Lernzettel - Hypothesentests

Klasse: Qualifikationsphase Abitur

Einstieg: Ist der Würfel aus dem 3D-Drucker fair?

Die Informatik-AG eines Gymnasiums in Bremen hat mit dem 3D-Drucker eigene Spielwürfel hergestellt. Weil die Augen als Vertiefungen gedruckt sind, vermutet Emil, dass die Sechs zu häufig fällt: Die Seite mit sechs Vertiefungen ist leichter, die gegenüberliegende Eins schwerer. Er wirft einen Würfel 120-mal und erhält 27 Sechsen statt der erwarteten 20. Ist das ein Beweis für einen gezinkten Würfel – oder kann so etwas auch bei einem fairen Würfel zufällig passieren?

Genau diese Frage beantwortet ein Hypothesentest. Er legt vorher eine Regel fest, ab welcher Trefferzahl man der Annahme „der Würfel ist fair“ nicht mehr glaubt, und er kontrolliert dabei die Wahrscheinlichkeit, sich zu irren. Hypothesentests werden in Medizin, Industrie, Umfrageforschung und Wissenschaft täglich eingesetzt.

Idee des Hypothesentests

Definition: Hypothesentest

Ein Hersteller behauptet, höchstens $20\,\%$ seiner Geräte seien defekt. Ein Kunde misstraut dieser Aussage. Er prüft eine Stichprobe von 100 Geräten und fragt sich, ab welcher Zahl defekter Geräte er der Behauptung nicht mehr glauben sollte. Genau diese Frage beantwortet ein **Hypothesentest**: Mit einer Stichprobe wird entschieden, ob eine Annahme (die **Nullhypothese** H_0) über eine unbekannte Wahrscheinlichkeit p beibehalten oder zugunsten der **Gegenhypothese** H_1 verworfen wird. Die Zufallsgröße X (Trefferzahl in der Stichprobe) ist unter H_0 binomialverteilt, $X \sim B(n; p_0)$.

Fehler 1. und 2. Art

Da die Entscheidung nur auf einer Stichprobe beruht, kann sie falsch sein. Es gibt zwei Fehlerarten:

- **Fehler 1. Art:** Man lehnt H_0 ab, obwohl H_0 wahr ist. Wahrscheinlichkeit $\alpha = P_{p_0}(X \in \text{Ablehnungsbereich})$.
- **Fehler 2. Art:** Man nimmt H_0 an, obwohl H_1 wahr ist. Wahrscheinlichkeit $\beta = P_{p_1}(X \notin \text{Ablehnungsbereich})$, berechnet für einen konkreten Wert p_1 aus H_1.

Beispiel (Medikament): H_0: „Das Medikament wirkt nicht besser als das alte.“ Fehler 1. Art: Man hält das Medikament für besser, obwohl es das nicht ist – ein unwirksames Mittel kommt auf den Markt. Fehler 2. Art: Man hält ein besseres Medikament für nicht besser – eine Chance wird verpasst. Die Nullhypothese wählt man so, dass der *schwerwiegendere* Fehler der Fehler 1. Art ist; denn nur dessen Wahrscheinlichkeit α wird kontrolliert.

Merke: Fehler 1. und 2. Art

Fehler 1. Art: H_0 wird abgelehnt, obwohl sie wahr ist (Wahrscheinlichkeit $\leq \alpha$). Fehler 2. Art: H_0 wird nicht abgelehnt, obwohl sie falsch ist (Wahrscheinlichkeit β).

Der Signifikanztest

Definition: Ablehnungsbereich

Beim Signifikanztest legt man das Signifikanzniveau α fest (häufig $5\,\%$ oder $1\,\%$) und bestimmt den **Ablehnungsbereich** so, dass die Wahrscheinlichkeit für den Fehler 1. Art höchstens α beträgt. Beim **rechtsseitigen Test** ($H_0 : p \leq p_0$, $H_1 : p > p_0$) besteht er aus großen Trefferzahlen: $\{k, k+1, \ldots, n\}$ mit der kleinsten Zahl k, für die $P(X \geq k) \leq \alpha$ gilt, wenn $X \sim B(n; p_0)$. Beim **linksseitigen Test** ($H_1 : p$) besteht er aus kleinen Trefferzahlen $\{0, \ldots, k\}$ mit der größten Zahl k, für die $P(X \leq k) \leq \alpha$ ist. Beim **zweiseitigen Test** ($H_1 : p \neq p_0$) verteilt man α auf beide Seiten, höchstens $\frac{\alpha}{2}$ je Seite.

Beispiel

$H_0 : p \leq 0{,}2$, $H_1 : p > 0{,}2$, $n = 100$, $\alpha = 5\,\%$. Unter $p_0 = 0{,}2$ ist $X \sim B(100; 0{,}2)$, $\mu = 20$. Es gilt $P(X \geq 28) \approx 0{,}0342 \leq 0{,}05$, aber $P(X \geq 27) > 0{,}05$. Der Ablehnungsbereich ist $\{28, \ldots, 100\}$. Die tatsächliche Wahrscheinlichkeit für den Fehler 1. Art beträgt also nur etwa $3{,}4\,\%$. Zählt der Kunde 27 defekte Geräte, kann er H_0 *nicht* verwerfen; bei 30 defekten verwirft er sie.

Satz: Annahme- und Ablehnungsbereich

Bei einem rechtsseitigen Test mit $H_0 : p \leq p_0$ wird H_0 abgelehnt, wenn $X \geq k$ gilt. Der kleinste Wert k mit $P_{p_0}(X \geq k) \leq \alpha$ legt den **Ablehnungsbereich** fest; α ist die Wahrscheinlichkeit für den Fehler 1. Art.

Fehler 2. Art berechnen

Angenommen, der wahre Wert ist $p_1 = 0{,}3$. Der Fehler 2. Art besteht darin, dass X trotzdem im Annahmebereich $\{0, \ldots, 27\}$ liegt: $\beta = P_{0{,}3}(X \leq 27) \approx 0{,}2964$. Je näher p_1 an p_0 liegt, desto größer ist β. Verkleinert man α, wird der Ablehnungsbereich kleiner, und β wächst. Beide Fehler lassen sich nur gleichzeitig verkleinern, indem man den Stichprobenumfang n erhöht.

Beispiel 1: Zweiseitiger Test

Eine Münze soll auf Fairness geprüft werden: $H_0 : p = 0{,}5$, $H_1 : p \neq 0{,}5$, $n = 100$, $\alpha = 5\,\%$. Wegen der zwei Seiten sucht man links die größte Zahl k_1 mit $P(X \leq k_1) \leq 0{,}025$ und rechts die kleinste Zahl k_2 mit $P(X \geq k_2) \leq 0{,}025$. Es ergibt sich $k_1 = 39$ und $k_2 = 61$. Der Ablehnungsbereich ist $\{0, \ldots, 39\} \cup \{61, \ldots, 100\}$; der

Annahmebereich $\{40, \ldots, 60\}$. Wegen der Symmetrie der Verteilung ($p_0 = 0{,}5$) sind beide Randwahrscheinlichkeiten gleich groß, etwa je $1{,}8\,\%$.

Vorgehen im Abitur

So gehst du vor

1. Zufallsgröße beschreiben und H_0, H_1 mit p formulieren.
2. Verteilung unter H_0 angeben: $X \sim B(n; p_0)$.
3. Ablehnungsbereich über Tabelle oder Taschenrechner bestimmen und die Ungleichung $P(X \geq k) \leq \alpha$ bzw. $P(X \leq k) \leq \alpha$ prüfen.
4. Entscheidungsregel formulieren, Stichprobenergebnis einordnen und die Entscheidung im Sachzusammenhang begründen.

Der Alternativtest

Definition: Alternativtest

Beim **Alternativtest** stehen zwei konkrete Werte zur Wahl: $H_0 : p = p_0$ gegen $H_1 : p = p_1$. Man legt eine Entscheidungsregel fest, zum Beispiel „Entscheidung für H_1, wenn $X \geq k$“, und berechnet beide Fehlerwahrscheinlichkeiten $\alpha = P_{p_0}(X \geq k)$ und $\beta = P_{p_1}(X$. Vergrößert man k, sinkt α und β steigt.

Weitere Beispiele

Beispiel 2: Rechtsseitiger Test

Ein Würfel soll auf zu viele Sechsen getestet werden ($n = 100$, $H_0: p \leq \frac{1}{6}$, $\alpha = 5\,\%$). Es gilt $P(X \geq 24) \approx 0{,}0379$ und $P(X \geq 23) \approx 0{,}0631$. Wie lautet die Entscheidungsregel?

Das kleinste k mit $P(X \geq k) \leq 0{,}05$ ist $k = 24$. Ablehnungsbereich $\{24; \ldots; 100\}$: Bei mindestens 24 Sechsen wird H_0 abgelehnt.

Den Ablehnungsbereich Schritt für Schritt bestimmen

Bei einem linksseitigen Test sucht man das größte k, für das $P(X \leq k) \leq \alpha$ gilt. Mit dem Taschenrechner erstellt man eine Wertetabelle von $P(X \leq k)$ in der Nähe von $\mu - 2\sigma$ und liest die Grenze ab. Notieren Sie im Abitur stets beide Nachbarwerte: den letzten, der noch unter α liegt, und den ersten, der darüber liegt. Nur so ist die Wahl der Grenze begründet.

Beispiel 4: Linksseitiger Test zur Keimfähigkeit

Ein Saatguthersteller garantiert eine Keimfähigkeit von mindestens $90\,\%$. Eine Gärtnerei vermutet, dass sie geringer ist, und sät 200 Samen.

Bestimmen Sie die Entscheidungsregel für $\alpha = 5\,\%$.

$H_0 : p \geq 0{,}9$, $H_1 : p$; $X \sim B(200; 0{,}9)$ im Grenzfall $p = 0{,}9$. Es ist $P(X \leq 172) \approx 0{,}0434 \leq 0{,}05$ und $P(X \leq 173) \approx 0{,}0672 > 0{,}05$. Ablehnungsbereich: $\{0; \ldots; 172\}$. Keimen höchstens 172 Samen, reklamiert die Gärtnerei mit einer Irrtumswahrscheinlichkeit von höchstens $5\,\%$.

Die Wahrscheinlichkeit für den Fehler 2. Art

Der Fehler 2. Art lässt sich nur berechnen, wenn man einen konkreten Wert p_1 aus der Gegenhypothese annimmt. Dann ist β die Wahrscheinlichkeit, dass die Testgröße trotzdem im Annahmebereich landet. Je weiter p_1 von p_0 entfernt ist, desto kleiner ist β: Große Abweichungen entdeckt der Test leicht, kleine kaum. Die Zahl $1 - \beta$ heißt Güte oder Trennschärfe des Tests.

Beispiel 5: Fehler 2. Art beim Saatgut

Wie groß ist beim Test aus Beispiel 4 die Wahrscheinlichkeit für einen Fehler 2. Art, wenn die Keimfähigkeit in Wahrheit nur $85\,\%$ beträgt?

Annahmebereich $\{173; \ldots; 200\}$, jetzt mit $p_1 = 0{,}85$: $\beta = P(X \geq 173) = 1 - P(X \leq 172) \approx 0{,}3166$. In etwa $32\,\%$ der Fälle bemerkt die Gärtnerei die schlechtere Ware nicht. Eine größere Stichprobe würde β verkleinern.

Was ein Test aussagt – und was nicht

Ein Hypothesentest ist ein Entscheidungsverfahren unter Unsicherheit. Liegt das Ergebnis im Ablehnungsbereich, spricht man von einem **signifikanten** Ergebnis: Die Daten wären unter H_0 sehr unwahrscheinlich. Das beweist nicht, dass H_0 falsch ist. Liegt das Ergebnis im Annahmebereich, ist H_0 ebenfalls nicht bewiesen; die Daten reichen nur nicht aus, um sie zu verwerfen. Deshalb legt man als H_1 die Aussage fest, die man mit kontrollierter Irrtumswahrscheinlichkeit belegen möchte. Wer etwa zeigen will, dass ein neues Medikament besser wirkt, wählt „wirkt nicht besser“ als Nullhypothese. Wichtig ist außerdem die Reihenfolge: Hypothesen, n und α werden vor dem Versuch festgelegt; die Regel nachträglich an das Ergebnis anzupassen, ist unwissenschaftlich.

Beispiel 6: Emils Würfel

Entscheiden Sie Emils Frage mit einem rechtsseitigen Test ($n = 120$, $\alpha = 5\,\%$).

$H_0 : p \leq \frac{1}{6}$, $H_1 : p > \frac{1}{6}$. Es gilt $P(X \geq 27) \approx 0{,}0597 > 0{,}05$ und $P(X \geq 28) \approx 0{,}0373 \leq 0{,}05$. Ablehnungsbereich $\{28; \ldots; 120\}$. Mit 27 Sechsen wird H_0 knapp nicht abgelehnt; Emil sollte mehr Würfe machen.

Wusstest du? Die Teetrinkerin

Der englische Statistiker Ronald A. Fisher beschrieb 1935 eine Kollegin, die behauptete, am Geschmack zu erkennen, ob zuerst Milch oder zuerst Tee in die Tasse gegossen wurde. Fisher plante dazu ein Experiment mit acht Tassen und überlegte

vorher, wie viele richtige Antworten nötig sind, damit Raten unwahrscheinlich wird. Daraus entstand die moderne Testtheorie, die heute zum Beispiel bei der Zulassung jedes Medikaments vorgeschrieben ist.

Der p-Wert als Alternative

Statt einen Ablehnungsbereich zu bestimmen, kann man auch fragen: Wie wahrscheinlich ist unter H_0 ein Ergebnis, das mindestens so extrem ist wie das beobachtete? Diese Wahrscheinlichkeit heißt p-Wert. Ist sie kleiner oder gleich α, wird H_0 abgelehnt. Beide Verfahren führen immer zur selben Entscheidung, der p-Wert zeigt aber zusätzlich, wie knapp sie ausgefallen ist.

Merke: Das Wichtigste auf einen Blick

- Nullhypothese H_0 (bisherige Annahme, Behauptung, die man widerlegen will) und Gegenhypothese H_1 (das, was man nachweisen möchte); Testgröße X = Anzahl der Treffer in einer Stichprobe vom Umfang n.
- Fehler 1. Art: H_0 wird abgelehnt, obwohl H_0 wahr ist (Wahrscheinlichkeit höchstens α). Fehler 2. Art: H_0 wird beibehalten, obwohl H_1 wahr ist (Wahrscheinlichkeit β).
- Rechtsseitiger Test ($H_1 : p > p_0$): Ablehnungsbereich $\{k; \ldots; n\}$ mit kleinstem k, für das $P_{p_0}(X \geq k) \leq \alpha$ gilt.
- Linksseitiger Test ($H_1 : p$): Ablehnungsbereich $\{0; \ldots; k\}$ mit größtem k, für das $P_{p_0}(X \leq k) \leq \alpha$ gilt.
- Zweiseitiger Test ($H_1 : p \neq p_0$): auf jeder Seite höchstens $\frac{\alpha}{2}$.
- β berechnet man mit einem konkreten p_1 aus H_1 als Wahrscheinlichkeit des Annahmebereichs; kleineres α bedeutet größeres β, größeres n verkleinert beide.
- Ein Testergebnis ist kein Beweis: Das Ablehnen von H_0 ist eine signifikante Aussage, das Beibehalten bedeutet nur, dass die Daten nicht gegen H_0 sprechen.
- Die Entscheidungsregel wird vor der Stichprobe festgelegt; die Wahl der Nullhypothese richtet sich danach, welcher Fehler schwerer wiegt.

Achtung: Typische Fehler

- Hypothesen vertauschen: H_0 enthält die Gleichheit ($p = p_0$, $p \leq p_0$ oder $p \geq p_0$).
- Den Ablehnungsbereich mit der falschen Seite (links statt rechts) formulieren.
- Grenze des Ablehnungsbereichs falsch bestimmen (Verschiebung um 1 bei $P(X \geq k) = 1 - P(X \leq k - 1)$).
- Fehler 1. und 2. Art verwechseln oder β mit der Wahrscheinlichkeit unter H_0 berechnen.
- Das Ergebnis des Tests als Beweis auffassen: Ein Test liefert eine Entscheidung mit Irrtumswahrscheinlichkeit, keinen Beweis.

Ottos Tipp: Formuliere immer zuerst H_0, H_1, n, α und die Verteilung unter H_0.

Danach ist die Rechnung nur noch Routine!

Ottos Tipp

Formulieren Sie die Entscheidung immer in einem Antwortsatz mit dem Sachbezug, zum Beispiel: „Bei mindestens 24 Sechsen wird angenommen, dass der Würfel zu oft die Sechs zeigt.“