Hypothese:

„Behauptung einer Tatsache, deren Überprüfung noch aussteht“ (Leutner in: Endruweit, Trommsdorff: Wörterbuch der Soziologie, 1989).
Statistischer Test: Überprüfung von Hypothesen über die Grundgesamtheit anhand einer Stichprobe
Idealtypische Vorgehensweise:
Wissenschaftlicher Fortschritt durch Falsifikation von Hypothesen

Empirischer Gehalt:

Nach dem Wissenschaftstheoretiker Karl Popper hat eine Hypothese umso mehr empirischen Gehalt, je mehr Möglichkeiten zur Falsifikation sie bietet (z.B. indem sie präziser formuliert ist). Je mehr Versuche der Falsifikation eine Hypothese besteht, umso höher ist ihr Bewährungsgrad. Wissenschaftlicher Fortschritt ergibt sich laut Popper also durch das „Aussortieren“ falscher Hypothesen.

Weiterentwicklung des Popperschen Falsifikationsprinzips v.a. durch Lagatos.

Statistische Testtheorie: Schließe von Stichprobe Experiment auf Grundgesamtheit 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

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

- Wahrscheinlichkeit für Zahl

$$p = P(X = 1)$$

- ,,Die Münze ist nicht fair'' heißt

$$p \neq 0,5$$

Überprüfung der Münzhypothese

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

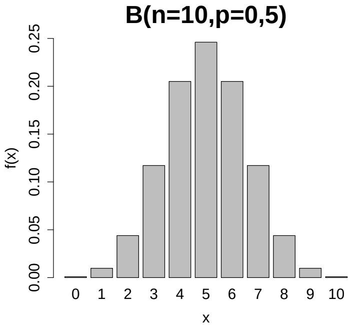
$$\sum_{i=1}^{n} X_i \quad \sim \quad B(n = 10, p)$$

- Welche Ergebnisse sind wahrscheinlich, falls die Münze fair ist?
- **Falls die Münze fair ist**, so ist die Anzahl ,,Zahl" binomialverteilt mit $p = 0,5$.

$$\sum_{i=1}^{10} X_i \quad \sim \quad B(n = 10, p = 0,5)$$

- **Falls die Münze fair ist**, so sollte $\sum_{i=1}^{10} X_i$ mit einer Wahrscheinlichkeit von 90 % nicht weit entfernt vom Erwartungswert 5 liegen.

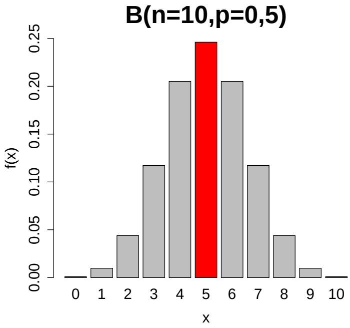
Binomialverteilung



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

$$\sum =$$

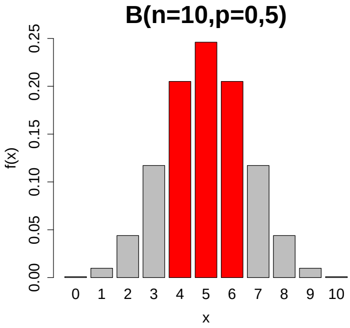
Binomialverteilung



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

$$\sum = 0.246$$

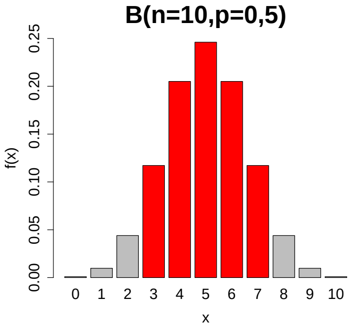
Binomialverteilung



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

$$\sum = 0.656$$

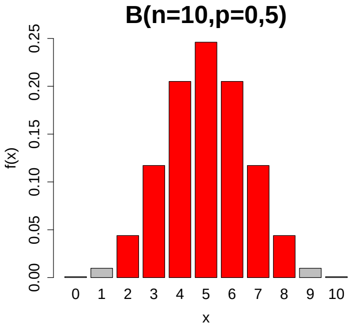
Binomialverteilung



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

$$\sum = 0.890$$

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 90% 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.

(Wir können uns natürlich irren.)

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

$$H_0 \quad : \quad p = 0,5$$

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

$$H_1 \quad : \quad p \neq 0,5$$

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:

$$T = \sum_{i=1}^{10} X_i \quad \sim B(n = 10, \mathbf{p} = \mathbf{0,5})$$

Statistischer Test: Annahme- und Ablehnbereich

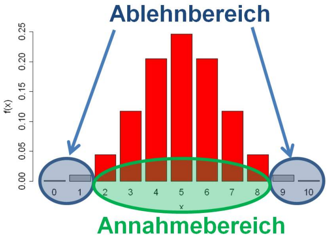
- Der Annahmebereich des Tests ist der Bereich, in dem die Prüfgröße T mit einer hohen Wahrscheinlichkeit (mindestens $1 - \alpha$) liegt.
  Beispiel: $\alpha = 0,1$ und

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

- 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: $\alpha = 0,1$ 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.

Unsere Entscheidung ist mit großer Sicherheit korrekt (zum Signifikanzniveau α).

Beispiel

Studie zur Einstellung der Münchner Bevölkerung zu psychisch Kranken (1989).
Wir betrachten eine Teilstudie: Kooperationsbereitschaft in der Befragung.

- „Theorie“: Aktive Stellung im öffentlichen Leben beeinflusst Kooperationsbereitschaft positiv.

$$\begin{array}{lll} \textit{Aktiv} & \leftrightarrow & \text{Altruismus} \\ & \leftrightarrow & \text{Interesse an öffentlichen Angelegenheiten} \\ & & \Rightarrow \text{eher bereit, die Rolle des Befragten einzunehmen} \end{array}$$

- Hypothese: „Unterscheidet sich die Koorperationsbereitschaft der aktiven Personen vom Rest der Bevölkerung?“

- Operationalisierung:

- Aktiv im öffentlichen Leben

  $\rightarrow$ Verbandsmitgliedschaft ja/nein = Variable X
- Kooperationsbereitschaft

  $\rightarrow$ antwortet freiwillig (Koorperativer)/nur auf „sanften Druck" (Primärverweigerer) = Variable Y

- Statistische Hypothesen: „Besteht ein Zusammenhang zwischen X und Y?"

  Statistisches Vorgehen:
  Kann die sog. *Nullhypothese* „Es besteht kein Zusammenhang zwischen X und Y" abgelehnt werden?

Herleitung / Motivation eines geeigneten Prüfverfahrens

Gegebene Daten (relative und absolute Häufigkeiten):

		kooperativ		
		ja	nein	
aktiv	ja	0.27 (95)	0.05 (17)	0.32 (112)
	nein.	0.53 (186)	0.15 (54)	0.68 (240)
		0.8 (281)	0.2 (71)	1 (352)

Vergleiche gegebene Tafel mit der „Unabhängigkeitstafel“
Wie würde denn die Tafel aussehen, wenn kein Zusammenhang bestünde?
Genauer: wie würde das Innere der Tabelle aussehen, wenn Unabhängigkeit (und die gleichen Randverteilungen) herrschen würde, also die Nullhypothese zutreffen würde?

		kooperativ		
		ja	nein	
aktiv	ja	0.256	0.064	0.32
	nein	0.544	0.136	0.68
		0.8	0.2	1

Die Häufigkeiten in der Unabhängigkeitstafel weichen von den tatsächlichen Daten ab. Vgl. Statistik I: Je stärker die Abweichung, desto stärker ist der Zusammenhang.

- Wie groß muss die Abweichung sein, um die Nullhypothese abzulehnen?
- Beachte: Die Daten entstammen einer Stichprobe, die mit einem Zufallsfehler behaftet ist. Selbst bei tatsächlich vorliegender Unabhängigkeit ist die Wskt., genau die Unabhängigkeitstafel zu beobachten, sehr gering.

Kardinalfrage der statistischen Testtheorie:

Weichen die tatsächlichen Daten von der bei Gültigkeit der Nullhypothese zu erwartenden Situation „überzufällig“ stark ab, d.h. so stark, dass man die Abweichung nicht mehr nur der Zufallsstreuung zuschreiben kann? Nur in diesem Fall ist die Nullhypothese abzulehnen.
Idee:

kleine Abweichung	$\Rightarrow$	nur Zufallsstreuung
große Abweichung	$\Rightarrow$	Zufallsstreung + inhaltlicher Unterschied
	$\Rightarrow$	Nullhypothese ablehen

Wann ist die Abweichung „groß genug“, d.h. überzufällig?

- Testen mit Hilfe des p-Wertes (Alternative: Testen mithilfe eines Ablehnbereichs, s.u.)
- Bestimme eine Zufallsvariable T, die in geeigneter Weise den Unterschied einer zufälligen Stichprobe zur Situation der Nullhypothese misst (hier: der χ^2-Abstand zwischen einer Stichprobe und der Unabhängigkeitstafel, vgl. Statistik I).
- Bestimme die Realisation t von T anhand der konkreten Daten (hier: χ^2=2.11).
- Berechne die Wahrscheinlichkeit, einen mindestens so extremen Wert von T zu beobachten, falls H_0 richtig ist:

$$p\text{-Wert} \quad := \quad P(T \geq t|H_0)$$

(hier: p-Wert=0.15).

Festlegung des Signifikanzniveaus α

Beim Testen sind folgende Entscheidungen möglich:

$$H_0\text{: ablehnen} \quad \text{oder} \quad H_0\text{: beibehalten}$$

Der Begriff einseitig/zweiseitig bezieht sich auf die Alternative, je nachdem ob die Alternative nur aus großen bzw. nur aus kleinen Werten besteht oder ob sowohl große als auch kleine Werte für die Alternative sprechen.
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“)

Typische Werte: üblich

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

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.

- Falls p-Wert $\leq$ einer aus substanzwissenschaftlichen Überlegungen abgeleiteten, vorgegebenen Schranke α (Signifikanznivea), dann H_0ablehnen, sonst beibehalten. (hier bei $\alpha = 0.05$ (üblicher Wert): p-Wert zu groß: Die Nullhypothese kann nicht abgelehnt werden.)

Der Richtwert α, das sogenannte *Signifikanzniveau* soll sicherstellen, dass die Nullhypothese nur in α der Fälle fälschlicherweise abgelehnt wird.

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.

Konstruktion eines parametrischen statistischen Tests

Aufstellen der substanzwissenschaftlichen Hypothese / inhaltliche Fragestellung
(z.B. Rot/Grün bekommt die absolute Mehrheit, das Einkommen von Akademikern beträgt mindestens 3000 Euro) Formulieren eines geeigneten statistischen Modells
Im Folgenden stets $X_1, \ldots, X_n$ i.i.d. Stichprobe sowie parametrisches Modell mit unbekanntem Parameter ϑ.
Anteil Rot/Grün: $B(1, \pi)$
Durchschnittseinkommen: $\mathrm{N}(\mu; \sigma^2)$.

Formulierung der statistischen Hypothesen

- Umformulieren der substantzwissenschaftlichen Hypothesen als Hypothesen über ϑ.
- Verglichen wird immer eine sog. *Nullhypothese* (H_0) mit einer sog. *Alternativhypothese* (H_1).
- Bei parametrischen Fragestellungen:
  a) Einseitige Testprobleme:

$$H_0 : \vartheta \leq \vartheta_0 \text{ gegen } H_1 : \vartheta > \vartheta_0$$

  (z.B. der Anteil von Rot/Grün ist kleiner gleich 50% oder größer)

$$H_0 : \vartheta \geq \vartheta_0 \text{ gegen } H_1 : \vartheta < \vartheta_0$$

  b) Zweiseitiges Testproblem:

$$H_0 : \vartheta = \vartheta_0 \text{ gegen } H_1 : \vartheta \neq \vartheta_0$$

  („Das Durchschnittseinkommen ist 3000 Euro„, Abweichungen nach unten oder oben möglich)

  ϑ_0 ist ein fester, vorgegebener Wert, der von inhaltlichem Interesse ist; zu unterscheiden von wahrem Wert ϑ.

Festlegen einer Testgröße und einer kritischen Region

Eine *Testgröße* T ist eine Zufallsgröße $T = g(X_1, \ldots, X_n)$, die „empfindlich gegenüber Abweichungen von H_0 ist". Die *Kritische Region* KR („Ablehnungsbereich") besteht aus potentiellen Werten von T, die gegen H_0 sprechen.
Ähnlich wie oben: Werte, die unter H_0 sehr unwahrscheinlich sind, sprechen gegen H_0.
Fällt die Beobachtung für T in KR, wird man sich gegen H_0 entscheiden.
Damit der Fehler 1. Art durch α beschränkt bleibt muss die kritische Region KR also so gewählt werden, dass

$$P(T \in KR | H_0) \leq \alpha$$

gilt, d.h. die Wahrscheinlichkeit, dass T in der kritischen Region liegt und damit zur Ablehnung von H_0 führt darf höchstens α sein, wenn H_0 stimmt.

Festlegen einer Testgröße und einer kritischen Region

Umgekehrt soll $P(T \in KR|H_1)$ möglichst groß sein, da dies die Wahrscheinlichkeit ist, die Nullhypothese H_0 abzulehnen, falls sie falsch ist. (Gegenwahrscheinlichkeit zur Wahrscheinlichkeit für den Fehler 2. Art, auch als *Power* oder *Güte* des Tests bezeichnet.)

Auswerten der Stichprobe

Berechnung der Realisation t der Testgröße T basierend auf der konkret vorliegenden Stichprobe.

Testentscheidung

Ist $t \in KR$, dann H_0 ablehnen, sonst nicht ablehnen.

Zweiseitiger approximativer Test auf den Anteilswert

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

$$\begin{aligned} H_0 &: \quad p = p_0 \\ H_1 &: \quad p \neq p_0 \end{aligned}$$

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

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

- Stichprobenumfang n ist genügend groß

Zweiseitiger approximativer Test auf den Anteilswert

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$p_0 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

H_0 wird abgelehnt, falls

$$R < p_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

oder

$$R > p_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

Fehler 1. Art und 2. Art

Zwei komplementäre Hypothesen über die Grundgesamtheit

1. Nullhypothese H_0 („Die Münze ist fair.")
2. Gegenhypothese H_1 („Die Münze ist nicht fair.")

Zwei komplementäre Testergebnisse

1. Nullhypothese H_0 wird abgelehnt.
2. Nullhypothese H_0 wird nicht abgelehnt.

Mögliche Ausgänge

	Hypothese	
	H_0 wahr	H_0 nicht wahr
Test lehnt H_0 ab	Fehler 1. Art (α-Fehler)	richtig
Test lehnt H_0 nicht ab	richtig	Fehler 2. Art (β-Fehler)

Fehleranalyse

- Mögliche Fehler
  1. Fehler 1. Art: Wir lehnen H_0 ab, obwohl H_0 wahr ist.
  2. Fehler 2. Art: Wir lehnen H_0 nicht ab, obwohl H_0 falsch ist.
- Wir können Fehlentscheidungen nicht ausschließen, aber es ist möglich, die Wahrscheinlichkeit von Fehlentscheidungen zu kontrollieren.
- **Statistische Tests kontrollieren den Fehler 1. Art**: Der Fehler 1. Art tritt mit einer Wahrscheinlichkeit von höchstens α auf.

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.

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

Zusammenfassung statistischer Test

Können wir Hypothesen (=Vermutungen) über die Grundgesamtheit anhand von Daten widerlegen?

- Ein **statistischer Test** führt zu eine Regel, anhand der wir entscheiden, ob wir eine Hypothese verwerfen oder beibehalten.
- Unsere Entscheidung kann falsch sein
  1. Fehler 1. Art: Die Nullhypothese wird fälschlicherweise abgelehnt.
  2. Fehler 2. Art: Die Nullhypothese wird fälschlicherweise beibehalten.

  Die Wahrscheinlichkeit des Fehlers 1. Art ist klein (höchstens α).
- Statistische Tests sind unsymmetrisch:
  1. Die Nullhypothese wird abgelehnt. $\rightarrow$ starke Evidenz für die Gegenhypothese.
  2. Die Nullhypothese wird nicht abgelehnt. $\rightarrow$ Die Daten sprechen nicht gegen die Nullhypothese.

Wiederholung: Prüfgröße, Teststatistik

Prüfgröße

Eine Prüfgröße (Teststatistik) ist eine zufällige Größe,

1. anhand der wir entscheiden, ob die Nullhypothese H_0 plausibel ist.
2. deren Verteilung wir (approximativ) kennen, falls die Nullhypothese H_0 zutrifft.

Beispiel Anzahl "Zahl" bei n Würfen. Unter H_0 gilt:

$$T = \sum_{i=1}^{n} X_i \quad \sim B(n, \mathbf{p} = \mathbf{0,5})$$

Falls $n > 30$, so gilt approximativ für den Anteil „Zahl" unter H_0

$$R = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N\left(\mathbf{0,5}; \frac{\mathbf{0,5} \cdot (1 - \mathbf{0,5})}{n}\right)$$

Wiederholung: Annahme- und Ablehnbereich

- Der Annahmebereich des Tests ist der Bereich, in dem die Prüfgröße mit einer hohen Wahrscheinlichkeit (mindestens $1 - \alpha$) liegt.
- α heißt das Signifikanzniveau des Tests.
- Der Ablehnbereich (kritische Bereich) ist der Bereich, in dem die Prüfgröße mit einer kleinen Wahrscheinlichkeit (höchstens α) liegt.

Wiederholung: Experiment und Entscheidung

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

Unsere Entscheidung ist mit großer Sicherheit korrekt (zum Signifikanzniveau α).

Wiederholung: Zweiseitiger approximativer Test auf den Anteilswert

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

$$\begin{aligned} H_0 &: \quad p = p_0 \\ H_1 &: \quad p \neq p_0 \end{aligned}$$

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

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

- Stichprobenumfang n ist genügend groß
(Faustregel: $np_0(1 - p_0) > 9$)

Wiederholung: Zweiseitiger approximativer Test auf den Anteilswert

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$p_0 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

H_0 wird abgelehnt, falls

$$R < p_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

oder

$$R > p_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

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

Beispiel: Münzwurf

- Nullhypothese: $p = p_0 = 0,5$ („Münze ist fair.")
- Signifikanzniveau: $\alpha = 0,05$
- $n = 50$ Münzwürfe
- Faustregel gültig?
  $50 \cdot 0,5 \cdot (1 - 0,5) = 50 \cdot 0,25 = 12,5 > 9 \rightarrow$ Normalverteilung
- Annahmebereich

$$
\begin{aligned}
p_0 \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} &= 0,5 \pm z_{1-\frac{0,05}{2}} \cdot \sqrt{\frac{0,5(1-0,5)}{50}} \\
&= 0,5 \pm z_{0,975} \cdot \sqrt{0,005} \\
&\approx 0,5 \pm 1,96 \cdot 0,07 \\
&= 0,5 \pm 0,14
\end{aligned}
$$

- H_0 wird beibehalten, falls

$$R \in [0,36; 0,64]$$

Einseitiger Test auf den Anteilswert

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

$$\begin{aligned} H_0 &: p \leq p_0 \\ H_1 &: p > p_0 \end{aligned}$$

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

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

- Stichprobenumfang n ist genügend groß
  (Faustregel: $np_0(1 - p_0) > 9$)

Einseitiger Test auf den Anteilswert

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$R \quad \leq \quad p_0 + z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

H_0 wird abgelehnt, falls

$$R \quad > \quad p_0 + z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

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

Einseitiger Test auf den Anteilswert

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

$$\begin{aligned} H_0 &: \quad p \geq p_0 \\ H_1 &: \quad p < p_0 \end{aligned}$$

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

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

- Stichprobenumfang n ist genügend groß
(Faustregel: $np_0(1 - p_0) > 9$)

Einseitiger Test auf den Anteilswert

Testentscheidung zum Signifikanzniveau α

Annahmebereich

$$R \quad \geq \quad p_0 - z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

H_0 wird abgelehnt, falls

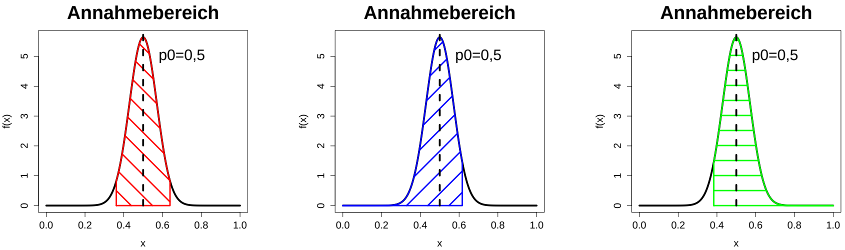
$$R \quad < \quad p_0 - z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

$z_{1-\alpha}$ 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 $\alpha = 0,05$

$H_0: p = 0,5$ $H_1: p \neq 0,5$	$H_0: p \leq 0,5$ $H_1: p > 0,5$	$H_0: p \geq 0,5$ $H_1: p < 0,5$



$[0,36; 0,64]$	$[0; 0,62]$	$[0,38; 1]$

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 (**nicht** Thema dieser Veranstaltung).
2. Viele Statistik-Programme geben als Ergebnis eines statistischen Tests nur den p-Wert aus.

p-Wert und Signifikanzniveau

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

Test auf den Erwartungswert

- Wir interessieren uns für den Erwartungswert μ einer metrischen Zufallsgröße.
  Beispiele: Alter, Einkommen, Körpergröße, ...
- Wir können einseitige oder zweiseitige Hypothesen formulieren.
- Beispiele
  - "Hörer und Hörerinnen der KW-Vorlesung sind im Schnitt 20 Jahre alt."
  - "Vor 10 Jahren betrug die Durchschnittsgröße von Studienanfängern und -anfängerinnen 167 cm. Heute sind sie im Schnitt größer als 167 cm."
  - ..

Zweiseitiger Gauss-Test auf den Erwartungswert μ

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

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

- 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 $n > 30$)

Zweiseitiger Gauss-Test auf den Erwartungswert μ

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.

Beispiel zweiseitiger Gauss-Test

- Nullhypothese: Das Durchschnittsalter der Hörer und Hörerinnen dieser Veranstaltung beträgt 20 Jahre.

$$\begin{aligned} H_0 &: \quad \mu = 20 \\ H_1 &: \quad \mu \neq 20 \end{aligned}$$

- Testgröße: Durchschnittsalter in der Donnerstagsübung (n=48 befragte Personen)
- Faustregel $n = 48 > 30$ ist erfüllt.

Beispiel zweiseitiger Gauss-Test

- Signifikanzniveau $\alpha = 0,1$
- Stichprobenvarianz in der Donnerstagsübung

$$S^2 \approx 4,54$$

- Annahmebereich

$$\begin{aligned}\mu_0 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{S^2}{n}} &= 20 \pm z_{1-0,05} \sqrt{\frac{4,54}{48}} \\ &\approx 20 \pm 1.64 \sqrt{0,09} \\ &\approx 20 \pm 0.49\end{aligned}$$

- Durchschnittsalter in der Donnerstagsübung

$$\overline{X} = 21,63$$

- H_0 wird verworfen.

Test auf den Erwartungswert mit SPSS

Beispiel zweiseitiger Test

- Nullhypothese: Das Durchschnittsalter der Hörer und Hörerinnen dieser Veranstaltung beträgt 20 Jahre.

$$H_0 \quad : \quad \mu = 20$$
$$H_1 \quad : \quad \mu \neq 20$$

- Testgröße: Durchschnittsalter in der Donnerstagsübung (n=48 befragte Personen)
- Faustregel $n = 48 > 30$ ist erfüllt
  $\longrightarrow$ zweiseitiger Gauß-Test
- Falls wir annehmen, dass das Alter normalverteilt ist, können wir auch den zweiseitigen t-Test anwenden.
  (Standardeinstellung in SPSS)

Nachtrag: Test auf den Erwartungswert mit SPSS

- Signifikanzniveau $\alpha = 0,1$
- Stichprobenvarianz in der Donnerstagsübung $S^2 \approx 4,54$
- Quantile

$$z_{0,95} = 1,64 \text{ (Gauss -Test)} \qquad t_{47;0,95} = 1,68 \text{ (t-Test)}$$

- Annahmebereich

$$[19,51; 20,49] \text{ (Gauss -Test)} \qquad [19,50; 20,50] \text{ (t-Test)}$$

- Durchschnittsalter in der Donnerstagsübung

$$\overline{X} \quad = \quad 21,63$$

$\longrightarrow H_0$ wird sowohl vom Gauss -Test als auch vom t-Test verworfen.

Dualität Annahmebereich und Konfidenzintervall

- Annahmebereich: Wir behalten H_0 bei, falls die Testgröße T in der Nähe von μ_0 liegt:

$$T = 21,63 \quad \overset{?}{\in} \quad [19,5; 20; 5]$$

- Äquivalente Formulierung über ein Konfidenzintervall: Wir behalten H_0 bei, falls μ_0 in der Nähe der Testgröße liegt

$$\mu_0 = 20 \quad \overset{?}{\in} \quad [21,13; 22; 13]$$

- SPSS: Wir behalten H_0 bei, falls 0 im Konfidenzintervall für die Differenz liegt

$$0 \quad \overset{?}{\in} \quad [1,63 - 0,5; 1,63 + 0,5] = [1,13; 2,13]$$

Ausgabe in SPSS

T-Test

[DatenSet1] C:\Dokumente und Einstellungen\kraemer\Eigene Dateien\Datensatz_Donnerstag.sav

Statistik bei einer Stichprobe

	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Alter	48	21,625	2,1301	,3075

Test bei einer Sichprobe

	Testwert = 20					
					90% Konfidenzintervall der Differenz	
	T	df	Sig. (2-seitig)	Mittlere Differenz	Untere	Obere
Alter	5,285	47	,000	1,6250	1,109	2,141

Vergleich:

$$0 \overset{?}{\in} [1,63 - 0,5; 1,63 + 0,5] = [1,13; 2,13]$$

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

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.

Einseitiger Gauss-Test auf den Erwartungswert μ

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

$$\begin{aligned} H_0 &: \quad \mu \geq \mu_0 \\ H_1 &: \quad \mu < \mu_0 \end{aligned}$$

- 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

Testentscheidung zum Signifikanzniveau α

Annahmebereich

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

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.

Ein- bzw. zweiseitiger t-Test auf den Erwartungswert

Was tun wir, falls die Faustregel $n > 30$ nicht erfüllt ist?

Zusätzliche Voraussetzung

- Zufallsgröße X ist normalverteilt.

Wir stellen keine Bedinung an die Stichprobengröße n.

t-Test

- Der einseitige bzw. der zweiseitige t-Test auf den Erwartungswert μ hat die gleiche Form wie der einseitige bzw. zweiseitige Gauss-Test.
- Der t-Test unterscheidet sich vom Gauss-Test dadurch, dass wir das Quantil z der Standardnormalverteilung durch das Quantil t der t-Verteilung mit $n-1$ Freiheitsgraden ersetzen.

Verbundene und unverbundene Stichprobe

Verbundene, abhängige Stichprobe

Zwei Stichproben heißen verbunden, falls an einem Merkmalsträger (z.B. einer Person) zwei vergleichbare Merkmale erhoben werden.
Man nennt verbundene Stichproben oft auch abhängige Stichproben.

Beispiel: Das Ziel einer medizinischen Studie ist es, die Wirkung eines cholesterin-senkenden Medikaments zu überprüfen.

- Unterteilung der Probanden und Probandinnen in 2 Gruppen: 1 Gruppe erhält das Medikament, 1 Gruppe erhält ein Placebo.
  unverbundene Stichprobe
- Alle Probanden und Probandinnen erhalten das Medikament. Von allen Personen wird der Cholesterinspiegel am Anfang und am Ende der Studie erhoben.
  verbundene Stichprobe

Approximativer Test auf Erwartungswert-Differenz bei unabhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Größen mit Erwartungswerten μ_X und μ_Y
- $X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ **unabhängige** Stichproben
- Testgröße: Differenz der Mittelwerte

$$\begin{aligned} T &= \overline{X} - \overline{Y} \\ &= \frac{1}{m}\sum_{i=1}^{m} X_i - \frac{1}{n}\sum_{i=1}^{n} Y_i \end{aligned}$$

- Faustregel: Stichprobenumfänge $m, n > 30$

Differenz von Erwartungswerten bei unabhängigen Stichproben

Annahmebereich

Für die beiden einseitigen Tests und den zweiseitigen Test auf die Differenz

$$
\begin{array}{lll}
H_0 : \mu_X - \mu_Y = d_0 & H_0 : \mu_X - \mu_Y \leq d_0 & H_0 : \mu_X - \mu_Y \geq d_0 \\
H_1 : \mu_X - \mu_Y \neq d_0 & H_1 : \mu_X - \mu_Y > d_0 & H_1 : \mu_X - \mu_Y < d_0
\end{array}
$$

ist der Annahmebereich

$$
d_0 \pm z_{1-\frac{\alpha}{2}} \cdot s \qquad]-\infty, d_0 + z_{1-\alpha} \cdot s] \qquad [d_0 - z_{1-\alpha} \cdot s, \infty[
$$

mit

$$
s = \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}
$$

Beispiel: Radio-Hördauer Ost-West

- Hören Personen in den alten Bundesländern im Schnitt mehr Radio?
  X : Hördauer im den alten Bundesländern, Y: Radiodauer in den neuen Bundesländern

$$
\begin{aligned}
H_0 &: \mu_X - \mu_Y \leq 0 \\
H_1 &: \mu_X - \mu_Y > 0
\end{aligned}
$$

- Befragung unter 253 Personen aus den alten Bundesländern und 932 Personen aus den neuen Bundesländern
  - unverbundene Stichproben $X_1, \ldots, X_{253}$ und $Y_1, \ldots, Y_{932}$
  - Stichprobengrößen $m = 253, n = 932 > 30$
- Durchschnittliche Hördauer:
  $11,4$ h (Standardabweichung $8,4$ h) in den alten Bundesländern
  $9,5$ h (Standardabweichung $8,4$ h) in den neuen Bundesländern

Beispiel: Radio-Hördauer Ost-West

- Signifikanzniveau: $\alpha = 0,1$
- Differenz der Radio-Hördauer

$$\overline{X} - \overline{Y} \quad = \quad 11,4 - 9,5 = 1,9$$

- Annahmebereich

$$\begin{aligned} \overline{X} - \overline{Y} &\leq z_{1-\alpha} \cdot \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}} \\ &= z_{0,9} \cdot \sqrt{\frac{8,4^2}{932} + \frac{8,4^2}{253}} \\ &\approx 1,28 \cdot 0,65 \\ &\approx 0,83 \end{aligned}$$

- H_0 wird abgelehnt, Personen aus den alten Bundesländern hören signifikant länger Radio.

Doppelter t-Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Größen mit Erwartungswerten μ_X und μ_Y
- $X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ **unabhängige** Stichproben
- Testgröße: Differenz der Mittelwerte

$$\begin{aligned} T &= \overline{X} - \overline{Y} \\ &= \frac{1}{m}\sum_{i=1}^{m} X_i - \frac{1}{n}\sum_{i=1}^{n} Y_i \end{aligned}$$

- X und Y sind **normalverteilt**.
- Die Varianzen sind gleich $\sigma_X^2 = \sigma_Y^2$

Doppelter t-Test auf die Erwartungswertdifferenz bei unabhängigen Stichproben

Annahmebereich

Für die beiden einseitigen t-Tests und den zweiseitigen t-Test auf die Differenz

$$H_0: \mu_X - \mu_Y = d_0 \qquad H_0: \mu_X - \mu_Y \leq d_0 \qquad H_0: \mu_X - \mu_Y \geq d_0$$
$$H_1: \mu_X - \mu_Y \neq d_0 \qquad H_1: \mu_X - \mu_Y > d_0 \qquad H_1: \mu_X - \mu_Y < d_0$$

ist der Annahmebereich

$$d_0 \pm t_{(m+n-2;1-\frac{\alpha}{2})} s_d \qquad]-\infty, d_0 + t_{(m+n-2;1-\alpha)} s_d] \qquad [d_0 - t_{(m+n-2;1-\alpha)} s_d, \infty[$$

mit

$$\begin{aligned} s_d &= s_{X,Y} \cdot \sqrt{\frac{1}{m} + \frac{1}{n}} \\ s^2_{X,Y} &= \frac{1}{m+n-2}\left((m-1)S_X^2 + (n-1)S_Y^2\right) \end{aligned}$$

$t_{(m+n-2;1-\frac{\alpha}{2})}$ und $t_{(m+n-2;1-\alpha)}$ sind die Quantile der t-Verteilung mit $m+n-2$ Freiheitsgraden.

Tests auf Erwartungswertdifferenz bei abhängigen Stichproben

1. Gegeben ist eine verbundene Stichprobe $X_1, \ldots, X_n$ und $Y_1, \ldots, Y_n$
2. Bilde die Differenz

$$W_i \quad = \quad X_i - Y_i \quad i = 1, \ldots, n$$

3. Führe einen Test auf den Erwartungswert von W durch (siehe letzte Vorlesung)
   - $n > 30 \longrightarrow$ Gauß-Test
   - W normalverteilt $\longrightarrow$ t-Test

Differenz von Anteilen bei unabhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Bernoulli-Größen mit

$$\begin{aligned} p_X &= P(X = 1) \\ p_Y &= P(Y = 1) \end{aligned}$$

- $X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ **unabhängige** Stichproben
- Testgröße: Differenz der Anteile

$$\begin{aligned} T &= R_X - R_Y \\ &= \frac{1}{m}\sum_{i=1}^{m} X_i - \frac{1}{n}\sum_{i=1}^{n} Y_i \end{aligned}$$

- Faustregel: Stichprobenumfänge $m, n > 30$

Differenz von Anteilen bei unabhängigen Stichproben

Annahmebereich

Für die beiden einseitigen Tests und den zweiseitigen Test auf die Differenz

$H_0 : p_X - p_Y = 0$	$H_0 : p_X - p_Y \leq 0$	$H_0 : p_X - p_Y \geq 0$
$H_1 : p_X - p_Y \neq 0$	$H_1 : p_X - p_Y > 0$	$H_1 : p_X - p_Y < 0$

ist der Annahmebereich

$[-z_{1-\frac{\alpha}{2}} \cdot s_r \; ; \; z_{1-\frac{\alpha}{2}} \cdot s_r]$	$]-\infty, z_{1-\alpha} \cdot s_r]$	$[-z_{1-\alpha} \cdot s_r, \infty[$

mit

$$s_r = \sqrt{r(1-r)\left(\frac{1}{m} + \frac{1}{n}\right)} \text{ und } r = \frac{m \cdot r_X + n \cdot r_Y}{m+n}$$

Beispiel: Ist Fernsehen informativ?

Weiterführung des Beispiels aus dem Thema ,,Schätzen"

- Beurteilen Personen aus den alten Bundesländern den Informationsgehalt im Fernsehen anders als Personen aus den neuen Bundesländern?
  zweiseitiger Test
  X: Person aus den alten Bundesländern hält Fernsehen für informativ
  Y: Person aus den neuen Bundesländern hält Fernsehen für informativ
- Signifikanzniveau: $\alpha = 0,05$
- Umfrage: 253 Personen aus den alten Bundesländern, 932 Personen aus den neuen Bundesländern:
  ,,Halten Sie Fernsehen für informativ? Ja/Nein"
  - unverbundene Stichproben $X_1, \dots, X_{253}$ und $Y_1, \dots, Y_{932}$
  - Stichprobengrößen $m = 253, n = 932 > 30$

Beispiel: Ist Fernsehen informativ?

- alte Bundesländer: 206 Personen halten Fernsehen für informativ
  neue Bundesländer: 747 Personen halten Fernsehen für informativ
- Anteile

$$R_X = \frac{206}{253} \approx 0,81 \quad \text{und} \quad R_Y = \frac{747}{932} \approx 0,80$$

- Standardfehler

$$\begin{aligned} r &= \frac{m \cdot r_X + n \cdot r_Y}{m+n} = \frac{253 \cdot 0,81 + 932 \cdot 0,8}{932 + 253} \\ &= \frac{950,53}{1185} \approx 0,802 \\ s_r &= \sqrt{r(1-r)(\frac{1}{m} + \frac{1}{n})} = \sqrt{0,802 \cdot (1 - 0,802) \left(\frac{1}{932} + \frac{1}{253}\right)} \\ &\approx 0,03 \end{aligned}$$

Beispiel: Ist Fernsehen informativ?

- Annahmebereich

$$\begin{aligned} \pm z_{1-\frac{\alpha}{2}} \cdot s_r &\approx \pm z_{0,975} \cdot 0,03 \\ &\approx \pm 1,96 \cdot 0,03 \\ &\approx \pm 0.059 \end{aligned}$$

- Differenz der Anteile in der Stichprobe

$$R_X - R_Y \approx 0,81 - 0,8 = 0,01$$

- H_0 wird beibehalten, der Unterschied zwischen alten und neuen Bundesländern ist nicht signifikant

Differenz von Anteilen bei abhängigen Stichproben

Voraussetzungen:

- X und Y sind zwei Bernoulli-Größen mit

$$\begin{aligned} p_X &= P(X = 1) \\ p_Y &= P(Y = 1) \end{aligned}$$

- $(X_1, Y_1), \ldots, (X_n, Y_n)$ **abhängige, verbundene** Stichproben
- Absolute Haufigkeiten werden in einer Kontingenztafel festgehalten

	Y=0	Y=1
X=0	n_{11}	n_{12}
X=1	n_{21}	n_{22}

- Test von McNemar! (mehr Details in Blatt 8)

Zusammenhang zwischen 2 kategorialen Merkmalen

- Sind zwei kategoriale Merkmale unabhängig?
- **Beispiele**
  - Gibt es einen Zusammenhang zwischen besuchter Schule (Hauptschule, Realschule, Gymnasium) und Fernsehkonsum (hoch/niedrig)?
  - Gibt es einen Zusammenhang zwischen Geschlecht (m/w) und der Affinität zu Fußball (Fan/kein Fan)?
  - ...
- empirische Untersuchung mittels Kontingenztafeln und Kennzahlen

Erinnerung: Randverteilungen

Die zu den einzelnen Merkmalen gehörigen empirischen Häufigkeitsverteilungen heißen **Randverteilungen** oder **marginale** Verteilungen.

Geschlecht \ Fußball	ja	nein	Summe
m	87	10	97
w	23	31	54
Summe	110	41	151

Randverteilung Geschlecht

m	w	
97	54	$\leftarrow$ Absolute Häufigkeiten
$\frac{97}{151}$	$\frac{54}{151}$	$\leftarrow$ Relative Häufigkeiten

Randverteilung Fußballfan

ja	nein	
110	41	$\leftarrow$ Absolute Häufigkeiten
$\frac{110}{151}$	$\frac{41}{151}$	$\leftarrow$ Relative Häufigkeiten

Erinnerung: Bedingte Verteilung

Unter der **bedingten Verteilung** von X gegeben $Y = B_j$ versteht man die Verteilung von X in der Teilgesamtheit der Untersuchungseinheiten, die die Ausprägung B_j des Merkmals Y aufweisen.

Geschlecht \ Fußball	ja	nein	Summe
m	87	10	97
w	23	31	54
Summe	110	41	151

Verteilung „Fußballfan“ bei Männern

ja	nein	Ges.
87	10	97
90%	10%	100%

Verteilung “Geschlecht“ bei Fußballfans

m	w	Ges.
87	23	110
79%	21%	100%

Erinnerung: Empirische Unabhängigkeit

- Falls die beiden Merkmale unabhängig voneinander sind, so sollte

  ,,relative Häufigkeit Fußballfan"
  ≈ ,,relative Häufigkeit Fußballfans unter Frauen"
  ≈,,relative Häufigkeit Fußballfans unter Männern"

- Formel:

$$\text{bedingte relative Häufigkeit } \frac{n_{ij}}{n_{\bullet j}} = \frac{n_{i\bullet}}{n} \text{ relative Randhäufigkeit}$$

- Daraus folgt

$$e_{ij} := \frac{n_{i\bullet} n_{\bullet j}}{n}$$

  ist die unter Unabhängigkeit erwartete Anzahl.

- Falls die Merkmale unabhängig sind, sollte gelten:

$$e_{ij} \approx n_{ij}$$

Erinnerung: Quadratische Kontingenz χ^2

- χ^2 (sprich: chi-quadrat) ist ein Maß für die empirische Unabhängigkeit zweier nominaler Merkmale
- Definition: Summe der mit den Unabhängigkeitszahlen e_{ij} normierten quadratischen Abweichungen der Besetzungszahlen n_{ij} von den Unabhängigkeitszahlen.

Quadratische Kontingenz

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

Es gilt: $0 \leq \chi^2 \leq \infty$

χ^2-Unabhängigkeitstest

- zwei Zufallsgrößen X und Y mit k bzw. l Ausprägungen

$$p_{ij} = P(X = i, Y = j)$$
$$p_{i\bullet} = P(X = i) \quad p_{\bullet j} = P(Y = j)$$

- Hypothesen

$$\begin{aligned} H_0 &: \quad \text{X und Y sind stochastisch unabhängig} \\ & \quad\; p_{ij} = p_{i\bullet} \cdot p_{\bullet j} \text{ für alle } i = 1, \dots, k,\ j = 1, \dots, l \\ H_1 &: \quad \text{X und Y sind stochastisch abhängig} \\ & \quad\; p_{ij} \neq p_{i\bullet} \cdot p_{\bullet j} \text{ für mindestens eine ij-Kombination} \end{aligned}$$

- Prüfgröße

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

- Faustregel $n_{ij} \geq 5$ für alle i, j

χ^2-Unabhängigkeitstest

Annahmebereich

Für den χ^2-Unabhängigkeitstest ist der Annahmebereich

$$\chi^2 \quad \leq \quad q_{1-\alpha,\,(k-1)(l-1)}$$

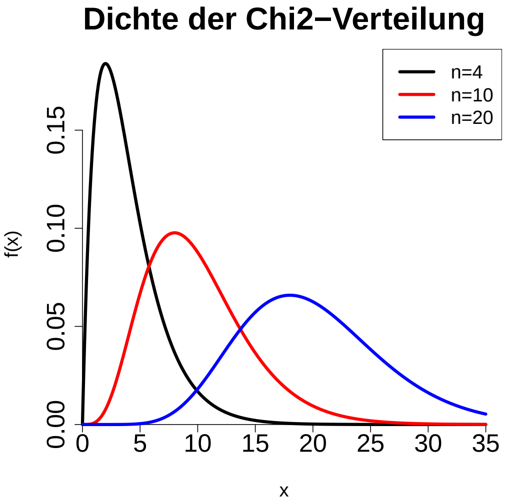
Die Nullhypothese wird abgelehnt, falls

$$\chi^2 \quad > \quad q_{1-\alpha,\,(k-1)(l-1)}$$

Dabei ist $q_{1-\alpha,\,(k-1)(l-1)}$ das

- $(1-\alpha)$-Quantil
- der χ^2-Verteilung
- mit $(k-1)\cdot(l-1)$ Freiheitsgraden.

Erinnerung: Dichte der χ^2-Verteilung mit n Freiheitsgraden



Beispiel

$$e_{ij} = \frac{n_{i\bullet} n_{\bullet j}}{n}$$

Geschlecht \ Fußball	ja	nein	Summe
m	87	10	97
w	23	31	54
Summe	110	41	151

Erwartete Besetzungszahlen bei Unabhängigkeit

	ja (j=1)	nein (j=2)
m (i=1)	$\frac{97 \cdot 110}{151} \approx 71$	$\frac{97 \cdot 41}{151} \approx 26$
w (i=2)	$\frac{54 \cdot 110}{151} \approx 39$	$\frac{54 \cdot 41}{151} \approx 15$

$$\begin{aligned}
\chi^2 &= \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \\
&\approx \frac{(87-71)^2}{71} + \frac{(10-26)^2}{26} + \frac{(23-39)^2}{39} + \frac{(31-15)^2}{15} \\
&\approx 37,09
\end{aligned}$$

Beispiel

- Signifikanzniveau: $\alpha = 0,01$
- Faustregel gültig?
  Besetzungszahlen $n_{ij} \geq 5$
- Bestimmung der Freiheitsgrade: $k = l = 2$

$$\text{Freiheitsgrade} \quad = \quad (k-1)\cdot(l-1) = (2-1)\cdot(2-1) = 1$$

- Ablesen des Quantils

$$q_{1-0,01;\,(2-1)(2-1)} \quad = \quad q_{0,09;\,1} \approx 6,63$$

- H_0 wird abgelehnt

Zusammenfassung Tests

Statistischer Test

Ein statistischer Test ist eine Untersuchung, ob man eine Hypothese über die Grundgesamtheit mit Hilfe einer Stichprobe widerlegen kann

Wichtige Tests

- einseitige Tests $\longleftrightarrow$ zweiseitiger Test
- Tests auf Anteile, Erwartungswerte und deren Differenzen
  - eine Stichprobe
  - zwei Stichproben $\rightarrow$ verbunden oder unverbunden
- Test auf Unabhängigkeit: χ^2-Unabhängigkeitstest

Interpretation eines Tests

- Statistische Signifikanz bedeutet: Das Ergebnis ist nicht durch Zufall erklärbar.
- Statistische Signifikanz bedeutet nicht unbedingt, dass der Unterschied relevant ist.
  Beispiel: Der Besuch eines kostenpflichtigen Tutoriums erhöht Ihre Chance, die Klausur zu bestehen, signifikant – um $0,01\%$!
  $\rightarrow$ Daher immer die Größe des Effekts (Schätzung) angeben!
- Ein statistischer Test liefert aber im Allgemeinen keinen kausalen Zusammenhang.