

## 8 Schliessende Statistik – Hypothesentests

### 8.1 Einführung

#### Problemstellung

Jemand hat eine Behauptung bzw. Hypothese aufgestellt. Diese soll nun durch ein statistisches Verfahren auf der Basis einer Stichprobe überprüft werden.

#### Beispiele

1. Ein Automobilhersteller gibt an, dass der mittlere Benzinverbrauch für einen bestimmten Autotyp  $\mu_0 = 8.2$  l/100 km beträgt. Stimmt diese Angabe?
2. Der Anteil unzufriedener Kunden einer Airline lag bei 32%. Eine Verbesserung des Kundenservices sollte zu einer Reduktion der unzufriedenen Kunden führen. Eine Umfrage nach einem halben Jahr unter 112 zufällig ausgewählten Kunden führt zu 26 unzufriedenen Kunden. Ist eine Verbesserung erkennbar?
3. Die typischen Laufzeiten eines Programms sind bekannt. Nun soll beurteilt werden, ob das System stabil läuft anhand der aktuell gemessenen Laufzeiten.
4. Zwei verschiedene Messmethoden für elektrische Widerstände sollen miteinander verglichen werden. Liefern sie dieselben Resultate?

Es reicht nun nicht, zu überprüfen, ob der Durchschnitt oder der Anteil oder die Verteilung der Stichprobe der Behauptung entspricht. Selbst wenn die Behauptung stimmt, müssen wir mit zufälligen Abweichungen rechnen. Die Frage lautet also:

*Ist es plausibel, die in der Stichprobe beobachteten Abweichungen von der Behauptung als zufällig zu betrachten? Liegen sie „im Rahmen“?*

Wir suchen objektive Kriterien für diese Entscheidung.

### 8.2 Vorgehen bei einem Hypothesentest

Im Folgenden wird Schritt für Schritt das Vorgehen bei einem Hypothesentest erklärt. Ergänzen Sie bei jedem Schritt die entsprechenden Angaben für das folgende Beispiel:

#### Beispiel 1: Kraftstoffverbrauch

Ein Automobilhersteller bringt ein neues PKW-Modell auf den Markt, dessen Benzinverbrauch  $X$  (in Liter pro 100 km Fahrleistung) eine *normalverteilte* Zufallsvariable mit dem Mittelwert  $\mu_0 = 8.2$  l/100 km sein soll. Die Redaktion einer Fachzeitschrift überprüft diese Angabe an  $n = 25$  zufällig ausgewählten Testfahrzeugen und kommt dabei zu folgendem Ergebnis:

$$\bar{x} = 9.1 \text{ l/100 km} \qquad s = 2.1 \text{ l/100 km}$$

Kann somit die Angabe des Herstellers bezüglich des Mittelwertes  $\mu$  aufrechterhalten werden?

#### 1. Schritt: Nullhypothese formulieren

Wir stellen die sogenannte *Nullhypothese*  $H_0$  auf, die die Behauptung beschreibt:

Die Nullhypothese wird in einem gewissen Sinn bevorzugt behandelt, nach dem Motto: "Im Zweifelsfall für den Angeklagten". D.h. wir gehen davon aus, dass  $H_0$  wahr ist, bis wir uns vom Gegenteil überzeugt haben.

## 2. Schritt: Alternativhypothese formulieren

Wir formulieren die sogenannte *Alternativhypothese*  $H_A$ . Wenn wir zum Schluss kommen, dass  $H_0$  nicht haltbar ist, nehmen wir an, dass  $H_A$  zutrifft.

Je nach Fragestellung gibt es auch andere Möglichkeiten für  $H_A$ :

Die Alternativhypothese ist weniger spezifisch als die Nullhypothese. In unserem Beispiel gibt sie keinen konkreten Wert für  $\mu$  an.

## 3. Schritt: Testvariable definieren

Wir bestimmen eine sogenannte *Testvariable*, eine dem konkreten Problem angepasste Stichprobenfunktion:

$$T = g(X_1, X_2, \dots, X_n)$$

In unserem Beispiel nehmen wir als Ausgangspunkt die Schätzfunktion für den Mittelwert:

Diese muss nun standardisiert werden (wie bei den Vertrauensintervallen):

(Schon hier merkt man die bevorzugte Behandlung von  $H_0$  – bei der Standardisierung gehen wir davon aus, dass  $\mu = \mu_0$  gilt.)

## 4. Schritt: Wahrscheinlichkeitsverteilung der Testvariable bestimmen

Wir bestimmen die Wahrscheinlichkeitsverteilung, der die Testvariable  $T$  folgt mit Tabelle 8.2.1 *unter der Voraussetzung, dass  $H_0$  wahr ist*.

In unserem Beispiel handelt es sich um eine .....

## 5. Schritt: Signifikanzzahl festlegen

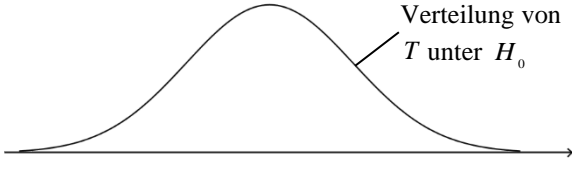
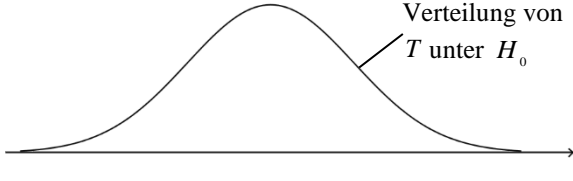
Wir wählen eine *Signifikanzzahl*:  $\alpha = 5\%$  (ebenfalls üblich ist 1%).

## 6. Schritt: Kritische Grenzen bestimmen

Wir bestimmen (aus der Tabelle für die Wahrscheinlichkeitsverteilung von  $T$ ) zwei *kritische Grenzen*  $c_u$  und  $c_o$  derart, dass gilt:

$$P(c_u \leq T \leq c_o) = 1 - \alpha$$

Bei einer zweiseitigen Alternativhypothese wird die Fläche des *kritischen Bereichs*, die  $\alpha$  entspricht, je zur Hälfte rechts und links abgetragen, bei einer einseitigen Alternativhypothese nur auf einer Seite.

	
2-seitige Alternativhypothese ( $H_A: \mu \neq \mu_0$ )	1-seitige Alternativhypothese ( $H_A: \mu > \mu_0$ )
$c_u =$ $c_o =$	$c_o =$

Bei der 1-seitigen Alternativhypothese  $H_A: \mu < \mu_0$  befindet sich der kritische Bereich dann stattdessen links mit  $c_u = -1.711$ .

## 7. Schritt: Testwert berechnen

Wir berechnen den *Testwert*  $\hat{t}$ , indem wir die Stichprobenwerte  $x_1, x_2, \dots, x_n$  in  $T = g(X_1, X_2, \dots, X_n)$  einsetzen.

## 8. Schritt: Testentscheidung:

Fällt der Testwert  $\hat{t}$  in den *Annahmebereich*, gilt also  $c_u \leq \hat{t} \leq c_o$ , dann wird  $H_0$  als plausibel angenommen.

Fällt der Testwert  $\hat{t}$  in den *kritischen Bereich*, gilt also  $\hat{t} < c_u$  oder  $c_o < \hat{t}$ , dann wird  $H_0$  als unwahrscheinlich abgelehnt.

## Bemerkung

Hypothesentests mit  $t$ -verteilter Testvariable werden allgemein als  $t$ -Tests bezeichnet.

### Vorgehen bei einem Parametertest

#### 1. Nullhypothese $H_0$ aufstellen

Um welchen Parameter geht es? Welchen Wert hat er angeblich? Oder werden zwei Parameter verglichen?

#### 2. Alternativhypothese $H_A$ aufstellen

Kommt es darauf an, in welche Richtung die Abweichung geht? Ist dies der Fall, so beschreibt  $H_A$  nur die relevante Alternative.

#### 3. Die richtige Zeile in der Tabelle 8.2.1 finden

Welcher Verteilung folgt die Grundgesamtheit? Um welche Nullhypothese geht es? Welcher Fall liegt vor?

#### 4. Kritische Grenze(n) bestimmen

Dabei müssen wir Folgendes berücksichtigen:

- Verteilung der Testvariablen gemäss Tabelle 8.2.1 (letzte Kolonne)
- Signifikanzniveau  $\alpha$
- Ist  $H_A$  einseitig oder zweiseitig? Wenn einseitig, auf welcher Seite befindet sich der kritische Bereich?

#### 5. Testwert berechnen

gemäss Tabelle 8.2.1 (vorletzte Kolonne).

#### 6. Testentscheidung fällen

Liegt der Testwert im Annahmebereich oder im kritischen Bereich?

## 8.2.1 Übersicht über verschiedene Parametertests

	Verteilung Grundges.	Null-hypothese	Fall	Schätzfunktion	Testvariable (standardisiert)	Verteilung der Testvariablen unter $H_0$
1	Normalverteilung	$\mu = \mu_0$	Varianz $\sigma^2$ bekannt	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$	$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	Standardnormalverteilung <b>T2</b>
2	Normalverteilung	$\mu = \mu_0$	Varianz $\sigma^2$ unbekannt *	$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$	$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	t-Verteilung mit $f = n - 1$ <b>T4</b>
3	2 Normalverteilungen	$\mu_1 - \mu_2 = 0$	Abhängige Stichproben, Varianzen $\sigma_1^2$ und $\sigma_2^2$ bekannt	$\bar{Z} = \bar{X} - \bar{Y}$	$U = \frac{\bar{Z}}{\sigma} \quad \text{mit} \quad \sigma^2 = \frac{\sigma_1^2 + \sigma_2^2}{n}$	Standardnormalverteilung <b>T2</b>
4	2 Normalverteilungen	$\mu_1 - \mu_2 = 0$	Abhängige Stichproben, Varianzen $\sigma_1^2$ und $\sigma_2^2$ unbekannt *	$\bar{Z} = \bar{X} - \bar{Y}$ $S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - Y_i - \bar{Z})^2$	$T = \frac{\bar{Z}}{S/\sqrt{n}}$	t-Verteilung mit $f = n - 1$ <b>T4</b>
5	2 Normalverteilungen	$\mu_1 - \mu_2 = 0$	Unabhängige Stichproben, Varianzen $\sigma_1^2$ und $\sigma_2^2$ bekannt	$Z = \bar{X} - \bar{Y}$	$U = \frac{Z}{\sigma} \quad \text{mit} \quad \sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$	Standardnormalverteilung <b>T2</b>
6	2 Normalverteilungen	$\mu_1 - \mu_2 = 0$	Unabhängige Stichproben, Varianzen $\sigma_1^2$ und $\sigma_2^2$ gleich, aber unbekannt *	$T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}}$		t-Verteilung mit $f = n_1 + n_2 - 2$ <b>T4</b>
7	Normalverteilung	$\sigma^2 = \sigma_0^2$		$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$	$Z = (n-1) \frac{S^2}{\sigma_0^2}$	Chi-Quadrat-Vert. mit $f = n - 1$ <b>T3</b>
8	Bernoulli-Verteilung	$p = p_0$		$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$	$U = \frac{\bar{X} - p_0}{\sqrt{p_0(1-p_0)/n}}$	Standardnormalvert. (näherungsweise) <b>T2</b>

\*) Falls  $n$  (bzw.  $n_1$  und  $n_2$ )  $> 30$  ist, so kann der entsprechende Fall für bekannte Varianzen angewendet werden; dabei dient  $s$  als Schätzwert für  $\sigma$

### Beispiel 2: Kundenzufriedenheit

Der Anteil unzufriedener Kunden einer Airline lag bei 32%. Eine Verbesserung des Kundenservices sollte zu einer Reduktion der unzufriedenen Kunden führen. Eine Umfrage nach einem halben Jahr unter 112 zufällig ausgewählten Kunden führt zu 26 unzufriedenen Kunden. Ist eine Verbesserung erkennbar?

Um welchen Parameter geht es hierbei?.....

1. Nullhypothese:
2. Alternativhypothese:
3. Welche Zeile der Tabelle 8.2.1 passt zu unserem Problem?
4. Die Testvariable folgt

Signifikanzniveau:

$H_A$ :

Kritische Grenze:

5. Testwert berechnen:

6. Testentscheidung:

### Bemerkung

Die Annahme, dass die Testvariable hier einer Standardnormalverteilung folgt ist nur bei grossen Stichprobenzahlen bzw. gemäss der Faustregel  $n\hat{p}(1 - \hat{p}) > 9$  als Näherung zulässig. Daher nutzen Statistikprogramme teils andere Verteilungsfunktionen für die Testvariable.

### Beispiel 3: Programmlaufzeiten

Es ist bekannt, dass die Laufzeit eines Programms (approximativ) normalverteilt ist mit  $\mu = 17s$  und einer bekannten Varianz von  $\sigma^2 = 2.25 s^2$ . Nun werden die aktuellen Programmlaufzeiten gemessen und es soll beurteilt werden, ob es Anpassungen am System benötigt, falls die Laufzeiten abweichen.

Die folgenden Laufzeiten wurden gemessen: 19.2s; 17.4s; 18.5s; 16.5s; 18.9s

1. Nullhypothese:
2. Alternativhypothese:
3. Welche Zeile der Tabelle 8.2.1 passt zu unserem Problem?
4. Die Testvariable folgt

Signifikanzniveau:

$H_A$ :

Kritische Grenze:

5. Testwert berechnen:

6. Testentscheidung:

### **Bemerkung**

Beim Beibehalten der Nullhypothese wurde also kein Widerspruch zum bisherigen Modell gefunden. Dies heisst nicht, dass die Nullhypothese richtig oder gar bewiesen ist – sie bleibt lediglich plausibel. Mehr dazu später...

### **Bemerkung**

Hypothesentests mit normalverteilter Testvariable werden allgemein als Gauss-Tests bezeichnet.

## **8.2.2 Hypothesentests für die Gleichheit der unbekannten Mittelwerte $\mu_1$ und $\mu_2$ zweier Normalverteilungen**

**Frage:** Sind die Mittelwerte  $\mu_1$  und  $\mu_2$  zweier normalverteilter Grundgesamtheiten gleich oder unterscheiden sie sich signifikant?

Wir entnehmen den beiden Grundgesamtheiten, die wir durch die Zufallsvariablen  $X$  und  $Y$  beschreiben, je eine Zufallsstichprobe:

$$x_1, x_2, \dots, x_{n_1} \quad \text{bzw.} \quad y_1, y_2, \dots, y_{n_2}$$

Nullhypothese  $H_0$ :

Alternativhypothese  $H_A$ :

### Definition

Zwei Stichproben heissen voneinander *abhängig*, wenn sie die folgenden Bedingungen erfüllen:

1. ....
2. ....
- .....

### Fall 1: Abhängige Stichproben

#### Beispiel 4

Vergleich zweier Messgeräte: Dieselben Bauelemente aus einer Serienproduktion werden auf zwei verschiedenen Messgeräten gemessen. Zu jedem Bauelement gehört also ein Messwert in der 1. Messreihe und ein Messwert in der 2. Messreihe. Die beiden Messreihen können als ..... Stichproben aufgefasst werden.

#### Beispiel 5

Gegeben sind die monatlichen Umsätze zweier Fluglinien in [Mio CHF] (wir nehmen an, dass Normalverteilungen zugrunde liegen). Unterscheiden sich die Umsätze dieser Fluglinien signifikant?

Monat.	Januar	Februar	März	April	Mai
Fluglinie A [Mio CHF]	14	9	10	12	15
Fluglinie B [Mio CHF]	9	10	12	10	9

mit den bekannten Varianzen  $\sigma_1^2 = 6.5$  [Mio CHF]<sup>2</sup> und  $\sigma_2^2 = 1.5$  [Mio CHF]<sup>2</sup>

1. Nullhypothese:
2. Alternativhypothese:
3. Welche Zeile der Tabelle 8.2.1 passt zu unserem Problem?
4. Die Testvariable folgt  
Signifikanzniveau:  
 $H_A$  :  
Kritische Grenze:
5. Testwert berechnen:



## 6. Testentscheidung:

### Beispiel 6

Beim Testen eines neuen Treibstoffs erhalten 5 PKW zunächst 10ℓ des herkömmlichen Treibstoffs (A) und anschliessend 10ℓ des neuen Treibstoffs (B). Zum Vergleich dient die Fahrleistung in km. Ist die mittlere Fahrleistung mit dem neuen Treibstoff signifikant höher als mit dem herkömmlichen? Wir nehmen an, dass die Fahrleistung für jeden Treibstoff normalverteilt ist. Als Irrtumswahrscheinlichkeit wählen wir  $\alpha = 5\%$ .

PKW-Nr.	1	2	3	4	5
Fahrleistung mit Treibstoff A [km]	95	106	110	115	94
Fahrleistung mit Treibstoff B [km]	96	115	115	114	100
Differenz [km]					

1. Nullhypothese:
2. Alternativhypothese:
3. Welche Zeile der Tabelle 8.2.1 passt zu unserem Problem?
4. Die Testvariable folgt  
Signifikanzniveau:  
 $H_A$ :  
Kritische Grenze:
5. Testwert berechnen:

## 6. Testentscheidung:

## Fall 2: Unabhängige Stichproben

### Beispiel 7

Besitzen die auf zwei verschiedenen Maschinen produzierten elektronischen Bauteile im Mittel die gleiche Lebensdauer? Verwenden Sie das Signifikanzniveau  $\alpha = 1\%$ .

$X$ : Lebensdauer eines auf Maschine A produzierten Bauteils

$Y$ : Lebensdauer eines auf Maschine B produzierten Bauteils

Wir entnehmen den beiden (normalverteilten) Grundgesamtheiten je eine Stichprobe und erhalten folgende Auswertung:

Maschine A:	$n_1 = 80$	$\bar{x} = 500\text{h}$	$s_1 = 50\text{h}$
Maschine B:	$n_2 = 50$	$\bar{y} = 480\text{h}$	$s_2 = 45\text{h}$

Da die Stichproben gross genug sind ( $> 30$ ), dürfen wir  $\sigma_1^2 = s_1^2$  und  $\sigma_2^2 = s_2^2$  annehmen.

1. Nullhypothese:
2. Alternativhypothese:
3. Welche Zeile der Tabelle 8.2.1 passt zu unserem Problem?
4. Die Testvariable folgt

Signifikanzniveau:

$$H_A$$

Kritische Grenzen:

5. Testwert berechnen:

6. Testentscheidung:

### Anwendung aus der Biomechanik: Testen eines dynamischen Stuhls (Bachelorarbeit)

In einer Projekt- und Bachelorarbeit wurde ein dynamischer Stuhl entwickelt. Dieser soll nun mit einem konventionellen, im Handel erhältlichen dynamischen Bürostuhl verglichen werden. Um nachzuweisen, dass der an der ZHAW entwickelte Stuhl besser ist, sitzen Versuchspersonen auf den beiden Stühlen Probe. Dabei werden einerseits Messungen im Bewegungslabor durchgeführt, andererseits geben die Versuchspersonen mittels Fragebogen über ihre subjektive Einschätzung des Sitzkomforts Auskunft (Noten von 1 bis 7).

(a) Welcher Hypothesentest muss hier verwendet werden?

.....

(b) Soll eine einseitige oder eine zweiseitige Hypothese getestet werden?

.....

Bei diesem Beispiel (wie auch bei vielen anderen Anwendungen aus der Biomechanik) ist die Erhebung der Daten ziemlich aufwändig. Deshalb möchte man mit einer möglichst kleinen Stichprobe arbeiten. Um zu bestimmen, wie gross die Stichprobe sein sollte, geht man so vor:

1. Man überlegt sich, welche Abweichung von  $H_0$  relevant ist. Das ist keine mathematische Frage!
2. Man schätzt die Varianz, entweder auf Grund von Daten aus anderen ähnlichen Erhebungen oder auf Grund einer (kleinen) Vorlaufstichprobe.
3. Man wählt die Stichprobengrösse  $n$  so, dass die kritische Grenze  $c_o$  kleiner ist als der Testwert, der einer relevanten Abweichung vom Parameterwert unter  $H_0$  entspricht.

Für unser Beispiel der subjektiven Einschätzung des Sitzkomforts sieht das so aus ( $\alpha = 5\%$ ):

1. Wir nehmen an, dass eine Abweichung von 1 Notenpunkt relevant ist.
2. Eine Vorlaufstichprobe mit 8 Versuchspersonen ergibt für die Differenzen zwischen den beiden Stühlen eine Varianz von  $\sigma^2 = 3.125$ .
3. Der Mittelwert  $\bar{z}$  der Differenzen befindet sich im kritischen Bereich, falls gilt:

In unserem Fall soll also  $n$  so gewählt werden, dass eine Abweichung von 1 Notenpunkt von  $\mu_0 = 0$  sicher im kritischen Bereich liegt, d.h. es soll gelten:

Da  $c_o$  auch von  $n$  abhängt, probieren wir ein paar Werte für  $n$  aus:

### 8.3 Mögliche Fehlerquellen bei einem Hypothesentest

Am Ende eines Hypothesentests wird eine Entscheidung gefällt: Entweder wird  $H_0$  angenommen oder abgelehnt. Die Entscheidung beruht auf der Auswertung einer Zufallsstichprobe. Auf dieser Grundlage kann aber nie eine absolut sichere Entscheidung gefällt werden; es besteht immer die Möglichkeit eines Irrtums.

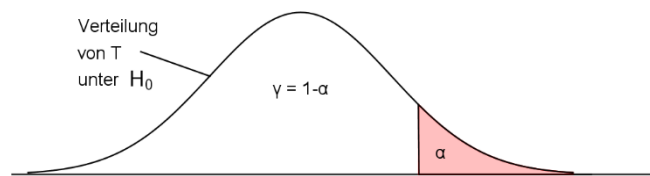
Es gibt zwei Arten von Fehlern:

Testentscheidung \ Realität	$H_0$ wird angenommen	$H_0$ wird abgelehnt
$H_0$ ist wahr		
$H_0$ ist falsch		

**Fehler 1. Art:  $H_0$  ist wahr und wird trotzdem abgelehnt.**

„Der Alarm geht los, obwohl es nicht brennt.“

Die Wahrscheinlichkeit für einen Fehler 1. Art wird mit  $\alpha$  bezeichnet.



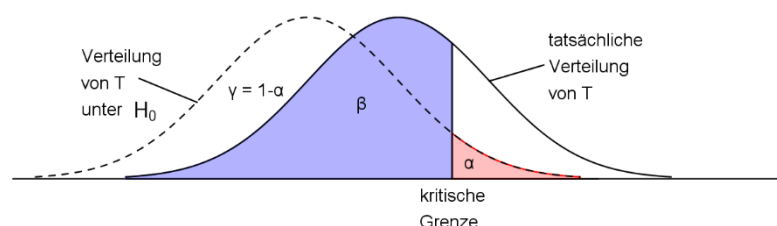
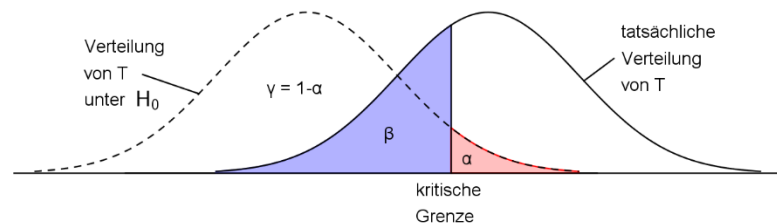
(Alle Illustrationen in diesem Abschnitt für 1-seitige Alternativhypothesen mit kritischem Bereich auf der rechten Seite)

**Fehler 2. Art:  $H_0$  ist falsch und wird trotzdem angenommen.**

„Der Alarm geht nicht los, obwohl es brennt.“

Die Wahrscheinlichkeit für einen Fehler 2. Art wird mit  $\beta$  bezeichnet.

Die Grösse von  $\beta$  ist abhängig vom tatsächlichen (unbekannten) Wert des Parameters:



Je näher der wahre Wert des Parameters am angenommenen Wert des Parameters ist, desto grösser ist die Wahrscheinlichkeit für einen Fehler 2. Art.

### Beispiel 8

Bei der Anlieferung von Ware wird eine Abnahmekontrolle durchgeführt.

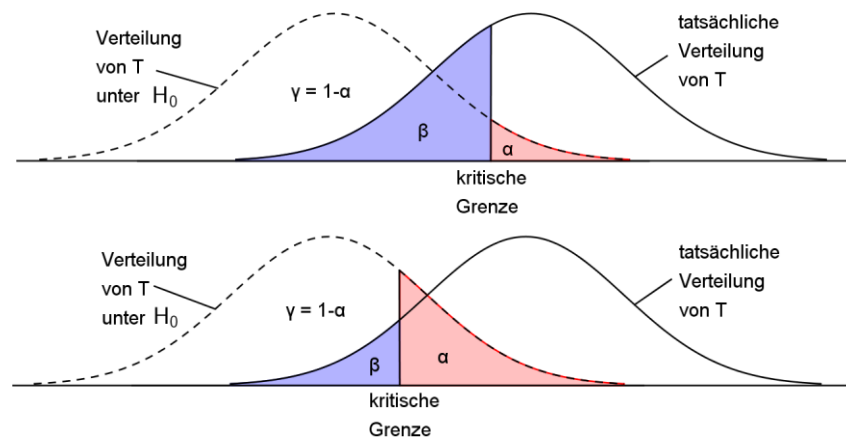
*Fehler 1. Art: ‚Produzentenrisiko‘*

Auf Grund der Stichprobe wird die Lieferung zurückgewiesen, obwohl der vereinbarte Ausschussanteil in der Lieferung nicht überschritten wurde.

*Fehler 2. Art: ‚Konsumentenrisiko‘*

Auf Grund der Stichprobe wird die Lieferung angenommen, obwohl der vereinbarte Ausschussanteil in der Lieferung überschritten wurde.

### Zusammenhang zwischen Fehlern 1. und 2. Art:

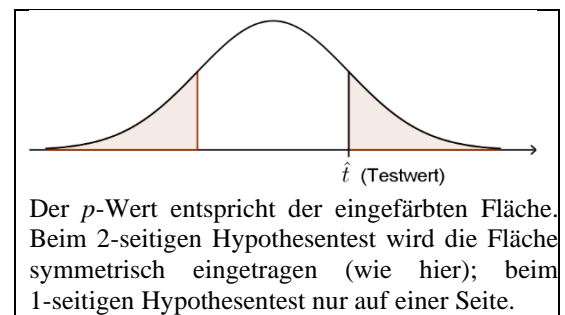


Eine *Verkleinerung* der Wahrscheinlichkeit für einen Fehler 1. Art ( $\alpha$ ) bewirkt bei gleicher Stichprobengrösse stets eine *Vergrößerung* der Wahrscheinlichkeit für einen Fehler 2. Art ( $\beta$ ) und umgekehrt.

## 8.4 Allgemeine Bemerkungen zu Hypothesentests

### (1) Der $p$ -Wert

Statt einer Ja/Nein-Entscheidung auf Grund einer vorgegebenen Irrtumswahrscheinlichkeit  $\alpha$  kann man auch den sogenannten  $p$ -Wert angeben: die Wahrscheinlichkeit, einen mindestens so extremen Testwert zu erhalten, wenn  $H_0$  wahr ist. Ist der  $p$ -Wert grösser oder gleich  $\alpha$ , wird die Nullhypothese angenommen. Ist der  $p$ -Wert kleiner als  $\alpha$ , wird die Alternative angenommen.



### Beispiel 9

In Produktionsfirmen wird vor der Auslieferung einer bestimmten Warenpartie deren Qualität auf Stichprobenbasis kontrolliert. Der Vertrag zwischen Produzenten und Kunde beinhaltet, dass der Kunde die Lieferung akzeptiert, sofern der Ausschussanteil unter 10% liegt mit einer Signifikanz von  $\alpha = 5\%$ . Aus Zeit- und Kostengründen können nicht alle Produkte getestet

werden, so dass der Ausschussanteil auf Basis einer Stichprobe im Umfang von  $n = 1000$  Stück geschätzt wird. In dieser Stichprobe befinden sich 102 Ausschussstücke. Damit ergäbe sich als Schätzer für den wahren, aber unbekannten Ausschussanteil in dieser Lieferung:

$$\hat{p} = \frac{102}{1000} = 0.102 = 10.2\%$$

Für die zugehörige standardisierte Zufallsvariable erhalten wir:

$$u = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} =$$

Wir bestimmen nun den zugehörigen  $p$ -Wert für den Testwert  $\hat{u} = 0.21$  mit  $\alpha = 0.05$  für einen einseitigen Test:

$p$ -Wert =

Die Nullhypothese wird angenommen, da der  $p$ -Wert grösser oder gleich  $\alpha = 0.05$  ist.

$p$ -Werte liefern dabei mehr Informationen als die Ja-Nein-Entscheidung bzgl. der Annahme bzw. Ablehnung der Nullhypothese: Man kann an ihnen ablesen, zu welchem Niveau der zugehörige Test die Nullhypothese gerade noch verworfen hätte. Diese letzte Interpretation birgt jedoch die Gefahr eines Missbrauchs insofern, als zunächst der Test durchgeführt, also der  $p$ -Wert berechnet werden kann und dann das Signifikanzniveau festgelegt wird und zwar gerade so, dass man  $H_0$  noch annehmen oder ablehnen kann.

## (2) ‚Signifikant‘ heisst ‚nicht zufallsbedingt‘

Wird bei einem Hypothesentest  $H_0$  abgelehnt, so sagt man: Die Abweichung von  $H_0$  ist *signifikant*, d.h. man kann davon ausgehen, dass sie nicht zufallsbedingt ist. Das heisst aber noch lange nicht, dass der Unterschied von Bedeutung ist! (engl. ‚significant‘ = bedeutsam).

### Beispiel 10

2500 Schüler aus der Stadt:	Durchschnitt von 26 P. in einem Test
2500 Schüler vom Lande:	Durchschnitt von 25 P. im selben Test

Der  $p$ -Wert für die Nullhypothese, dass tatsächlich ein Unterschied besteht, beträgt  $5/10^4$ . Nur: Ist dieser Unterschied relevant!?

Bei einer grossen Stichprobe kann schon ein kleiner Unterschied signifikant sein.

## (3) Unterschiede werden nicht erklärt

Ein Test stellt nur fest, dass ein Unterschied nicht zufällig ist; er erklärt den Unterschied nicht.

### Beispiel 11

Jemand behauptet, mit parapsychologischen Kräften einen Würfel beeinflussen zu können. Tatsächlich würfelt er in einem Experiment so viele Sechsen, dass man mit einem Hypothesentest nachweisen kann, dass das wohl nicht zufällig ist. Als er aber Einsen würfeln soll, gibt es immer noch zu viele Sechsen – der Sechserüberschuss war also tatsächlich nicht zufällig, aber basiert keineswegs auf parapsychologischen Fähigkeiten, sondern auf einem gefälschten Würfel!

Der Test interpretiert das Resultat nicht. Er kontrolliert auch nicht das Design der Studie.

#### **(4) Hypothesentests basieren auf Zufallsstichproben**

All diese Methoden basieren auf der Annahme, dass wir eine Zufallsstichprobe betrachten, und nicht eine ‚praktische‘ Stichprobe wie ‚die ersten hundert Leute, die mir begegnen und bereit sind, meinen Fragebogen auszufüllen‘.

#### **(5) Gegenüberstellung Hypothesentest – Vertrauensintervall**

Beim *Hypothesentest* ist der behauptete Wert des Parameters der Ausgangspunkt für die Bestimmung des Annahmebereichs; soll ein *Vertrauensintervall* bestimmt werden, so geht man vom Schätzwert für den betrachteten Parameter aus.

## 8.5 Lernziele für dieses Kapitel

- ☐ Ich kann an Beispielen erklären, was ein *Hypothesentest* ist.
- ☐ Ich kann *Nullhypothesen* und *Alternativhypothesen* aufstellen.
- ☐ Ich kann Tabelle 8.2.1 einsetzen, um Parametertests zu definieren.
- ☐ Ich kann die Quantiltabellen der Standardnormalverteilung, der Student-t-Verteilung und der Chi Quadrat Verteilung anwenden, um die kritischen Grenzen bei den zugehörigen Hypothesentests zu berechnen.
- ☐ Ich kann die Entscheidung begründen, weswegen Nullhypothesen angenommen oder abgelehnt werden.
- ☐ Ich kann Hypothesentests für die Gleichheit von unbekannten Mittelwerten aufstellen.
- ☐ Ich kann die Begriffe *abhängige* und *unabhängige* Stichproben an Beispielen erklären.
- ☐ Ich kann den  $p$ -Wert einer Nullhypothese bestimmen.
- ☐ Ich kann aufgrund des  $p$ -Wertes entscheiden, ob die Nullhypothese anzunehmen oder abzulehnen ist.