



4. Testtheorie

4.1. Nullhypothese, Gegenhypothese und Entscheidung

Hypothesen – Annahmen über die Verteilung oder über einzelne Parameter der Verteilung eines Merkmals in einer Grundgesamtheit.

Es kann durch Testen *nicht* festgestellt werden, ob eine Hypothese *richtig* oder *falsch* ist, sondern nur ob sie beibehalten oder verworfen werden kann. „Annahme“ einer statistischen Hypothese bedeutet immer



nur: Die vorliegende statistische Evidenz reicht nicht aus, um die Hypothese zu verwerfen.

Beim Testen stehen sich zwei Behauptungen gegenüber – die Nullhypothese (H_0) und die Gegenhypothese oder Alternativhypothese (H_1).

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

Der Test führt zur Entscheidung über die Annahme oder Ablehnung der Nullhypothese.

Dabei sind vier Fälle des Zusammentreffens von Realität und Testentscheidung möglich:



Realität \ Testentscheidung	H_0 ist richtig	H_0 ist falsch
	H_0 beibehalten	H_0 verworfen
H_0 beibehalten	o.k.	Fehler 2. Art β -Fehler
H_0 verworfen	Fehler 1. Art α -Fehler	o.k.

Quelle: Schira (2003), S. 474.

Fehler 1. Art:

Die Nullhypothese wird verworfen, obwohl sie richtig ist.

Fehler 2. Art:

Die Nullhypothese wird angenommen, obwohl sie falsch ist.



4.2 Tests für Mittel- und Anteilswerte (Ein-Stichproben-Fall)

4.2.1. Heterogener Fall (Erwartungswerte)

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0 \text{ (zweiseitige Fragestellung)}$$

$$\left. \begin{array}{l} \mu > \mu_0 \\ \mu < \mu_0 \end{array} \right\} \text{(einseitige Fragestellung)}$$

Aufgrund einer Stichprobe soll entschieden werden, ob die Nullhypothese zu verwerfen ist. Meistens wird man einen Stichprobenmittelwert finden, der von dem hypothetischen Wert μ_0 abweicht. Dies bedeutet



jedoch nicht, dass die Nullhypothese gleich verworfen werden soll. Die Abweichung könnte auch **zufällig** sein, d.h. durch ungünstige Auswahl der Stichprobenelemente beeinflusst werden.

Erst wenn die Abweichung einen bestimmten *kritischen Wert* überschreitet, kann sie als *signifikant von Null verschieden* angesehen werden, die Nullhypothese wird dann verworfen zu Gunsten von Alternativhypothese.

Zur Testentscheidung wird die Verteilung des Stichprobenmittelwertes herangezogen:



Die Nullhypothese wird beibehalten, solange das mit dieser Stichprobe geschätzte Konfidenzintervall um den Stichprobenmittelwert den hypothetischen Wert μ_0 enthält.

Mit der Verteilung des Stichprobenmittelwertes $\bar{X}$ bei Gültigkeit von H_0 kann für $\bar{X}$ ein **Annahmebereich** und ein **Ablehnungsbereich** so bestimmt werden, dass die Wahrscheinlichkeit, mit der der Stichprobenmittelwert in den Ablehnungsbereich fällt (bei Gültigkeit der Nullhypothese), höchstens α beträgt.



Die Wahrscheinlichkeit des Fehlers 1. Art α heißt auch **Signifikanzniveau** des Tests.

Prüfgröße (Teststatistik):

$$T = \frac{|\bar{x} - \mu_0|}{\sigma_{\bar{x}}}$$

Die Prüfgröße T ist bei kleinen Stichproben t -verteilt mit $n-1$ Freiheitsgraden und bei großen Stichproben ($n > 30$) standardnormalverteilt.



Bei kleinen Stichproben wird die t -Verteilung dann herangezogen, wenn die *Varianz geschätzt werden muss*. Voraussetzung dafür ist aber, dass man eine Normalverteilung des Merkmals in der Grundgesamtheit annehmen kann.

Die **Entscheidungsregel** lautet

- bei großen Stichproben:

$$\left. \begin{array}{l} T = \frac{|\bar{x} - \mu_0|}{\sigma_{\bar{x}}} > z[1 - \alpha / 2] \text{ bei zweiseitigen Tests} \\ T = \frac{|\bar{x} - \mu_0|}{\sigma_{\bar{x}}} > z[1 - \alpha] \text{ bei einseitigen Tests} \end{array} \right\} \Rightarrow H_0 \text{ verwerfen!}$$



- bei kleinen Stichproben:

$$\left. \begin{array}{l} T = \frac{|\bar{X} - \mu_0|}{\hat{\sigma}_{\bar{X}}} > t_{n-1}[1 - \alpha / 2] \text{ bei zweiseitigen Tests} \\ T = \frac{|\bar{X} - \mu_0|}{\hat{\sigma}_{\bar{X}}} > t_{n-1}[1 - \alpha] \text{ bei einseitigen Tests} \end{array} \right\} \Rightarrow H_0 \text{ verwerfen!}$$

4.2.2. Homograde Fall (Anteilswerte)

In der Nullhypothese wird behauptet, dass der Anteilswert einer Grundgesamtheit (P) gleich dem hypothetischen Anteilswert P_0 ist:



$$H_0 : P = P_0$$

$$H_1 : P \neq P_0 \text{ (zweiseitige Fragestellung)}$$

$$\left. \begin{array}{l} P > P_0 \\ P < P_0 \end{array} \right\} \text{ (einseitige Fragestellung)}$$

Die Entscheidungsregel:

$$\left. \begin{array}{l} T = \frac{|h - P_0|}{\sigma_H} > z[1 - \alpha / 2] \text{ bei zweiseitigen Tests} \\ T = \frac{|h - P_0|}{\sigma_H} > z[1 - \alpha] \text{ bei einseitigen Tests} \end{array} \right\} \Rightarrow H_0 \text{ verwerfen!}$$

h : Anteilswert der Stichprobe



Anmerkung: die Standardabweichung darf nicht mit dem Anteilswert der Stichproben (h) geschätzt werden, sondern ist mit dem hypothetischen Wert P_0 zu berechnen.

$$\sigma_H = \sqrt{\frac{P_0(1 - P_0)}{n}}$$

Bei **großen Stichproben** kann die **Standardnormalverteilung** verwendet werden.

Bei **kleinen Stichproben** sollte man die **Binomialverteilung** nehmen.



Der Binomialtest

1. linksseitiger Test:

$$H_0 : P \geq P_0$$

$$H_1 : P \leq P_0$$

Die Nullhypothese wird verworfen, wenn ein kritischer Wert x_{unten} unterschritten wird: $P(X < x_{unten}) \leq \alpha$

2. rechtsseitiger Test:

$$H_0 : P \leq P_0$$

$$H_1 : P \geq P_0$$



Die Nullhypothese wird verworfen, wenn ein kritischer Wert x_{oben} überschritten wird.

$$P(X > x_{oben}) \leq \alpha$$

3. zweiseitiger Test:

$$H_0 : P = P_0$$

$$H_1 : P \neq P_0$$

In diesem Fall braucht man zwei kritische Werte. Die Irrtumswahrscheinlichkeit wird aufgeteilt.

$$P(X < x_{unten}) \leq \alpha / 2 \text{ und } P(X > x_{oben}) \leq \alpha / 2$$



4.3 Tests für Varianzen

Nullhypothese: die Varianz einer Grundgesamtheit σ^2 weicht nicht von einem vorgegebenen hypothetischen Wert σ_0^2 ab.

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 \neq \sigma_0^2 \text{ (zweiseitige Fragestellung)}$$

$$\left. \begin{array}{l} \sigma^2 > \sigma_0^2 \\ \sigma^2 < \sigma_0^2 \end{array} \right\} \text{ (einseitige Fragestellung)}$$



Die Testgröße $\frac{n \cdot S^2}{\sigma_0^2} = \chi_{n-1}^2$ ist Chi-Quadrat-verteilt mit $n-1$ Freiheitsgraden, wenn das Merkmal in der Grundgesamtheit normalverteilt ist.

Die Entscheidungsregel:

$$T = \frac{n \cdot S^2}{\sigma_0^2} \left\{ \begin{array}{l} < \chi_{n-1}^2 [\alpha / 2] \text{ oder} \\ > \chi_{n-1}^2 [1 - \alpha / 2] \end{array} \right\} \Rightarrow H_0 \text{ verwerfen !}$$

mit $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ (Stichprobenvarianz)



Die Nullhypothese wird verworfen, wenn die Testgröße das untere Quantil unterschreitet oder das obere überschreitet.

4.4 Vergleich zweier Erwartungswerte (Zwei-Stichproben-Fall, heterograd)

Seien $\bar{x}_1$ und $\bar{x}_2$ Mittelwerte aus zwei *unabhängigen* Stichproben. Es soll getestet werden, ob die beiden Stichproben aus derselben Grundgesamtheit stammen (oder wenigstens aus Grundgesamtheiten mit gleichem Mittelwert).



$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2 \text{ (zweiseitige Fragestellung)}$$

$$\left. \begin{array}{l} \mu_1 > \mu_2 \\ \mu_1 < \mu_2 \end{array} \right\} \text{ (einseitige Fragestellung)}$$

Bei **großen Stichproben** (n_1 und n_2 müssen groß sein) gilt der zentrale Grenzwertsatz und man kann **die Normalverteilung** anwenden.



Die Entscheidungsregel lautet:

$$\left. \begin{array}{l} T = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma_{\Delta}} > z[1 - \alpha / 2] \text{ bei zweiseitigen Tests} \\ T = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma_{\Delta}} > z[1 - \alpha] \text{ bei einseitigen Tests} \end{array} \right\} \Rightarrow H_0 \text{ verwerfen!}$$

mit $\sigma_{\Delta} = \sqrt{\sigma_1^2 / n + \sigma_2^2 / n}$



Bei **kleinen Stichproben** (wenn σ_{Δ} geschätzt werden muss) ist die Testgröße **t-verteilt** mit n_1+n_2-2 Freiheitsgraden. Voraussetzung:

$$\sigma_1 = \sigma_2 = \sigma .$$

$$\left. \begin{array}{l} T = \frac{|\bar{x}_1 - \bar{x}_2|}{\hat{\sigma}_{\Delta}} > t_{n_1+n_2-2} [1 - \alpha / 2] \text{ bei zweiseitigen Tests} \\ T = \frac{|\bar{x}_1 - \bar{x}_2|}{\hat{\sigma}_{\Delta}} > t_{n_1+n_2-2} [1 - \alpha] \text{ bei einseitigen Tests} \end{array} \right\} \Rightarrow H_0 \text{ verwerfen!}$$

$$\text{mit } \hat{\sigma}_{\Delta} = \sqrt{\hat{\sigma}_1^2 / n + \hat{\sigma}_2^2 / n}$$



4.5 Vergleich zweier Anteilswerte (homograd)

Hypothesen:

$$H_0 : P_1 = P_2$$

$$H_1 : P_1 \neq P_2 \text{ (zweiseitige Fragestellung)}$$

$$\left. \begin{array}{l} P_1 > P_2 \\ P_1 < P_2 \end{array} \right\} \text{ (einseitige Fragestellung)}$$

Die Entscheidungsregel:

$$\left. \begin{array}{l} T = \frac{|p_1 - p_2|}{\hat{\sigma}_R} > z[1 - \alpha / 2] \text{ bei zweiseitigen Tests} \\ T = \frac{|p_1 - p_2|}{\hat{\sigma}_R} > z[1 - \alpha] \text{ bei einseitigen Tests} \end{array} \right\} \Rightarrow H_0 \text{ verwerfen!}$$



mit $\hat{\sigma}_R = \sqrt{p(1-p)(1/n_1 + 1/n_2)}$

4.6. *F*-Verteilung

Seien Q_n und Q_m zwei unabhängige Chi-Quadrat-verteilte Zufallsvariablen. Die Verteilung des Quotienten Z nennt man dann *F*-Verteilung mit n und m Freiheitsgraden:



$$Z = \frac{\frac{Q_n}{n}}{\frac{Q_m}{m}} \sim F_{n,m} \quad \text{mit } n: \text{Anzahl der Zählerfreiheitsgrade und } m: \text{Anzahl}$$

der Nennerfreiheitsgrade

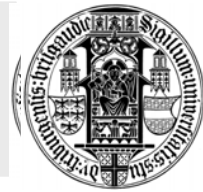
Der Erwartungswert und die Varianz werden wie folgt bestimmt:

$$E[Z] = \frac{m}{m-2} \quad \text{falls } m \geq 3$$

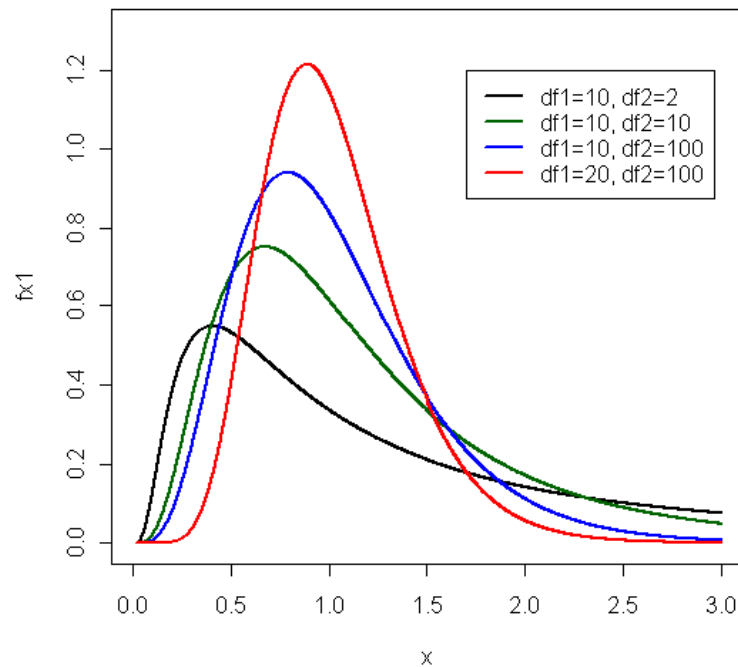


$$V[Z] = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)} \quad \text{falls } m \geq 5$$

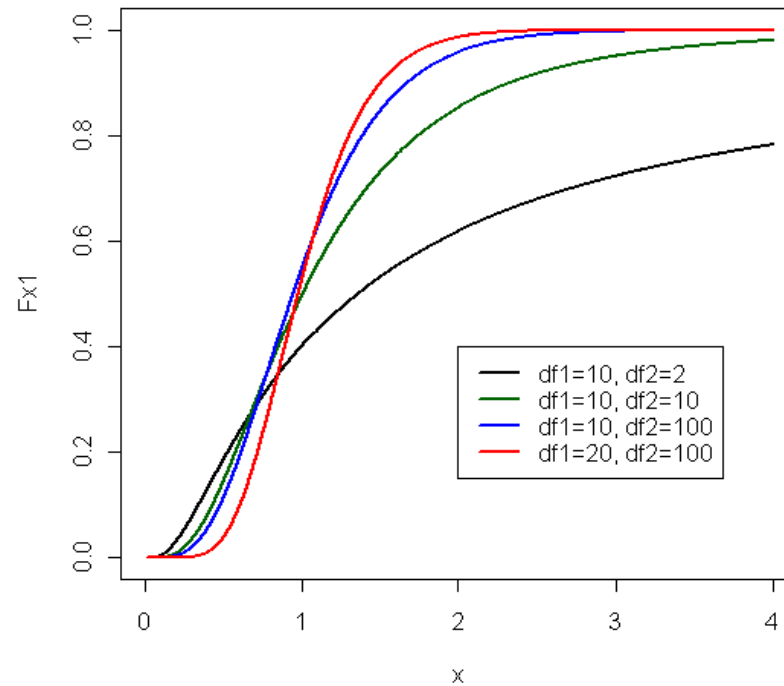
Die F -Verteilung ist eine stetige Verteilung. Sie ist asymmetrisch rechts-schief (linkssteil).



F-Verteilung



Verteilungsfunktion



Für große m lässt sich die F -Verteilung durch die Chi-Quadrat-Verteilung approximieren.



Die F -Verteilung wird zum Vergleich von Streuungen verwendet.

4.7 Vergleich zweier Varianzen

Annahme: die Stichprobenvarianzen s_1^2 und s_2^2 stammen aus zwei verschiedenen unabhängigen Stichproben.

In der Nullhypothese wird behauptet, dass die beiden Stichproben aus derselben Grundgesamtheit stammen:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$$

Die Testgröße ist F -verteilt mit den Freiheitsgraden $n_1 - 1$ und $n_2 - 2$.



Voraussetzung: das Merkmal ist in der Grundgesamtheit normalverteilt mit der Varianz σ^2 .

Die Nullhypothese wird verworfen, wenn die Testgröße das untere $\alpha/2$ -Quantil unterschreitet oder das obere $(1 - \alpha/2)$ -Quantil überschreitet.

Die Entscheidungsregel:

$$T = \frac{\frac{n_1}{n_1 - 1} s_1^2}{\frac{n_2}{n_2 - 1} s_2^2} \left\{ \begin{array}{l} < F[\alpha / 2] \text{ oder} \\ > F[1 - \alpha / 2] \end{array} \right\} \Rightarrow H_0 \text{ verwerfen!}$$



Im Zähler und im Nenner sind erwartungstreue Schätzwerte der Varianzen.

4.8 Signifikanzniveau und Überschreitungswahrscheinlichkeit

Bisher: das Signifikanzniveau α wurde immer vor dem Testen festgelegt.

Nachteil:

für die Testentscheidung spielt es dann keine Rolle, ob die Teststatistik (Prüfgröße T) weit im Annahmehereich oder nur knapp an der kritischen Grenze liegt.



Alternative Möglichkeit wäre die Abweichung zwischen hypothetischem Wert des Parameters und dem Stichprobenergebnis zu berücksichtigen und die Wahrscheinlichkeit anzugeben, mit der das Stichprobenergebnis bei Gültigkeit der Nullhypothese den kritischen Wert über- oder unterschreiten würde.

Vorteil:

Vermeidung der „willkürlichen“ Festlegung eines Signifikanzniveaus und effizientere Nutzung der Stichprobeninformationen.

**Definition:**

Die Überschreitungswahrscheinlichkeit (P-Wert, probability value) ist das Signifikanzniveau, bei dem die Testgröße T auf den kritischen Wert fällt.

P -Wert wird auch als empirisches Signifikanzniveau bezeichnet.

Je kleiner der P -Wert, desto eher wird man die Nullhypothese verwerfen.



4.9 Macht und Trennschärfe eines Tests

Ein Test ist umso trennschärfer, je zuverlässiger man mit ihm erkennen kann, ob eine Hypothese richtig oder falsch ist.

Idealer Test: $\alpha = 0 = \beta$.

Nur in diesem Fall könnte die Nullhypothese mit Wahrscheinlichkeit Eins verworfen oder angenommen werden.

In der Realität ist jedoch ein solcher Test unmöglich!



Ist für den Test ein kleines Signifikanzniveau α (z.B. von 1%) gewählt, wird es immer problematisch die Nullhypothese zu verwerfen, wenn θ nur wenig größer als θ_0 ist.

Wie schnell bei einem konkreten Test die Ablehnungswahrscheinlichkeit zunimmt, ist für seine Brauchbarkeit ganz entscheidend, deswegen wird diese Eigenschaft **die Macht oder die Güte eines Tests** genannt.

Definition:

Die Funktion $G(\theta) = P(H_0 \text{ verwerfen} | \theta)$, welche die Wahrscheinlichkeit die Nullhypothese zu verwerfen in Abhängigkeit vom wahren Parameter θ angibt, heißt **Machtfunktion** oder **Gütefunktion** des Tests.

Die Machtfunktion G hängt von den Parametern μ , μ_0 , σ , n und α ab.