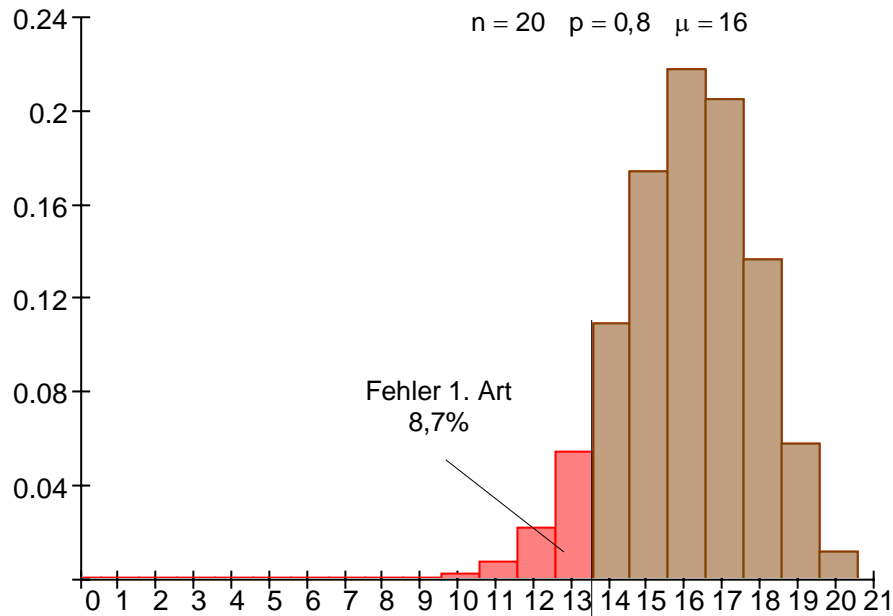
$\Rightarrow \alpha = 0,087 < 0,1$ ist die Wahrscheinlichkeit für einen Fehler 1. Art.

Annahmebereich: $A = \{14, 15, \dots, 20\}$ Ablehnungsbereich: $\bar{A} = \{0, 1, \dots, 13\}$

$P_{0,6}(X \geq 14) = P_{0,6}(X \leq 20) - P_{0,6}(X \leq 13) = 1 - 0,750 = 0,25$ (Tabelle 2)

$\Rightarrow \beta = 0,25$ 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 etwa 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 $n = 20$ und $p = 0,6$

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

Tabelle 3:Kumulierte Binomialverteilung für $n = 20$ und $p = 0,8$

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

Wann ist was wie zu testen?

Aufstellen der Nullhypothese (H_0) und der Alternativhypothese (H_1).

Die Nullhypothese hängt von der jeweiligen Interessenlage ab.

Interessengruppe I behauptet $p > p_0 \Rightarrow H_0 : p \leq p_0 \Rightarrow H_1 : p > p_0$

Interessengruppe II behauptet $p < p_0 \Rightarrow H_0 : p \geq p_0 \Rightarrow H_1 : p < p_0$

In beiden Fällen wird für die Berechnung $p = p_0$ verwendet.

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	$H_0 : p \geq p_0$ zu testen ist. Alternative $H_1 : p < p_0$
Rechtsseitiger Test falls	$H_0 : p \leq p_0$ zu testen ist. Alternative $H_1 : p > p_0$
Beidseitiger Test falls	$H_0 : p = p_0$ zu testen ist. Alternative $H_1 : p \neq p_0$

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