Stochastik für Info SoSe 2023 TU Berlin
Statistische Tests

Hanno Gottschalk

June 28, 2023

Inhaltsverzeichnis der Vorlesung

Grundlagen der Testtheorie

Worum geht es beim Testen?

Starrsinn ist. . .

Meine Meinung steht fest verwirren Sie mich nicht mit Fakten!

Statistik ist. . .

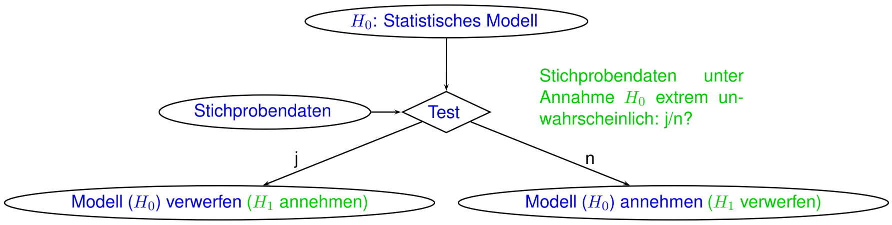
Meine Meinung steht zwar fest, aber ich lasse mich vom Gegenteil überzeugen, wenn die Daten SEHR eindeutig gegen meine Meinung sprechen.

'Meinung' = Nullhypothese H_0

'sehr eindeutig' = signifikanter Ausgang des Tests

Formalisieren des Tests

Statistischer Test



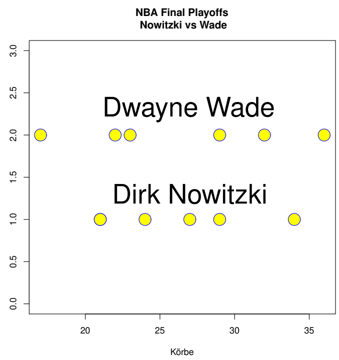
- H_0: Ein statistisches Modell für 'Experiment' mit Parameter θ_0
- Stichprobendaten: Tatsächlicher Ausgang des 'Experiments'
- Testentscheidung: Vergleich des *beobachteten* Wertes einer *Statistik* mit deren Quantil unter H_0

Testen ohne Rechnen - Nowitzki vs Wade

H_0 Beide Spieler waren in den Play-Offs 2011 Spielen gleich gut. . .

'Experiment': NBA-Playoffs 2011

Unterschied nur schwer auszumachen, bleiben erstmal bei H_0



Signifikanzniveau und Fehler 1. Art

Frage: Was genau heisst 'extrem unwahrscheinlich'?

- Stelle Nullhypothese H_0 incl. stat. Modell und Parameter θ_0 auf und formuliere Gegenhypothese H_1
- Gebe *Fehlerniveau* $1 > \alpha > 0$ vor, z.B. $\alpha = 5\% = 0.05$
- Kostruiere eine Statistik S (etwa Mittelwert), die in Bezug auf Problemstellung aussagekräftig ist
- Konstruiere einen *Ablehnungsbereich* A für die Statistik so dass A die Werte enthält, die am schechtesten zu H_0 passen so dass

$$P(S \in A|H_0) = \alpha. \tag{1}$$

Signifikanzniveau und Fehler 1. Art II

- Führe Experiment durch und berechne S für experimentelle Daten
- Führe Testentscheidung durch: $S \in A \Rightarrow H_0$ verwerfen und H_1 annehmen, sonst H_0 beibehalten

Def: (i) Der *Fehler erster Art* ist, dass ich H_0 verwerfe, obwohl H_0 richtig war

(ii) α ist die W-keit für den Fehler erster Art unter H_0 und wird *Signifikanzniveau* genannt.

- Je kleiner das Signifikanzniveau, desto schwieriger ist es H_0 zu wiederlegen
- Übliche Signifikanzniveaus $\alpha = 10\%, 5\%, 1\%, 0.1\%$.

Null- und Gegenhypothese in der empirischen Forschung

In der empirischen Forschung nehmen Sie das *Gegenteil* dessen, was Sie beweisen möchten, als Nullhypothese an.

Beispiel: Wollen 'beweisen': Heilerfolg bei Einnahme von Medikament ist im Mittel *größer* als ohne.

Nullhypothese H_0: Heilerfolg bei Vergabe von Medikament kleiner oder gleich Heilerfolg ohne Vergabe dieses Medikamentes.

Wirksamkeit von Medikament gilt dann als erwiesen, wenn der in der Studie ermittelte mittlere Heilerfolg im Ablehnungsbereich von H_0 ist

$\Rightarrow$ Müssen entweder annehmen, dass ein sehr seltenes Ereignis (W-keit α) eingetreten ist, oder dass H_1 richtig ist — Medikament wirkt

Fehler 2. Art

Def.: Der *Fehler 2. Art* oder β-Fehler tritt dann auf, wenn H_0 beibehalten wird, obwohl H_1 richtig war.

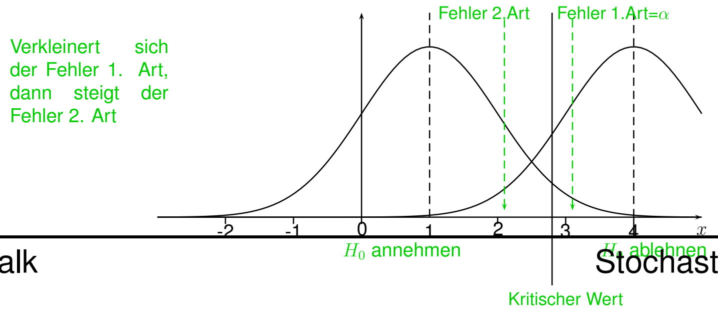
Testergebnis/Wahrheit	H_0	H_1
H_0	o.k.	Fehler 2.Art
H_1	Fehler 1. Art	o.k.

Da H_1 i.A. NICHT mit einem statstischen Modell verbunden ist, kann der β-Fehler in der Regel nicht quantifiziert werden

Je kleiner α, desto größer β — und umgekehrt

Fehler 1. und 2. Art im Alternativtest

Def: Ein *Alternativtest* liegt dann vor, wenn sowohl H_0 als auch H_1 mit einem stat. Modell verbunden sind, und man sich zwischen diesen beiden Möglichkeiten entscheiden muss.



Tests als Entscheidungsproblem

Oft muss man aufgrund von empirischen Daten *eine Entscheidung treffen*.

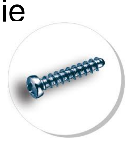
Beispiel:

- H_0 Die Standard-Balliste ist mindestens genauso gut, wie irgendeine andere - ich bleib dabei!
- Ich will mir den Ärger, eine neue Balliste einzuführen (H_1), wenn die neue gar nicht besser ist, nur in $\alpha = 10\%$ der Fälle einbrocken!

Im Entscheidungsproblem ist α das Risiko, sich fälschlicher Weise von der *baseline* H_0 zu entfernen. Kleines α führt also zu *konservativer* Entscheidungsfindung.

Beispiel Qualitätskontrolle

- Bei der Qualitätskontrolle werden aus der Produktion von Schrauben immer wieder Stichproben von $n = 200$, Schrauben entnommen. Emp. arithm. Mittel $\bar{x} = 1.004$cm.
- Die normale Maschinenungenauigkeit führe zu einer Standardabweichung von 0.01 cm.
- Die Nullhypothese H_0 besagt: H_0: Die Schrauben haben die durchschnittliche Länge von 1cm.
- Man will die Maschine nicht unnötig anhalten. Deshalb möchte man, dass die Maschine, wenn sie richtig eingestellt ist, nur bei jeder 100 Stichprobe fälschlicherweise angehalten wird ($\alpha = 1\%$).



Beispiel Qualitätskontrolle II

Schritt 1: Formulierung der Nullhypothese:

- Für $X =$ Schraubenlänge nehmen wir also die Verteilung $N(1, (0.01)^2)$ an.
- Jede einzelne der 200 Schraubenentnahmen X_j sei unabhängig (Produktmodell) $\Rightarrow$

$$\bar{X} = \frac{1}{200}\sum_{j=1}^{n} X_j \sim N\left(1, \frac{(0.01)^2}{200}\right) \Rightarrow Z = \frac{\sqrt{200}}{0.01}(\bar{X} - 1) \sim N(0,1)$$

Nullhypothese H_0**:** $\mu_X = 1$

Gegenhypothese H_1**:** $\mu_X \neq 1$

Beispiel Qualitätskontrolle III

Schritt 2: Konstruktion des Ablehnungsbereiches.

Statistik: $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$, $Z = \frac{\sqrt{200}}{0.01}(\bar{X} - 1)$

c die kritische Abweichung, bei der H_0 nicht mehr geglaubt wird. Durch *Standadisieren* erhält man Ablehnungsbereich

$$P(|\bar{X} - 1| > c) = P\left(|Z| > c\frac{\sqrt{200}}{0.01}\right) = \alpha = 0.01$$

Also: Also (mit $\alpha = 0.01$)

$$\begin{aligned} z_{1-\alpha/2} &= c\frac{\sqrt{200}}{0.01} \Leftrightarrow \\ c &= z_{1-\alpha/2}\frac{0.01}{\sqrt{200}} = 2.5758\frac{0.001}{\sqrt{200}} = 0.00182 \end{aligned}$$

Beispiel Qualitätskontrolle IV

Schritt 3: Testentscheidung

Empirisch gefundene Abweichung des arithmetischen Mittels:

$$|\bar{x} - 1| = |1.004 - 1| = 0.004 > 0.00182 = c$$

$\Rightarrow$ Die Teststatistik liegt im Ablehnungsbereich

Die Nullhypothese H_0 wird verworfen, H_1 wird angenommen — Die Maschine ist nicht ganz präzise eingestellt!

Gaußtest

Def.: Es sei X eine normalverteilte Z.V. mit bekannter Varianz σ_X^2. Die Hypothese

Nullhypothese: $H_0 : \mu_X = \mu_0$
wird bei einer Stichprobe vom Umfang n gegen die
Gegenhypothese: $H_1 : \mu_X \neq \mu_0$
zum Signifikanzniveau $\alpha > 0$ getestet

Ergebnis der Stichprobe sei das arithmetische Mittel $\bar{x}$ (Teststatistik).

Dann verläuft die **Testentscheidung** folgendermaßen:
H_0 wird angenommen, falls $|\mu_0 - \bar{x}| \leq z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}$ und H_0 wird abgelehnt falls $|\mu_0 - \bar{x}| > z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}$

Gaußtest und Konfidenzbereich

Mittels der beidseitigen Konfidenzintervalle zum Konfi-Niveau $(1-\alpha)$ formuliert man das:

Es gilt: H_0 wird genau dann angenommen wenn μ_0 im 2-seitigen Konfidenzintervall ist

Denn: H_0 wird angenommen, falls $\bar{x} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}} \leq \mu_0 < \bar{x} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}$ und ansonsten abgelehnt.

Einseitiger Gaußtest auf Überschreitung

Ebenfalls gibt es *einseitige Gaußtests* (vorauss. wie Gaußtest):

Es sei X eine normalverteilte Z.V. mit bekannter Varianz σ_X^2. Die Hypothese besagt, dass der wahre Erwartungswert μ_X *unter* einem kritischen Wert μ_0 liegt:
Nullhypothese: $H_0 : \mu_X \leq \mu_0$

Sie wird mittels einer Stichprobe vom Umfang n zum Signifikanzniveau α getestet gegen die
Gegenhypothese $H_1 : \mu_X > \mu_0$

Ergebnis der Stichprobe sei das arithmetische Mittel $\bar{x}$.

Testentscheidung:
H_0 wird angenommen, falls $\mu_0 \geq \bar{x} - z_{1-\alpha}\frac{\sigma_X}{\sqrt{n}}$ und H_0 wird abgelehnt falls $\mu_0 < \bar{x} - z_{1-\alpha}\frac{\sigma_X}{\sqrt{n}}$

Einseitiger Gaußtest auf Unterschreitung

Voraussetzungen wie unter Gaußtest

Nullhypothese $H_0 : \mu_X \geq \mu_0$

Gegenhypothese $H_1 : \mu_X < \mu_0$

Testentscheidung:
H_0 wird angenommen, falls $\mu_0 \leq \bar{x} + z_{1-\alpha}\frac{\sigma_X}{\sqrt{n}}$ und H_0 wird abgelehnt falls $\mu_0 > \bar{x} + z_{1-\alpha}\frac{\sigma_X}{\sqrt{n}}$

Vorüberlegungen

Der t-Test ist ein Test für gaußverteilte Z.V. auf den Erwartungswert μ_X, bei dem die *Varianz unbekannt* ist.

Unterschied zum Gaußtest:

- Verteilung durch die Hypothese H_0 nur teilweise festgelegt
- Varianz muß *aus der Stichprobe geschätzt* werden

Dieser Unterschied führt zu demselben Effekt, wie bei den Konfi-Intervallen:

- $z_{1-\alpha/2}$ bzw. $z_{1-\alpha}$-Quantile müssen durch $t_{1-\alpha/2}$ bzw. $t_{1-\alpha}$-Quantile ersetzt werden.
- Ersetze σ_X durch empirische Standardabweichung $\hat{\sigma}$

t-Test: Die Testentscheidungen

Es sei $X \sim N(\mu_X, \sigma_X^2)$ mit μ_X und σ_X^2 unbekannt (alle t-Quantile mit $n-1$ Freiheitsgrafen)

t-Test

$$H_0 : \mu_X = \mu_0 \text{ gegen } H_1 : \mu_X \neq \mu_0, \; H_0 \text{ annehmen falls } \mu_0 = \bar{x} \pm t_{1-\alpha/2}\frac{s}{\sqrt{n}}$$
$$H_0 : \mu_X \leq \mu_0 \text{ gegen } H_1 : \mu_X > \mu_0, \; H_0 \text{ annehmen falls } \mu_0 \geq \bar{x} - t_{1-\alpha}\frac{s}{\sqrt{n}}$$
$$H_0 : \mu_X \geq \mu_0 \text{ gegen } H_1 : \mu_X < \mu_0, \; H_0 \text{ annehmen falls } \mu_0 \leq \bar{x} + t_{1-\alpha}\frac{s}{\sqrt{n}}$$

H_0 wird ansonsten abgelehnt und H_1 angenommen.

(Hier $s = \hat{\sigma}$ empirische Standardabweichung)

Falls $n > 30$ kann man den t-Test auch ohne die Normalverteilungshypothese als *approximativen* Test durchführen.

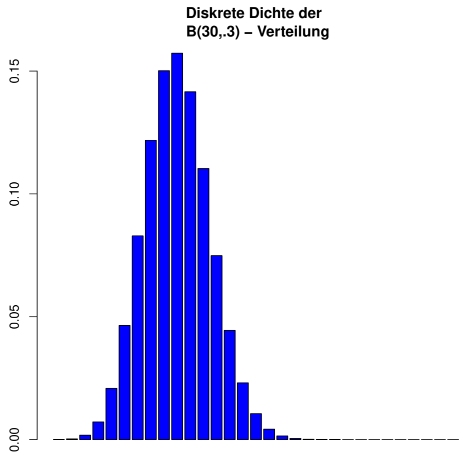
Vorüberlegungen

Betrachten wir eine Grundgesamtheit, in der jedes Element eine Eigenschaft E hat oder nicht.

- Der $p =$*Anteil* der Elemente mit Eigenschaft E
- $X =$ Anzahl der Elemente in einer Zufallsstichprobe vom Umfang n, welche Eigenschaft E haben
- $X \sim B(n, p)$ ist also die *Zählvariable der Zufallsstichprobe*
- $\bar{X} = X/n = \frac{1}{n}\sum_{j=1}^{n} X_j$, $X_j = 1$ falls j Eigenschaft E hat $X_j = 0$ sonst

Faustregel: $n \geq 30$ und $np > 10$ sowie $n(1-p) > 10 \Rightarrow \bar{X} \sim N(p, \frac{p(1-p)}{n})$ approximativ

Dichte Bonomialverteilung



Approximativer Binomialtest

Def.: *approximativer Binomialtest* auf Anteilswert: Für $n > 30$ und $pn > 10$ ist (approximativ)

Approximativer Binomialtest

$$H_0 : p = p_0 \text{ gegen } H_1 : p \neq p_0,\ H_0 \text{ annehmen falls } p_0 = \bar{x} \pm z_{1-\alpha/2}\sqrt{\tfrac{p_0(1-p_0)}{n}}$$
$$H_0 : p \leq p_0 \text{ gegen } H_1 : p > p_0,\ H_0 \text{ annehmen falls } p_0 \geq \bar{x} - z_{1-\alpha}\sqrt{\tfrac{p_0(1-p_0)}{n}}$$
$$H_0 : p \geq p_0 \text{ gegen } H_1 : p < p_0,\ H_0 \text{ annehmen falls } p_0 \leq \bar{x} + z_{1-\alpha}\sqrt{\tfrac{p_0(1-p_0)}{n}}$$

H_0 wird ansonsten abgelehnt und H_1 angenommen.

Bemerkung: Auch ein *exakter Binomialtest* existiert, bei dem der Ablehnungsbereich mittels der Binomialverteilung konstruiert wird. . .

Binomial shoot out (?) - das ist die Lage

Sie sitzen ganz entspannt im Saloon von Tombstone bei einer Partie Würfeln

Ihr Gegenüber (Typ: 'Dunkle Sonnenbrille') hat eine Glückssträne und würfelt

6	6	6	3	6	2	6	4	1	6	6	4	5	4	6
3	4	4	5	4	6	2	6	4	5	5	2	1	6	6
5	4	4	3	4	2	5	6	1	1	6	6	2	2	2
5	5	5	2	1	5	3	2	2	2	1	3	1	6	4

Ist das noch Glück. . .

. . . oder schon Falschspiel?

Binomial shoot out (?) - man ist nicht gerne ungerecht...

Sie möchten nur sehr ungerne jemand fälschlicher Weise erschiessen (Irrtumswahrscheinlichkeit 10%)...

... aber 16 mal die Sechs in 60 Würfen, ist das normal?

Nullhypothese H_0**:** $1/6 = p_0 \geq p = p(W = 6)$, $H_1 : p > 1/6$

Testentscheidung:

$$\frac{16}{60} - z_{0.9}\sqrt{\frac{\frac{1}{6}\cdot\frac{5}{6}}{60}} = 0.205 > 1/6 = 0.166\bar{6}$$

H_0 wird verworfen