5 Testen

In diesem Kapitel werden verschiedene Verfahren zum Testen von Hypothesen vorgestellt.

5.1 Signifikanztest

A1 Sechsenwürfel?

Peter und Pauline spielen „Mensch ärgere dich nicht“. Jeder würfelt mit seinem Lieblingswürfel, bis es Streit gibt, weil Pauline so häufig gewinnt. Peter behauptet, Pauline hätte einen „Sechsenwürfel“, bei dem die „6“ häufiger als bei einem „normalen“ Würfel erscheint. Darum würfeln sie nun fünfzigmal mit Paulines Würfel und zählen dabei 13-mal die „6“. Pauline meint, dies sei normal. Peter ist anderer Ansicht.

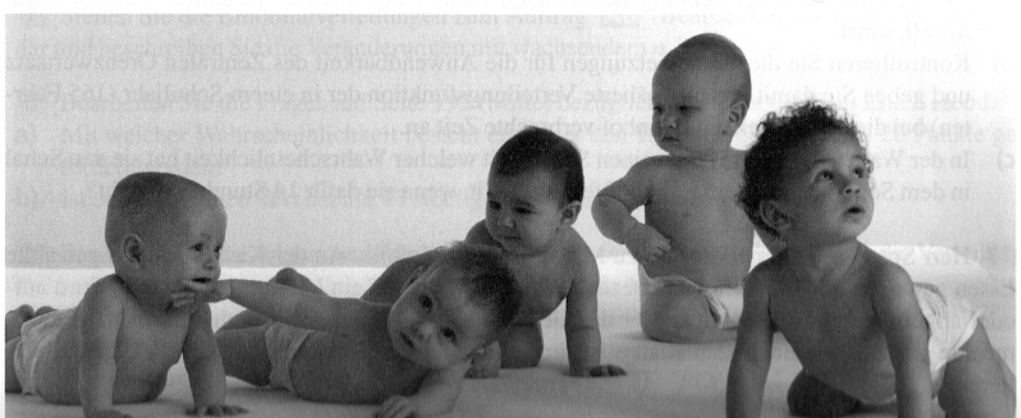
A2 Neue Therapie

Zur Behandlung einer bestimmten Erkrankung wird ein Medikament eingesetzt, das bei 40 % der Behandelten erfolgreich ist.

Ein neues – und hoffentlich erfolgreicheres – Medikament soll mit 100 zufällig ausgewählten Erkrankten getestet werden. Entwickeln Sie ein Entscheidungsverfahren.

A3 Männerüberschuss

Bisher ging man davon aus, dass etwa 50 % der Neugeborenen männlich sind. In einer Stadt waren unter 100 Neugeborenen 60 männlichen Geschlechts. Werden wir eine Männergesellschaft?



Sechsenwürfel?

Wir bearbeiten Auftrag A1. Bisher haben wir Vorgänge untersucht, bei denen die Wahrscheinlichkeit für das Eintreten eines Ereignisses E bekannt war. Beim Würfeln haben wir beispielsweise angenommen, dass ein idealer Würfel vorliegt, dass also gilt:

$$P\,(\text{Eine 6 gewürfelt}) = \tfrac{1}{6}.$$

Peter vermutet jedoch, dass die Wahrscheinlichkeit für das Ereignis {Eine 6 gewürfelt} bei Paulines Würfel größer als $\frac{1}{6}$ ist. Dagegen hält Pauline ihren Würfel für ideal; sie geht also aus von der **Vermutung** oder **Hypothese** $p = \frac{1}{6}$.

Nachdem sie jeweils 50-mal gewürfelt haben, steigert sich der Streit sogar noch, weil sie unterschiedlicher Ansicht über das Ergebnis – bei Pauline erscheint 13-mal die „6“ – sind.

Ein Entscheidungsverfahren mithilfe von Wahrscheinlichkeiten

Peter und Pauline haben eine **Stichprobe** vom Umfang 50 durchgeführt. Kann man durch eine solche Stichprobe entscheiden, wer Recht hat?

Die Stichprobe stellt eine Bernoulli-Kette der Länge 50 dar, die die Ergebnisse $\{6\}$ und $\overline{\{6\}}$ mit $P(\{6\}) = p$ und $P(\overline{\{6\}}) = 1 - p$ hat.

Hinsichtlich der unbekannten Wahrscheinlichkeit p bestehen zwei Hypothesen:

Pauline vermutet: $p = \frac{1}{6}$; Peter hält dagegen: $p > \frac{1}{6}$.

Pauline nimmt also an, dass die Wahrscheinlichkeitsverteilung ihres Würfels mit der eines idealen Würfels übereinstimme, also kein Unterschied zwischen vorliegender und angenommener Verteilung bestehe, der Unterschied „gleich null“ sei. Deshalb heißt Paulines Hypothese auch **Nullhypothese**; man schreibt:

$$H_0\!: \;\; p = p_0 = \tfrac{1}{6}.$$

Peter vermutet dagegen: $p > \frac{1}{6}$. Diese Hypothese wird **Alternativhypothese** genannt und meist mit H_1 bezeichnet; man schreibt:

$$H_1\!: \;\; p > p_0 = \tfrac{1}{6}.$$

Ob man sich für H_0 oder H_1 entscheidet, hängt vom Ergebnis der jeweiligen Stichprobe ab, also von der Anzahl X der dabei auftretenden Sechsen.

X ist damit eine **binomialverteilte Zufallsgröße** mit den Wahrscheinlichkeiten $B(50;\, p;\, k)$, wobei die Grundwahrscheinlichkeit p unbekannt ist.

Die Wahrscheinlichkeit, mindestens 13 Sechsen bei 50 Würfen zu erzielen, kann aus der Tabelle auf Seite 342 abgelesen werden:

$$P(X \geq 13) = 1 - F\left(50;\, \tfrac{1}{6};\, 12\right) = 1 - 0{,}9373 = 0{,}0627,$$

d. h., bei 6,27 % aller 50er Wurfreihen sind mindestens 13 Sechsen zu erwarten; dagegen kann man bei 93,73 % aller Wurfreihen höchstens 12 Sechsen erwarten. Das scheint für Peters Standpunkt und für die Ablehnung der Nullhypothese zu sprechen.

Pauline gibt allerdings zu bedenken: Lehne man H_0 ab, so beruhe das in 6,27 % aller Fälle auf einem Irrtum. Diese **Irrtumswahrscheinlichkeit** von $\alpha = 6{,}27\,\%$ erscheint ihr viel zu groß.

Allgemein ist es üblich, eine Irrtumswahrscheinlichkeit zwischen 1 % und 5 % anzuerkennen. Wählt man als Irrtumswahrscheinlichkeit $\alpha = 5\,\%$, so kann man aus der Bedingung

$$P(X \geq K) \leq \alpha = 0{,}05 \quad (K \in \mathbb{N})$$

bereits **vor** Erhebung einer Stichprobe den **Ablehnungsbereich** $\{K, K+1, \ldots, 50\}$ für die Zufallsgröße X ermitteln. Dazu liest man aus der Tabelle auf Seite 342 den minimalen Werte für K ab, der der Bedingung

$$1 - F\left(50; \tfrac{1}{6}; K-1\right) \leq 0{,}05, \quad \text{also} \quad F\left(50; \tfrac{1}{6}; K-1\right) \geq 0{,}95$$

genügt. Man erhält die **kritische Zahl** $K = 14$, also als Ablehnungsbereich

$$\{14, 15, 16, \ldots, 50\}.$$

Da 13 nicht im Ablehnungsbereich liegt, kann die Nullhypothese für Paulines Würfel nicht abgelehnt werden. Peters Ansicht steht also auf schwachen Füßen.

Das beschriebene Verfahren zur Prüfung einer Hypothese heißt **Signifikanztest** [1].

Fehlentscheidungen beim Signifikanztest

Man wird die Nullhypothese beibehalten, wenn die Zufallsgröße X nicht im Ablehnungsbereich liegt; man wird sie ablehnen, wenn X im Ablehnungsbereich liegt. In beiden Fällen kann die Entscheidung jedoch auch falsch sein:

- **Lehnt** man die Nullhypothese **irrtümlich ab**, obwohl die Behauptung der Nullhypothese tatsächlich zutrifft, dann begeht man einen **Fehler 1. Art**.
- **Behält** man die Nullhypothese **irrtümlich bei**, obwohl die Behauptung der Nullhypothese tatsächlich nicht zutrifft, dann begeht man einen **Fehler 2. Art**.

Beispiele:

1. Angenommen, man hätte bei der Stichprobe mit einem idealen Würfel 20-mal eine Sechs gewürfelt und deshalb die Nullhypothese abgelehnt, dann läge ein Fehler 1. Art vor.
2. Angenommen, man hätte bei der Stichprobe mit einem nicht idealen Würfel 10-mal eine Sechs gewürfelt und deshalb die Nullhypothese beibehalten, dann läge ein Fehler 2. Art vor.

Somit ergeben sich bei einem Signifikanztest folgende Fehlermöglichkeiten:

	Die Behauptung von H_0 trifft zu	Die Behauptung von H_0 trifft nicht zu
H_0 wird abgelehnt	Fehler 1. Art	richtige Entscheidung
H_0 wird beibehalten	richtige Entscheidung	Fehler 2. Art

[1] significare (lat.), Zeichen geben, anzeigen, andeuten, zu erkennen geben

Die Irrtumswahrscheinlichkeit α bezieht sich nur auf einen Fehler 1. Art; sie ist eine obere Schranke für die Wahrscheinlichkeit, eine Nullhypothese irrtümlich abzulehnen, aber keine Abschätzung für die Wahrscheinlichkeit, bei diesem Test überhaupt einen Fehler zu begehen.

Bei unserem Beispiel „Mensch ärgere dich nicht" ist der kritische Wert $K = 14$, also ist der Ablehnungsbereich

$$\{14, 15, 16, \ldots, 50\} \quad \text{(Bild 201/1)}.$$

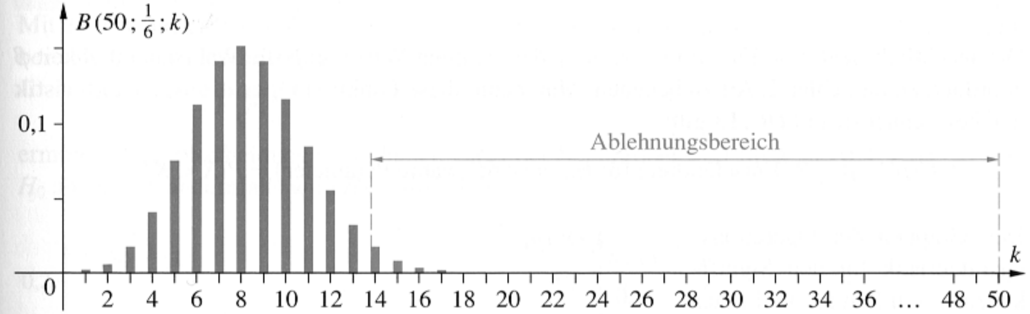


Bild 201/1 : Ablehnungsbereich für H_0: $p = p_0 = \frac{1}{6}$

Damit können wir die Wahrscheinlichkeit für einen Fehler 1. Art ermitteln:

$$P(X \geq 14) = 1 - F\left(50, \tfrac{1}{6}, 13\right) = 1 - 0{,}9693 = 0{,}0307 = 3{,}07\,\%.$$

Wahrscheinlichkeit eines Fehlers 2. Art

Die Wahrscheinlichkeit für einen Fehler 2. Art bezeichnet man mit β; es ist

$$\beta = B(50; p; 0) + B(50; p; 1) + B(50; p; 2) + \cdots + B(50; p; 13) = F(50; p; 13).$$

Da sich H_1: $p > \frac{1}{6}$ nicht wie H_0: $p = \frac{1}{6}$ auf einen einzigen, sondern auf unendlich viele Werte für p bezieht, kann man β nicht allgemein, sondern nur für Beispielwerte berechnen.

Beispiel: $p = 0{,}3 \Rightarrow \beta = F(50; 0{,}3; 13) = 0{,}3279 = 32{,}79\,\%$ (Bild 201/2). ■

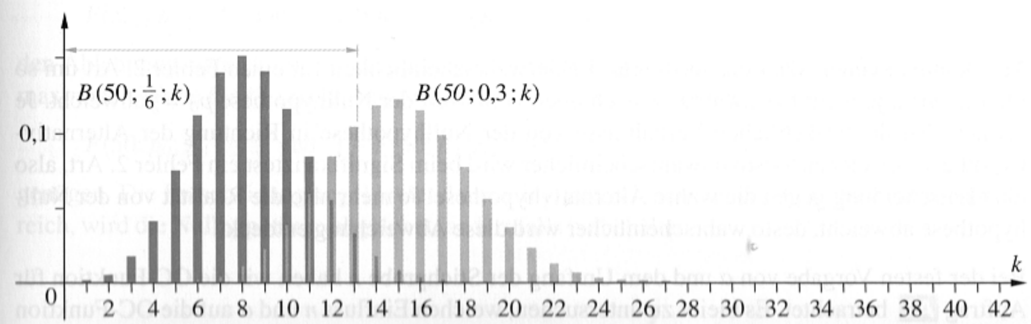


Bild 201/2 : Die grünen Säulen stellen den Fehler 2. Art für $p = 0{,}3$ dar.

Linksseitiger und zweiseitiger Signifikanztest

Peter nörgelt weiter; er meint, bei seinem Würfel falle die 6 zu selten. Die Alternativhypothese zu Paulines Nullhypothese H_0: $p = \frac{1}{6}$ wäre in diesem Fall

$$H_1: p < \frac{1}{6}.$$

Man wird die Nullhypothese in diesem Fall ablehnen, wenn in einer Stichprobe nur wenige Sechsen auftreten. Ein solcher Test heißt deshalb **linksseitiger Signifikanztest**.

Mit der Irrtumswahrscheinlichkeit $\alpha = 0{,}05$ erhält man die kritische Zahl $K = 3$ des Ablehnungsbereichs $\{0; 1; 2; 3\}$, die man aus der Eigenschaft

$$P(X \leq K) = F\left(n; \tfrac{1}{6}; K\right) \leq \alpha$$

ermittelt. Die Entscheidungsregel lautet: Liegt der Stichprobenwert für X in $\{0; 1; 2; 3\}$, so wird H_0 abgelehnt, andernfalls wird H_0 beibehalten.

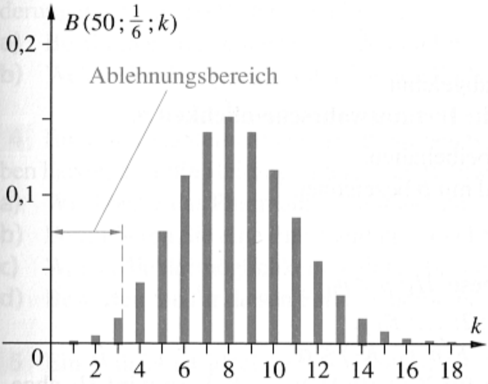


Bild 203/1: linksseitiger Signifikanztest

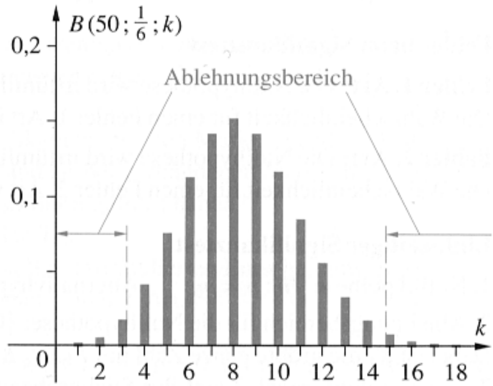


Bild 203/2: zweiseitiger Signifikanztest

Es gibt auch Probleme, bei denen man nicht von vorne herein weiß, in welche Richtung eine Wahrscheinlichkeit p von einer angenommenen Wahrscheinlichkeit p_0 abweicht, ob also $p < p_0$ oder $p > p_0$ ist. Es steht dann der Nullhypothese H_0: $p = p_0$ die Alternativhypothese H_1: $p \neq p_0$ gegenüber. Der sich ergebende Test heißt **zweiseitiger Signifikanztest**.

Dabei verteilt man die Irrtumswahrscheinlichkeit α gleichmäßig auf sehr kleine und sehr große Werte der binomialverteilten Zufallsgröße X und erhält aus den Bedingungen

$$P(X \leq K_1) \leq \tfrac{\alpha}{2} \quad \text{und} \quad P(X \geq K_2) \leq \tfrac{\alpha}{2} \quad (K_1, K_2 \in \mathbb{N})$$

den Ablehnungsbereich $\{0, \ldots, K_1\} \cup \{K_2, \ldots, n\}$. Aus der Tabelle von Seite 342 liest man den maximalen Werte für K_1 und den minimalen für K_2 ab, die den Bedingungen

$$F(50; p_0; K_1) \leq \tfrac{\alpha}{2} \quad \text{und} \quad 1 - F(50; p_0; K_2 - 1) \leq \tfrac{\alpha}{2}$$

genügen. Die Entscheidungsregel lautet dann: Liegt das Stichprobenergebnis im Ablehnungsbereich, wird die Nullhypothese abgelehnt, andernfalls beibehalten.

Wir fassen zusammen

Bei vielen praktischen Problemen ist die Wahrscheinlichkeit $p = P(E)$ für ein Ereignis E nicht bekannt; man vermutet aber, dass p einen bestimmten Wert p_0 hat. Zur Überprüfung dieser Vermutung kann man einen **Signifikanztest** durchführen. Dazu werden zunächst zwei **Hypothesen** aufgestellt: Eine **Nullhypothese** H_0 mit der Annahme, dass $p = p_0$ ist, und eine **Alternativhypothese** H_1 mit der Annahme, dass dem nicht so ist.

Man gibt dann eine kleine **Irrtumswahrscheinlichkeit** α vor – üblich sind $\alpha = 0{,}05$, $\alpha = 0{,}01$ oder $\alpha = 0{,}001$ – und ermittelt den zugehörigen **Ablehnungsbereich** für die Nullhypothese.

Schließlich führt man als **Stichprobe** ein Bernoulli-Experiment mit den Parametern n (Stichprobenumfang) und p (unbekannte Wahrscheinlichkeit $P(E)$) durch und ermittelt so einen Wert für die **binomialverteilte** Zufallsgröße X.

Liefert die Stichprobe für X einen Wert aus dem Ablehnungsbereich, lehnt man die Nullhypothese H_0 ab. Liefert die Stichprobe dagegen für X einen Wert, der außerhalb dieses Ablehnungsbereichs liegt, so behält man die Nullhypothese bei.

Fehler beim Signifikanztest

Fehler 1. Art: Die Nullhypothese wird irrtümlich abgelehnt.
Die Wahrscheinlichkeit für einen Fehler 1. Art ist die **Irrtumswahrscheinlichkeit** α.

Fehler 2. Art: Die Nullhypothese wird irrtümlich beibehalten.
Die Wahrscheinlichkeit für einen Fehler 2. Art wird mit β bezeichnet.

Linksseitiger Signifikanztest

1. Nullhypothese $H_0\colon p = p_0$ Alternativhypothese $H_1\colon p < p_0$
2. Ablehnungsbereich für die Nullhypothese: $\{0; 1; 2; \ldots; K_1\}$.
Dabei ist K_1 die größte ganze Zahl mit $P(X \leq K_1) = F(n; p_0; K_1) \leq \alpha$.
3. Entscheidungsregel: Liegt der Stichprobenwert für X in $\{0; 1; 2; \ldots; K_1\}$, so wird H_0 abgelehnt, andernfalls wird H_0 beibehalten.

Rechtsseitiger Signifikanztest

1. Nullhypothese $H_0\colon p = p_0$ Alternativhypothese $H_1\colon p > p_0$
2. Ablehnungsbereich für die Nullhypothese: $\{K_2; K_2+1; K_2+2; \ldots; n\}$.
Dabei ist K_2 die kleinste ganze Zahl mit $P(X \geq K_2) = 1 - F(n; p_0; K_2 - 1) \leq \alpha$.
3. Entscheidungsregel: Liegt der Stichprobenwert für X in $\{K_2; K_2+1; K_2+2; \ldots; n\}$, so wird H_0 abgelehnt, andernfalls wird H_0 beibehalten.

Zweiseitiger Signifikanztest

1. Nullhypothese $H_0\colon p = p_0$ Alternativhypothese $H_1\colon p \neq p_0$
2. Ablehnungsbereich für die Nullhypothese: $\{0; 1; 2; \ldots; K_1\} \cup \{K_2; K_2+1; K_2+2; \ldots; n\}$.
Dabei ist K_1 die größte ganze Zahl mit $P(X \leq K_1) = F(n; p_0; K_1) \leq \frac{\alpha}{2}$ und K_2 die kleinste ganze Zahl mit $P(X \geq K_2) = 1 - F(n; p_0; K_2 - 1) \leq \frac{\alpha}{2}$.
3. Entscheidungsregel: Liegt der Stichprobenwert für X in $\{0; 1; \ldots; K_1\} \cup \{K_2; K_2+1; \ldots; n\}$, so wird H_0 abgelehnt, andernfalls wird H_0 beibehalten.