Signifikanztests

Mit einem Signifikanztest kann eine Vermutung (Hypothese) geprüft werden, die nicht direkt bewiesen werden kann.

Beispiele:
Ist ein neues Medikament wirksamer als Placebo (enthält keinen Wirkstoff)?
Hat eine Lieferung die versprochene Qualität?

In den Beispielen kommt häufig die Binomialverteilung vor. Dazu eine kurze Zusammenfassung (Quelle: ETH Malthuis):

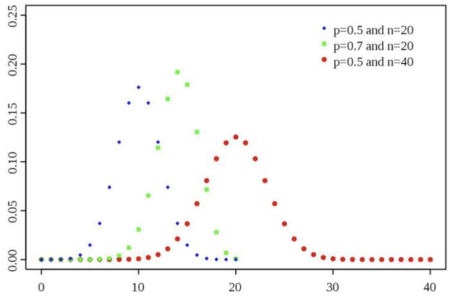
$X \sim Bin(n,p)$

"X ist binomial-verteilt mit Parametern n und p"

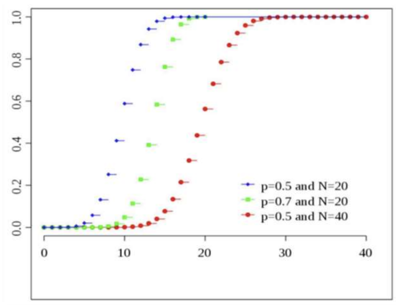
$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x},\ x \in \{0,1,\ldots,n\}$

$E(X) = n \cdot p,\ Var(X) = n \cdot p \cdot (1-p)$

"probability mass function" (pmf)



"cumulative distribution function" (cdf)



1 Test für einen Anteilswert.

Das folgende Beispiel stammt von Sir Ronald Aylmer Fisher (1890-1962), der als Begründer der modernen mathematisch orientierten Statistik gilt. Es illustriert das Vorgehen:

Die englische Lady

Die englische Lady trinkt ihren Tee stets mit einem Zusatz Milch. Eines Tages verblüfft sie ihre Teerunde mit der Behauptung, sie könne allein am Geschmack entscheiden, ob zuerst die Milch oder zuerst der Tee eingegossen worden sei. Sie behauptet, dass sie im Vergleich zum blinden Raten öfter die richtige Eingiess-Reihenfolge treffen würde. Wie könnte die Behauptung der Lady überprüft werden?

1.
Alternativhypothese H_A: Die Lady hat die behauptete Fähigkeit, d.h. ihre Trefferwahrscheinlichkeit p ist grösser als $\frac{1}{2}$

Die Alternativhypothese H_A drückt eine Vermutung aus, die nicht direkt bewiesen werden kann. In den folgenden Beispielen betrifft die Vermutung den Parameter p der Binomialverteilung.

2.
Der Alternativhypothese H_A wird die Nullhypothese H_0 gegenübergestellt:
Die Nullhypothese behauptet die Abwesenheit eines Effekts, den man eigentlich nachweisen möchte:

Nullhypothese H_0: Die Lady rät nur: $p = \frac{1}{2}$

3.
Planen eines Zufallsexperiments

Die Lady erhält n = 20-mal je 2 Tassen Tee überreicht, jeweils eine vom Typ Milch vor Tee, die andere vom Typ Tee vor Milch. Die Reihenfolge dieser beiden Tassen wird durch den Wurf einer Laplace-Münze festgelegt

4.
Es wird für die Zufallsvariable X (im Beispiel die Anzahl der richtigen Antworten) eine **Entscheidungsregel (Verwerfungsbereich V)** festgelegt oder das sogenannte **Signifikanzniveau** (meistens 5%) vorgegeben:

Entscheidungsregel: X sei die Anzahl der Treffer

$X \geq 15$	Entscheid für H_A
$X < 15$	Entscheid für H_0

5.
Das Zufallsexperiment wird durchgeführt und der Wert der Zufallsvariable bestimmt.

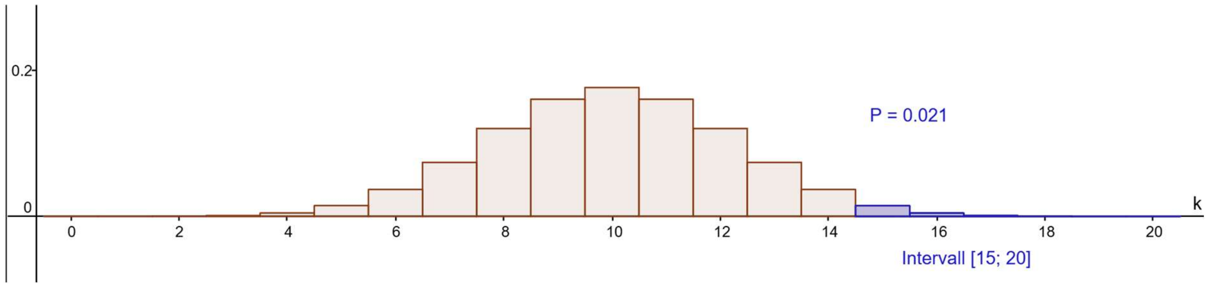
6.
Fehlermöglichkeiten:

Fehler 1. Art α:

H_0 wird verworfen, obwohl sie richtig ist (d.h. man attestiert der Dame eine Fähigkeit, die nicht vorhanden ist (die Dame erzielt also mindestens 14 Treffer durch Raten)

Fehler 2. Art.
H_0 wird beibehalten, obwohl H_A richtig ist. Die Dame hat tatsächlich die behauptete Begabung, erzielt aber weniger als 14 Treffer.

Die Entscheidungsregel 4. wird oft so gewählt, dass die Wahrscheinlichkeit α des Fehlers erster Art kleiner als 0.05 ist. Ein solcher Test heisst Signifikanztest zum **Signifikanzniveau α = 5%**. Bei dieser Vorgehensweise wird der Fehler 1. Art im Vergleich zum Fehler 2. Art als schwerwiegender erachtet. Die meisten Tests werden in der Hoffnung auf eine signifikante Ablehnung der Nullhypothese angelegt.



Ein weiteres Beispiel:

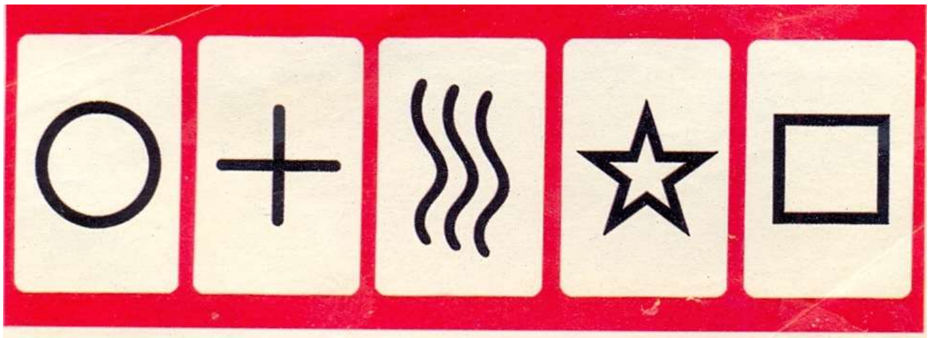
ESP: Extra Sensory Perception, Parapsychologie

Gibt es Menschen, die über eine Begabung verfügen, die sich naturwissenschaftlich nicht erklären lässt?

1.
Alternativhypothese H_A: Die Person verfügt über eine besondere Begabung: $p > \frac{1}{5}$

2.
Nullhypothese H_0: Die Person hat keine besondere Begabung, rät nur: $p > \frac{1}{5}$

3.
Planen eines Zufallsexperiments
Der amerikanische Parapsychologe B. Rhyne (1895 bis 1980) hat folgendes Experiment vorgeschlagen:
Die Testperson muss versuchen, aus den 5 Karten in der Abbildung eine ausgewählte Karte zu identifizieren, ohne sie gesehen zu haben. Es werden $n = 20$ Versuche durchgeführt.


ASW-Symbole: Kreis, Kreuz, Welle, Stern, Quadrat

4.
Testgrösse X: Anzahl der richtig geratenen Karten ist binomialverteilt.

Entscheidungsregel: X sei die Anzahl der Treffer

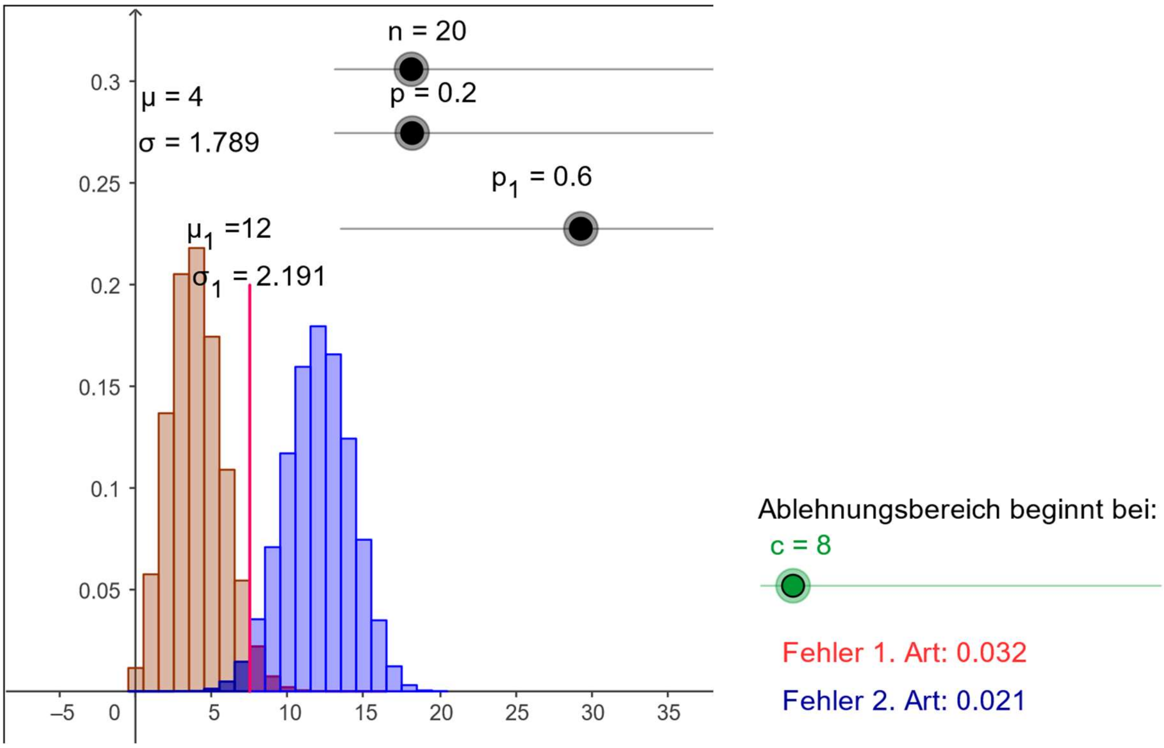
$X > 8$ Entscheid für H_A
$X \leq 8$ Entscheid für H_0

5.
Das Zufallsexperiment wird durchgeführt und der Wert der Zufallsvariable bestimmt.

6.
Fehlermöglichkeiten:

Fehler 1. Art:
Die Wahrscheinlichkeit α, dass einer "normalen" Person irrtümlich besondere Begabung zugeschrieben wird ($p = 0.2$, aber X > 8) die ist 3.2%
Je grösser der Grad aussersinnlicher Begabung ist (das heisst, je grösser p ist), umso kleiner ist die Wahrscheinlichkeit, diese besondere Begabung nicht anzuerkennen.

Fehler 2. Art:
Es handelt sich um die Wahrscheinlichkeit β..Bei einer Testperson wird eine besondere Begabung nicht anerkannt.
Je grösser der Grad aussersinnlicher Begabung ist (das heisst, je grösser p ist), umso kleiner ist die Wahrscheinlichkeit, diese besondere Begabung nicht festzustellen.
Das Risiko 2. Art wird mit β bezeichnet. Es hängt von der Trefferwahrscheinlichkeit der Testperson ab:
Im Beispiel wurde $p' = 0.6$ gewählt.
Verschärft man die Entscheidungsregel (Z > 9) so verkleinert sich zwar der Fehler 1. Art, aber der Fehler 2. Art vergrössert sich.
Will man beide Fehler verkleinern, dann bedingt dies eine Erhöhung der Anzahl n.



Quelle: Unterrichtsmaterialien geogebra (M. Raggl)

Ist wie im folgenden Beispiel $n > \frac{9}{p \cdot q}$, dann kann die Binomialverteilung durch die Normalverteilung approximiert werden.

Multiplechoice-Test

Bei einem Multiple-choice-Test werden einer Kandidatin n Fragen mit z.B. 4 möglichen Antworten vorgelegt, die entweder richtig oder falsch sein können. Die Kandidatin hat alle richtigen Antworten anzukreuzen. In dem folgenden Beispiel wir d eine einfachere Version behandelt.

1.
Alternativhypothese H_A: Die Kandidatin ist eine Könnerin $p > \frac{1}{2}$ (Test einseitig); ihre Trefferwahrscheinlichkeit p ist grösser als $\frac{1}{2}$

2.
Nullhypothese H_0: Die Kandidatin rät nur $p = \frac{1}{2}$

3.
Planen eines Zufallsexperiments
Einer Kandidatin werden $n = 100$ Fragen mit zwei Auswahlantworten vorgelegt, von denen genau eine richtig ist. Es soll geprüft werden, ob die Kandidatin eine Könnerin ist ($p > 0.5$) oder nur rät ($p = 0.5$).

4.
Die Testgrösse X: Anzahl der richtig geratenen Karten.
Die Zufallsvariable X ist binomialverteilt mit dem Erwartungswert $\mu = 50$ und der Standardabweichung $\sigma = 5$. Da die Faustregel erfüllt ist, kann die Binomialverteilung durch die Normalverteilung approximiert werden.
Beim gebräuchlichen Signifikanzniveau $\alpha = 5\%$, ist u so zu wählen, dass $\Phi(u) = 0.95$ gilt. Für die kritische Grenze erhält man damit:

$$50 + 1.645 \cdot \sigma \;=\; 50 + 1.645 \cdot 5 \;=\; 58.2$$

Beantwortet die Kandidatin also mindestens 59 Fragen richtig, dann wird sie als Könnerin eingestuft.

Entscheidungsregel:

Entscheidungsregel:	$X \geq 59$	Entscheid für H_A
	$X < 59$	Entscheid für H_0

5.
Durchführung des Zufallsexperiments
Das Zufallsexperiment wird durchgeführt und der Wert der Zufallsvariable bestimmt.

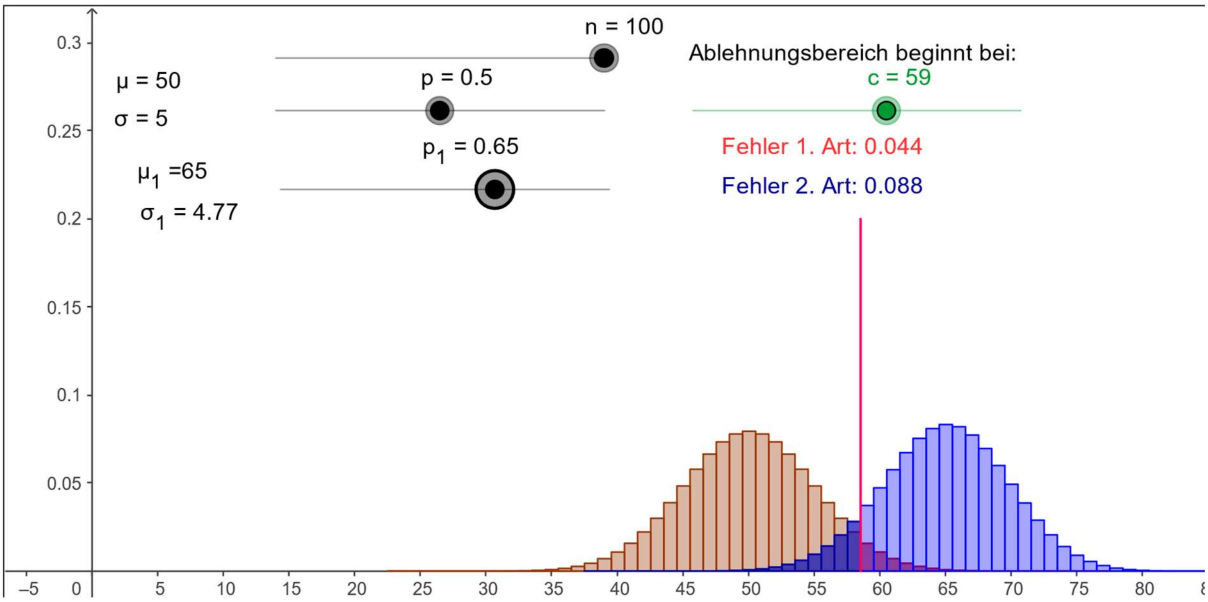
6.
Fehlermöglichkeiten:

Fehler 1. Art α:
Die Nullhypothese wird abgelehnt, obwohl sie richtig ist.
Im Beispiel bedeutet dies: Eine ratende Kandidatin beantwortet, obwohl sie nur rät, mindestens 59 Fragen richtig und wird damit fälschlicherweise als Könnerin eingestuft

Fehler 2. Art: β
Die Nullhypothese ist falsch, wird aber fälschlicherweise beibehalten.
Im Beispiel bedeutet dies: Eine Könnerin (z.B. $p_1 = 0.65$) hat Pech, beantwortet zu wenig Fragen richtig und wird deshalb als Raterin eingestuft.

In der Abbildung sind die beiden Wahrscheinlichkeiten für den Fehler 1. und 2. Art (unter der Annahme $p_1 = 0.65$ zu erkennen.

Die Wahrscheinlichkeit, dass eine ratende Kandidatin die Prüfung besteht (d.h. mindestens 59 Fragen richtig beantwortet ist höchstens 5%.
Die Wahrscheinlichkeit, dass eine Kandidatin mit $p_1 = 0.65$ die Prüfung nicht besteht (d.h. weniger als 59 Fragen richtig beantwortet) ist 8.8%.



Zusammenfassung Signifikanztest:

1.
Alternativhypothese H_A
Aufgrund der Daten ergibt sich eine bestimmte Vermutung (Hypothese) über den unbekannten Parameter einer Wahrscheinlichkeitsverteilung (z.B. über die Erfolgswahrscheinlichkeit bei einem Bernoulliversuch).

2.
Nullhypothese H_0
Die Nullhypothese behauptet Abwesenheit eines Effekts, den man eigentlich nachweisen möchte. Die Nullhypothese möchte man eigentlich widerlegen.

3.
Planen eines **Zufallsexperiment**s

4.
Festlegen einer **Entscheidungsregel** gemäss der Vorgabe des Signifikanzniveaus

5.
Durchführung des Zufallsexperiments

6.
Fehlermöglichkeiten:

Fehler 1. Art:
H_0 wird verworfen, obwohl H_0 richtig ist. Die zugehörige Wahrscheinlichkeit wird mit α bezeichnet, α heisst Signifikanzniveau. Übliches Signifikanzniveau ist $\alpha = 5\%$.

Fehler 2. Art:
H_0 wird beibehalten, obwohl H_0 falsch ist. Der Wert der Testgrösse fällt nicht in den Verwerfungsbereich. Die Wahrscheinlichkeit für diesen Fehler wird mit β bezeichnet.
β kann in der Regel nur durch zusätzliche Annahmen bestimmt werden.

Beispiele für weitere Anwendungen

Ist das neue Medikament besser als Placebo?	Nullhypothese: Sie sind beide gleich gut
Ist ein konkreter Würfel gefälscht?	Nullhypothese: Es handelt sich um einen Laplace-Würfel.
Statistische Qualitätskontrolle:	
Der Ausschussanteil ist zu gross?	Die Lieferung ist in Ordnung

Die beiden Fehler in zwei anderen Kontexten:

In der Justiz: Justizirrtum:	
Ein Schuldiger wird freigesprochen	Ein Unschuldiger wird bestraft
In der Produktion:	
Ein Verfahren wird überschätzt	Ein Verfahren wird unterschätzt
Ein giftiger Pilz wird als essbar eingestuft	Ein essbarer Pilz wird als giftig eingestuft

Je nach den möglichen Auswirkungen des Fehlers wird man versuchen den Fehler 1. oder 2. Art oder beide zu minimieren

Mehrfaches Testen
Testet man eine Hypothese mehrfach mit voneinander unabhängigen Signifikanztests, dann erhöht sich die Wahrscheinlichkeit, dass der kritische Wert übertroffen wird, womit die Nullhypothese fälschlicherweise abgelehnt wird (Fehler 1. Art). Dazu ein Beispiel:

Die Wahrscheinlichkeit, dass von vier voneinander unabhängigen Tests mit dem Signifikanzniveau $\alpha = 0.05$ mindestens einer zufällig signifikant ist, beträgt bereits
$1 - 0.95^4 \approx 0.185$