Vorlesung: Statistik II für Wirtschaftswissenschaft

Prof. Dr. Helmut Küchenhoff

Institut für Statistik, LMU München

Sommersemester 2017

Hypothese

„Behauptung einer Tatsache, deren Überprüfung noch aussteht“ (Leutner in: Endruweit, Trommsdorff: Wörterbuch der Soziologie, 1989).

Statistischer Test: Überprüfung von Hypothesen anhand einer Stichprobe

Idealtypische Vorgehensweise
Wissenschaftlicher Fortschritt durch Falsifikation von Hypothesen

Statistische Testtheorie

Schließe von Stichprobe oder Experiment auf Grundgesamtheit bzw. Allg. Gesetz

Vorgehen:

- Inhaltliche Hypothese aufstellen
- Operationalisierung
- Inhaltliche Hypothese in statistische Hypothese „Übersetzen“
- Statistischer Test

Bemerkungen

- **Statistische Tests:**
  Die am häufigsten verwendete Art statistischer Inferenz
- **Statistische Signifikanz:**
  Zentrales Argument bei vielen empirischen Arbeiten
- **Voraussetzung für Testverfahren:**
  Zufallsstichprobe oder Experiment

Ist ein beobachtetes Phänomen in Stichproben ein **reines Zufallsprodukt** oder **mit großer Sicherheit** auf einen **realen Effekt** zurückzuführen?
$\longrightarrow$ Dazu notwendig:
Formale Entscheidungsregel = Statistischer Test

Beispiel: Münzdrehen (2€)

Zeitungsberichte: 2€ Münzen nicht „fair“

Münzhypothese

- Vermutung:
  2€- Münze nicht fair
- Überprüfung: 10-Mal die Münze werfen, Anzahl „Zahl" notieren

Mögliche Ergebnisse des Experiments

- 5-Mal "Zahl"
  $\longrightarrow$ deutet nicht auf eine unfaire Münze hin
- 10-Mal "Zahl"
  $\longrightarrow$ verdächtig, die Münze ist vermutlich nicht fair
- 0-Mal "Zahl"
  $\longrightarrow$ verdächtig, die Münze ist vermutlich nicht fair
- 8-Mal "Zahl"
  $\longrightarrow$?? mehr Zahlwürfe als erwartet. Zufall? Oder Münze nicht fair?

Münzhypothese

- Vermutung:
  2€- Münze nicht fair
- Statistische Formulierung:
  X Bernoulli-Variable

$$X = \begin{cases} 1 & \text{"}Zahl\text{"} \\ 0 & \text{"}Adler\text{"} \end{cases}$$

- Wahrscheinlichkeit für Zahl

$$\pi = P(X = 1)$$

- „Die Münze ist nicht fair" heißt

$$\pi \neq 0.5$$

Überprüfung der Münzhypothese

- Experiment: Wir werfen $n = 10$-Mal die Münze

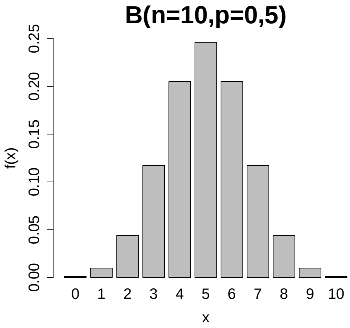
$$\sum_{i=1}^{n} X_i \quad \sim \quad B(n = 10, \pi)$$

- Welche Ergebnisse sind wahrscheinlich, falls die Münze fair ist?
- **Falls die Münze fair ist**, so ist die Anzahl „Zahl" binomialverteilt mit $p = 0.5$.

$$\sum_{i=1}^{10} X_i \quad \sim \quad B(n = 10, \pi = 0.5)$$

- **Falls die Münze fair ist**, so sollte $\sum_{i=1}^{10} X_i$ mit einer Wahrscheinlichkeit von 95 % nicht weit entfernt vom Erwartungswert 5 liegen.

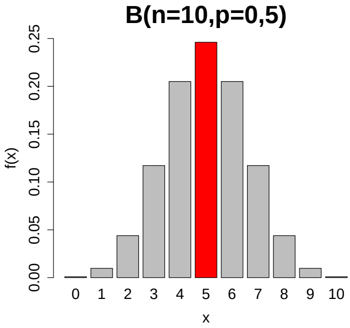
Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001

$$\sum =$$

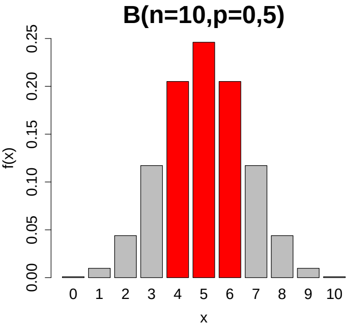
Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117	0.205	0.246 0.246	0.205	0.117	0.044	0.010	0.001

$$\sum = 0.246$$

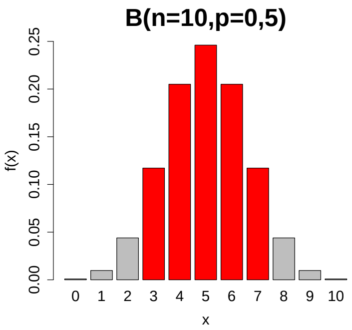
Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117	0.205 0.205	0.246 0.246	0.205 0.205	0.117	0.044	0.010	0.001

$$\sum = 0.656$$

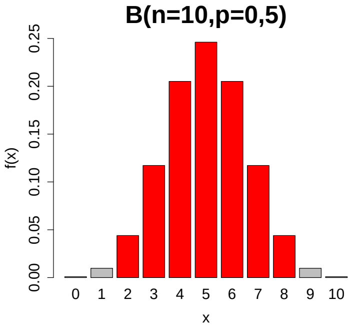
Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117 0.117	0.205 0.205	0.246 0.246	0.205 0.205	0.117 0.117	0.044	0.010	0.001

$$\sum = 0.890$$

Binomialverteilung



x	0	1	2	3	4	5	6	7	8	9	10
f(x)	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001
			0.044	0.117	0.205	0.246	0.205	0.117	0.044		

$$\sum = 0.978$$

Münzhypothese

- Falls die Münze fair ist, so liegt die Anzahl "Zahl" bei $n = 10$ Würfen mit einer Wahrscheinlichkeit von mindestens 95% im Bereich

$$\{2, 3, 4, 5, 6, 7, 8\}$$

- Falls die Anzahl "Zahl" im Bereich $\{0, 1, 9, 10\}$ liegt, kann dies zwei Ursachen haben.
  1. Ein sehr unwahrscheinliches Ereignis ist eingetreten.
  2. Unsere Annahme, dass die Münze fair ist, stimmt nicht.

Entscheidungsregel, statistischer Test

Falls die Anzahl "Zahl" im Bereich $\{0, 1, 9, 10\}$ liegt, verwerfen wir die Vermutung, dass die Münze fair ist und gehen davon aus, dass die Münze nicht fair ist.

Statistischer Test: Hypothese

Statistischer Test

Untersuchung, ob man eine Hypothese über die Grundgesamtheit mit Hilfe einer Stichprobe widerlegen kann.

- Nullhypothese $H_0 =$ Hypothese, die widerlegt werden soll.
  Beispiel: Die Münze ist fair

$$H_0 \quad : \quad \pi = 0.5$$

- Gegenhypothese $H_1 =$ Alternative zur Nullhypothese.
  Beispiel: Die Münze ist nicht fair

$$H_1 \quad : \quad \pi \neq 0.5$$

Statistischer Test: Prüfgröße, Teststatistik

- Eine Prüfgröße (Teststatistik) T ist eine zufällige Größe,
  1. anhand der wir entscheiden, ob die Nullhypothese H_0 plausibel ist.
  2. deren Verteilung wir kennen, falls die Nullhypothese H_0 zutrifft.
- Beispiel: Anzahl "Zahl" bei $n = 10$ Würfen. Unter H_0 gilt:

$$T = \sum_{i=1}^{10} X_i \quad \sim B(n = 10, \pi = \mathbf{0.5})$$

Statistischer Test: Annahme- und Ablehnbereich

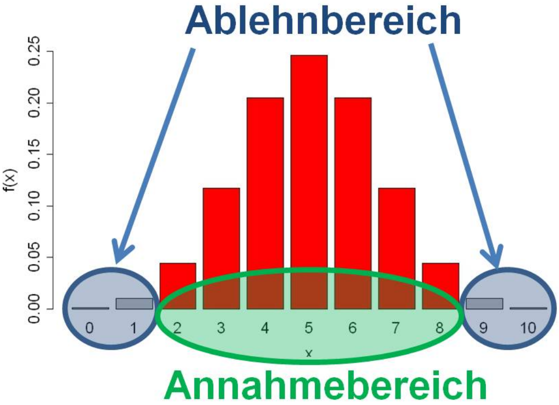
- Der Annahmebereich des Tests ist der Bereich, in dem die Prüfgröße T mit einer hohen Wahrscheinlichkeit (mindestens $1 - \alpha$) liegt.
  Beispiel: $\alpha = 0.05$ und

$$\text{Annahmebereich} = \{2, 3, 4, 5, 6, 7, 8\}$$

- α heißt das Signifikanzniveau des Tests.
- Der Ablehnbereich (kritische Bereich) ist der Bereich, in dem die Prüfgröße T mit einer kleinen Wahrscheinlichkeit (höchstens α) liegt.
  Beispiel: $\alpha = 0.05$ und

$$\text{Ablehnbereich} = \{0, 1, 9, 10\}$$

Beispiel Annahme- und Ablehnbereich



Statistischer Test: Experiment und Entscheidung

Wir ziehen eine Stichprobe und berechnen den Wert der Teststatistik T.

1. Fall: Der Wert der Teststatistik liegt im Annahmebereich.
$\longrightarrow$ Wir behalten die Nullhypothese H_0 bei.

2. Fall: Der Wert der Teststatistik liegt im Ablehnbereich.
$\longrightarrow$ Wir lehnen die Nullhypothese H_0 zugunsten der Gegenhypothese H_1 ab.

Festlegung des Signifikanzniveaus α

Beim Testen sind folgende Entscheidungen möglich:

H_0: ablehnen oder H_0: beibehalten

Damit sind zwei verschiedene Arten von Fehlern möglich:

Wahrheit Aktion	H_0 beibehalten	H_0 ablehnen
H_0 wahr	$\surd$	Fehler 1.Art
H_0 falsch	Fehler 2. Art	$\surd$

Man kann nicht beide Fehlerwahrscheinlichkeiten gleichzeitig kontrollieren! (Tradeoff!)
$\Rightarrow$ asymmetrische Vorgehensweise:

Der Fehler 1. Art wird kontrolliert durch die Angabe einer Oberschranke α („Signifikanzniveau“)

Signifikanzniveau

Übliche Werte für den Fehler erster Art sind:

$$\alpha = 0.1, \quad \alpha = 0.05, \quad \alpha = 0.01 \quad \alpha = 0.001$$

- Implizit wird also der Fehler 1. Art als schwerwiegender betrachtet.
- „konservative Perspektive“: Nullhypothese erst ablehnen, wenn wirklich nicht mehr mit den Daten verträglich.
- z.B. in der Medizin: H_0: keine Wirkung.
  $\Rightarrow$ Nur wenn die Wirkung des Medikaments überzeugend ist, soll es zugelassen werden.

Fehler 1. Art (α-Fehler):

- Die Nullhypothese wird abgelehnt, obwohl sie in Wirklichkeit richtig ist. Z.B.: Man behauptet, es bestünde ein Zusammenhang, obwohl in Wirklichkeit kein Zusammenhang besteht.

- Der Fehler 1. Art soll klein sein (üblich sind 5% oder 10%). Allerdings kann man nicht fordern, dass der Fehler 1. Art bei 0% liegen soll, sonst würde man die Nullhypothese nie ablehnen können.

  $\Rightarrow$ Fehler 2. Art

Fehler 2. Art (β-Fehler):

- Die Nullhypothese wird beibehalten, obwohl sie in Wirklichkeit falsch ist.

- Ein guter statistischer Test garantiert bei einem vergegebenen niedrigen Signifikanzniveau (als Schranke für den Fehler 1. Art) auch einen möglichst geringen Fehler 2. Art.

Folgerungen

- Die Nullhypothese wird höchstens mit Wahrscheinlichkeit α fälschlicherweise verworfen.
- Die Wahrscheinlichkeit für den Fehler 2. Art können wir nicht kontrollieren.

$\downarrow$

Ungleichbehandlung beider Fehlerarten
$\rightarrow$ Grund für Formulierung eigentlicher Forschungsfrage
als statistische Alternative:
Entscheidung für H_1 durch α statistisch abgesichert!

Veranschaulichung

- Ein Angeklagter steht vor Gericht.
- Hypothesen
  H_0: „Angeklagter ist unschuldig“
  und
  H_1: „Angeklagter ist schuldig“
- Urteil: schuldig/nicht schuldig
- H_0 und H_1 sind so formuliert, da das Gericht die Schuld des Angeklagten beweisen muss, und nicht der Angeklagte seine Unschuld.

$$\downarrow$$

- Fehler 1. Art: Unschuldiger wird verurteilt
- Fehler 2. Art: Schuldiger wird nicht verurteilt

p-Wert

p-Wert

Der p-Wert ist die Wahrscheinlichkeit, dass die Testgröße

- den beobachteten Wert oder einen noch extremeren Wert („weiter weg von H_0") annimmt
- unter der Bedingung, dass H_0 wahr ist.

Bemerkungen

1. Für die Berechnung der p-Werte benötigt man eine Statistik-Software oder Tabellen.
2. Viele Statistik-Programme geben als Ergebnis eines statistischen Tests nur den p-Wert aus.

p-Wert Bestimmung: Zweiseitiger Test

- $P_{H_0}(10 \text{ Zahl}) + P_{H_0}(0 \text{ Zahl}) = 0.002$
  10 Zahl $\Rightarrow$ p–Wert 0.002
- $P_{H_0}(9 \text{ Zahl}) = 0.01$
  $\longrightarrow P_{H_0}(\text{mindestens 9 Zahl oder höchstens 1 Zahl })$
  $= 0.001 + 0.01 + 0.01 + 0.001 = 0.022$
  9 Zahl $\Rightarrow$ p–Wert $= 0.022$
- $P_{H_0}(8 \text{ Zahl}) = 0.044$
  $\longrightarrow P_{H_0}(\text{mindestens 8 Zahl oder höchstens 2 Zahl})$
  $= 2 * (0.001 + 0.01 + 0.044) = 0.110$
  8 Zahl $\Rightarrow$ p–Wert $= 0.110$
- $P_{H_0}(7 \text{ Zahl}) = 0.117$
  $\longrightarrow P_{H_0}(\text{mehr als 7 Zahl oder höchstens 3 Zahl})$
  $= 2 * (0.001 + 0.01 + 0.044 + 0.117) = 0.344$
  7 Zahl $\Rightarrow$ p–Wert $=0.344$

Testentscheidung durch p-Wert

p-Wert und Signifikanzniveau

Die Nullhypothese wird genau dann abgelehnt, wenn der p-Wert kleiner oder gleich α ist.

Das ermöglicht ein direktes Ablesen der Testentscheidung aus entsprechenden Computerprogrammen. Daher wird der p-Wert meist zu den Test angegeben.
Illustration mit R
Je kleiner der p–Wert desto weniger passen die Daten zur Nullhypothese

p-Wert: Interpretation

- Wahrscheinlichkeit betrifft das Auftreten der Daten und nicht die Wahrscheinlichkeit von H_0
- p-Wert ist **kein Maß** für die Stärke des Effekts. Daher sollten Begriffe wie "hochsignifikant" eher vermieden werden.
- Angabe des p-Wertes immer mit Schätzung des Effekts und Konfidenzintervall
- Bei kleinen p-Werten sollte nur $p < 0.001$ o.ä. angegeben werden.

Testtheorie

Motivation

Die Prüfung einer statistischen Hypothese H_0 erfolgt mit statistischen Tests.

Ausgangspunkt ist die *Beobachtung* einer Zufallsvariablen in einer *zufälligen Stichprobe* oder einem *Experiment* .

Mittels der daraus gewonnenen *Schätzungen* der unbekannten Parameter will man zu einer *Aussage* über die Glaubwürdigkeit der Hypothese H_0 gelangen.

Testtheorie

Definition Hypothesenraum

Der statistische Test stellt eine Methode dar, Verteilungsannahmen über eine Zufallsvariable X anhand einer konkreten Stichprobe zu überprüfen.

Die Menge aller für die Zufallsvariable X in Frage kommenden Verteilungen wird als *Hypothesenraum* Ω bezeichnet. Diese Menge ist *vor* der Durchführung eines Test festzulegen.

Definition parametrisches Testproblem

Betrachtet man einen Hypothesenraum Ω, der nur Verteilungen *einer* Familie (z.B. Normalverteilung) enthält, so ist die Festlegung von Ω äquivalent zur Festlegung des Parameterraums Θ, der alle möglichen Werte eines Verteilungsparameters θ enthält. In diesem Fall spricht man von einem *parametrischen Testproblem*.

Testtheorie

Definition Nullhypothese und Alternative

Bei einem parametrischen Testproblem wird der Hypothesenraum (Parameterraum) in zwei Teilmengen aufgeteilt:

Nullhypothese die zu testende Hypothese, die durch den Test widerlegt werden soll: $H_0 = \{\theta | \theta \in \Theta_0\}$

Alternative diejenige Hypothese, die durch den Test gezeigt werden soll: $H_1 = \{\theta | \theta \in \Theta_1\}$

Dabei gilt stets: $\Theta_0 \cap \Theta_1 = \emptyset$

Definition Signifikanztest

Ein Test heißt *Signifikanztest*, wenn die Nullhypothese direkt an die Alternative „grenzt“, d.h., wenn die minimale Distanz zwischen beiden Hypothesen gleich Null ist (z.B. $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$).

Testtheorie

Definition Testgröße

Die Funktion $T(\mathbf{X}) = T(X_1, ..., X_n)$ der Stichprobenvariblen $\mathbf{X} = (X_1, ..., X_n)$ heißt *Testgröße* oder *Prüfgröße*.

Für die konkrete Stichprobe $(x_1, ..., x_n)$ ergibt sich $t = T(x_1, ..., x_n)$ als Realisation der Zufallsgröße $T(\mathbf{X})$.

Definition kritischer Bereich und Annahmebereich

Der Wertebereich der Zufallsgröße $T(X_1, ..., X_n)$ wird in zwei Teilbereiche zerlegt:

kritischer Bereich K H_0 wird abgelehnt, falls

$$t = T(x_1, ..., x_n) \in K$$

Annahmebereich $\bar{K}$ H_0 wird beibehalten, falls

$$t = T(x_1, ..., x_n) \notin K$$

Testtheorie

Definition Fehler 1. und 2. Art

Bei der Durchführung eines statistischen Tests können folgende vier Situationen auftreten:

	H_0 ist richtig	H_0 ist falsch
H_0 wird beibehalten	richtige Entscheidung	*Fehler 2. Art*
H_0 wird abgelehnt	*Fehler 1. Art*	richtige Entscheidung

Definition Signifikanzniveau und Niveau-α-Test

Bei der Konstruktion eines Tests gibt man sich für die Wahrscheinlichkeit des Fehlers 1. Art eine Schranke α vor (z.B. $\alpha = 0,05$), die nicht überschritten werden darf.
Diese Schranke bezeichnet man als *Signifikanzniveau* des Tests.
Der zugehörige Test heißt dann *Signifikanztest zum Niveau* α oder kurz *Niveau-α-Test*.

Testtheorie

ein- und zweiseitige Tests

Fall	Null-hypothese	Alternativ-hypothese	Testproblem
(a)	$\theta = \theta_0$	$\theta \neq \theta_0$	zweiseitig
(b)	$\theta \geq \theta_0$	$\theta < \theta_0$	einseitig
(c)	$\theta \leq \theta_0$	$\theta > \theta_0$	einseitig

Testtheorie

allgemeines Vorgehen bei Tests

1. Verteilungsannahme über die Zufallsvariable X
2. Formulierung von H_0 und H_1
3. Vorgabe der Irrtumswahrscheinlichkeit α
4. Konstruktion bzw. Wahl einer geeigneten Testgröße $T(X) = T(X_1, ..., X_n)$ als Funktion der Stichprobenvariablen X
5. Wahl eines kritischen Bereichs K mit $P_\theta(T(X) \in K) \leq \alpha$ für alle $\theta \in \Theta_0$
6. Berechnung der Realisierung $t = T(X_1, ..., X_n)$ der Testgröße anhand der konkreten Stichprobe $(x_1, ..., x_n)$
7. Entscheidungsregel:
    - $t \in K$: H_0 ablehnen, damit H_1 signifikant
    - $t \notin K$: H_0 beibehalten

Testentscheidung mit p-Werten

Beim Einsatz von Statistiksoftware zum Prüfen von Hypothesen werden unsere üblichen Schritte – insbesondere der kritische Wert – nicht angezeigt. Statt dessen wird der konkrete Wert der Teststatistik und der zugehörige **p-value** oder die sog. **Signifikanz** ausgegeben.

Die Testentscheidung lautet dann: H_0 ablehnen, falls der p-value kleiner oder gleich dem vorgegebenem Signifikanzniveau α ist, ansonsten H_0 nicht ablehnen.

Zweiseitiger approximativer Test auf den Anteilswert

- X Bernoulli-Variable mit $\pi = P(X = 1)$.
- Zweiseitige Hypothese über den Anteilswert p

$$\begin{aligned} H_0 &: \quad \pi = \pi_0 \\ H_1 &: \quad \pi \neq \pi_0 \end{aligned}$$

- Testgröße: Anteil in der Stichprobe $X_1, \ldots, X_n$

$$\hat{\pi} \quad = \quad \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Stichprobenumfang n ist genügend groß

Zweiseitiger approximativer Test auf den Anteilswert

Hypothesen: $H_0 : \pi = \pi_0$ versus $H_1 : \pi \neq \pi_0$

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$\pi_0 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

H_0 wird abgelehnt, falls

$$\hat{\pi} < \pi_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

oder

$$\hat{\pi} > \pi_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

Beispiel: Münzwurf bei Stern TV 2002

- Nullhypothese: $\pi = \pi_0 = 0.5$ („Münze ist fair.")
- Signifikanzniveau: $\alpha = 0.01$
- $n = 800$ Münzwürfe
  $\rightarrow$ Normalverteilung
- Annahmebereich

$$\begin{aligned} \pi_0 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}} &= 0.5 \pm z_{1-\frac{0.01}{2}} \cdot \sqrt{\frac{0.5(1-0.5)}{800}} \\ &= 0.5 \pm 0.054 \end{aligned}$$

- H_0 wird beibehalten, falls: $\hat{\pi} \in [0.454; 0.546]$

Wert bei Stern TV (2002) : 501/800

Dualität Test und Konfidenzintervall

- Annahmebereich: Wir behalten H_0 bei, falls die Testgröße $\hat{\pi}$ in der Nähe von π_0 liegt:
- Äquivalente Formulierung über ein Konfidenzintervall: Wir behalten H_0 bei, falls π_0 in der Nähe der Testgröße liegt
- Wir behalten H_0 bei, falls π_0 im Konfidenzintervall für den Anteil liegt
- Dabei hängen das Konfindenzniveau γ und das Signifikanzniveau α wie folgt zusammen:
  $1 - \alpha = \gamma$
- Dies gilt sehr allgemein für zweiseitige Test und Konfidenzintervalle
- Dies Prinzip kann zur Konstruktion von Konfidenzintervallen verwendet werden

Einseitiger Test auf den Anteilswert

- X Bernoulli-Variable mit $\pi = P(X = 1)$.
- Einseitige Hypothese über den Anteilswert π

$$\begin{aligned} H_0 &: \quad \pi \leq \pi_0 \\ H_1 &: \quad \pi > \pi_0 \end{aligned}$$

- Testgröße: Anteil in der Stichprobe $X_1, \ldots, X_n$

$$\hat{\pi} \quad = \quad \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Stichprobenumfang n ist genügend groß
(Faustregel: $n\pi_0(1 - \pi_0) > 9$)

Einseitiger Test auf den Anteilswert

Hypothesen: $H_0 : \pi \leq \pi_0$ vs. $H_1 : \pi > \pi_0$

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$\hat{\pi} \quad \leq \quad \pi_0 + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

H_0 wird abgelehnt, falls

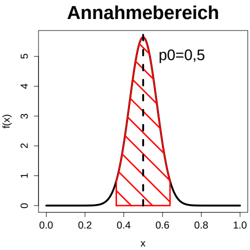
$$\hat{\pi} \quad > \quad \pi_0 + z_{1-\alpha} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

$z_{1-\alpha}$ ist das $(1-\alpha)$-Quantil der Standardnormalverteilung.

Vergleich einseitige Tests und zweiseitiger Test

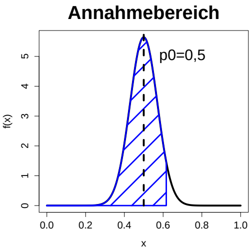
Test auf Anteil mit einer Stichprobe der Größe $n = 50$ und Signifikanzniveau $\alpha = 0.05$

$$H_0 : \pi = 0.5$$
$$H_1 : \pi \neq 0.5$$



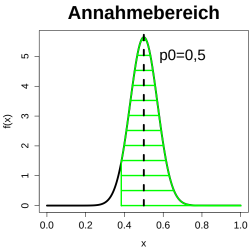
$$[0.36; 0.64]$$

$$H_0 : \pi \leq 0.5$$
$$H_1 : \pi > 0.5$$



$$[0; 0.62]$$

$$H_0 : \pi \geq 0.5$$
$$H_1 : \pi < 0.5$$



$$[0.38; 1]$$

p-Wert: Ein- und zweiseitiger Test

In vielen Fällen halbiert sich der p-Wert bei einseitiger Fragestellung.

Beispiel: Binomialtest 10 Experimente T Anzahl Treffer.
Ergebnis: $T = 9$

- zweiseitiger Test: $H_0 : \pi = 0.5$ gegen $H_1 : \pi \neq 0.5$.

  Extrem oder gleich extreme Werte von T: 9,10,1,0

  p-Wert =
  $P_{H_0}(T \in \{9, 10, 0, 1\}) = 0.01 + 0.001 + 0.01 + 0.001 = 0.022$

- einseitiger Test: $H_0 : \pi \leq 0.5$ gegen $H_1 : \pi > 0.5$

  Extrem oder gleich extreme Werte von T: 9, 10

  p-Wert $= P_{H_0}(T \in \{9, 10\}) = 0.11$

Test auf den Erwartungswert

- Wir interessieren uns für den Erwartungswert μ einer metrischen Zufallsgröße.
  Beispiele: Alter, Einkommen, Körpergröße, Scorewert ...
- Wir können einseitige oder zweiseitige Hypothesen formulieren.
- Beispiele
  - "Vor 10 Jahren betrug die Durchschnittsgröße von Studienanfängern und -anfängerinnen 167 cm. Heute sind sie im Schnitt größer als 167 cm."
  - Die Differenz zwischen Gewicht vor und nach einer Diät ist 0.

Zweiseitiger Gauss-Test auf den Erwartungswert μ

- X Zufallsgröße mit Erwartungwert μ.
- Zweiseitige Hypothese über μ:

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \end{aligned}$$

- Testgröße: Mittelwert in der Stichprobe $X_1, \ldots, X_n$:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Stichprobenumfang n ist genügend groß
  (Faustregel $n > 30$)

Zweiseitiger Gauss-Test auf den Erwartungswert μ

Hypothesen: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$\mu_0 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}$$

mit

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

H_0 wird abgelehnt, falls

$$\overline{X} < \mu_0 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} \quad \text{oder} \quad \overline{X} > \mu_0 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}}$$

$z_{1-\frac{\alpha}{2}}$ ist das $(1-\alpha/2)$-Quantil der Standardnormalverteilung.

Zweiseitiger Gauss-Test: p–Wert

Der p-Wert ergibt sich zu :

$$2 \cdot \left[1 - \Phi\left(|\overline{X} - \mu_0| / \sqrt{\frac{S^2}{n}}\right)\right]$$

Φ ist die Verteilungsfunktion der Standard Normalverteilung

$$(\overline{X} - \mu_0) / \sqrt{\frac{S^2}{n}}$$

wird als t-Wert (oder z-Wert) bezeichnet.

Beispiel: Verändert sich der Blutdruck nach einer Intervention

Beispiel zweiseitiger Test

- Nullhypothese: Die Blutdruckdifferenz ist 0.

$$\begin{aligned} H_0 &: \quad \mu = 0 \\ H_1 &: \quad \mu \neq 0 \end{aligned}$$

- Testgröße: Durchschnittlicher Bluddruck
- Faustregel $n = 62 > 30$ ist erfüllt
  $\longrightarrow$ zweiseitiger Gauß-Test

Einseitiger Gauss-Test auf den Erwartungswert μ

- X Zufallsgröße mit Erwartungwert μ.
- Einseitige Hypothese über μ:

$$\begin{aligned} H_0 &: \quad \mu \leq \mu_0 \\ H_1 &: \quad \mu > \mu_0 \end{aligned}$$

- Testgröße: Mittelwert in der Stichprobe $X_1, \ldots, X_n$:

$$\overline{X} \quad = \quad \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Stichprobenumfang n ist genügend groß
  (Faustregel $n > 30$)

Einseitiger Gauss-Test auf den Erwartungswert μ

Hypothesen: $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$\overline{X} \quad \leq \quad \mu_0 + z_{1-\alpha}\sqrt{\frac{S^2}{n}}$$

H_0 wird abgelehnt, falls

$$\overline{X} \quad > \quad \mu_0 + z_{1-\alpha}\sqrt{\frac{S^2}{n}}$$

$z_{1-\alpha}$ ist das $(1-\alpha)$-Quantil der Standardnormalverteilung.

Zusammenfassung

- Signifikanztest weiteres zentrales Instrument der statistischen Inferenz
- Konstruktion über Nullhypothese (i.d.R. Gegenteil der Forschungshypothese)
- Statistische Signifikanz entspricht Falsifizierung der Nullhypothese
- Enger Zusammenhang mit Konfidenzintervallen