

# **Statistik I — bis jetzt**

## **K 1 Einführung**

## **K 2 Beschreibende Statistik**

## **K 3 Graphiken**

## **K 4 Schätzen**

und jetzt

## **K 5 Testen**

5.1 Tests im allgemeinen

5.2 Binomial Tests

## K5 Statistische Tests

### Beispiel: Exemestane und Brustkrebs

Tamoxifen über 5 Jahre ist für gewisse Frauen mit Brustkrebs eine Standardbehandlung. Wissenschaftler haben eine neue Behandlung vorgeschlagen:

2 bis 3 Jahre Tamoxifen und danach Exemestane.

In einer Doppelblindstudie von Coombes et al sind folgende Resultate berichtet worden (NEJM 350(11) 1081-92).

4742 Patientinnen insgesamt

2362 bekamen nach 2 bis 3 Jahren Exemestane

2380 bekamen noch Tamoxifen

Tote

93 in der Exemestane Gruppe                    %

106 in der Tamoxifen Gruppe                    %

Ereignisse

266 in der Exemestane Gruppe                    %

449 in der Tamoxifen Gruppe                    %

Zeigen diese Resultate, dass die Behandlung mit Exemestane besser ist?

## 5.1 Statistische Tests im allgemeinen

z.B. Hilft Vitamin C gegen Erkältungen?

Ist der neue Ferrari schneller als der alte?

Sind finnische Schüler besser als deutsche?

- Was ist ein statistischer Test?
- Warum werden statistische Tests durchgeführt?
- Wann werden statistische Tests durchgeführt?
- Wie werden statistische Tests durchgeführt?
- Wer führt statistische Tests durch?
- Welche Folgen haben statistische Tests?

# Terminologie für Test-Theorie

Hypothese

$H_0$  Nullhypothese

$H_1$  Alternativhypothese

Teststatistik (Prüfgröße)

Verteilung der Teststatistik

Testniveau  $\alpha$

Annahmereich/Ablehnungsbereich

Entscheidungsregel (kritischer Wert)

Statistische Signifikanz

$p$ -Wert

Fehler Typ I  $P(H_0 \text{ verwerfen} | H_0 \text{ wahr})$

Fehler Typ II  $P(H_0 \text{ akzeptieren} | H_1 \text{ wahr})$

Gütefunktion (beim Test eines Parameters)

$g(\theta) = P(H_0 \text{ verwerfen} | \theta \text{ der wahre Parameter})$

Operationscharakteristik eines Tests  $OC = 1 - g(\theta)$

# **Statistische Testprozeduren**

## **Theorie**

Nullhypothese  $H_0$

## **Studie entwerfen**

Definition der Grundgesamtheit

Stichprobenverfahren

Stichprobengröße

## **Daten holen**

Daten erheben

## **Daten organisieren**

Häufigkeitstabellen, Histogramme usw.

## **Resultate zusammenfassen**

Teststatistik berechnen

## **Teststatistik mit erwarteten Werten vergleichen**

Verteilung der Teststatistik

$p$ -Wert, Signifikanz

## Signifikanz der Teststatistik

Wir berechnen den  $p$ -Wert

$$P_{H_0}(\text{Teststatistik} > \text{beobachtete Wert der Statistik})$$

(Im allgemeinen sucht man die Wahrscheinlichkeit des Wertebereichs, worin alle Werte eine kleinere Wahrscheinlichkeit als der beobachtete Wert haben.)

Traditionelle Deutung

(nach Fisher nur als Richtlinie, aber ...):

$p$ -Wert  $> 0.05 \Rightarrow H_0$  sollte nicht verworfen werden

$p$ -Wert  $< 0.05 \Rightarrow H_0$  kann vielleicht verworfen werden

$p$ -Wert  $< 0.01 \Rightarrow H_0$  kann verworfen werden

Vieles hängt von den nachfolgenden Wirkungen ab, wie man  $p$ -Werte interpretiert.

z.B.

neues Medikament genehmigt

oder

alte Behandlung verboten

## 5.2 Binomial Test (eine Stichprobe)

### Nullhypothese

$$p = p_0$$

### Teststatistik

Anzahl Erfolge  $R$  aus  $n$

### Verteilung unter $H_0$

$$P(R = r) = \binom{n}{r} p_0^r (1 - p_0)^{n-r}$$

### Annahmebereich

$$[a, b] \text{ um } np_0$$

### Ablehnungsbereich

$$[0, a - 1] \cup [b + 1, n]$$

### Typ I Fehler (Test-Niveau)

$$P(R \notin [a, b] | H_0)$$

Das Test-Niveau wird für einen diskreten Test selten genau  $\alpha$  sein. Eine Möglichkeit ist ein randomisierter Test: falls  $R = a$  oder  $R = b$ , akzeptiert man  $H_0$  mit Wahrscheinlichkeiten, die so gewählt sind, dass das Test-Niveau  $= \alpha$ . Das wird mathematisch genau sein, ist aber in der Praxis unzufriedenstellend.

### **Alternativhypothese**

$$p = p_1$$

### **Typ II Fehler**

$$P(R \in [a, b] | H_1)$$

### **Gütefunktion**

$$g(p_1) = P(R \in [a, b] | p = p_1)$$

**Binomial Test Beispiel**  
**Außersinnliche Wahrnehmung**  
**Farbe einer Karte erraten**  
(cf. u.a. Jung und Pauli)

**Nullhypothese**  $p = 0.5$

**Teststatistik** Anzahl Erfolge  $R$  aus 6

**Verteilung unter  $H_0$**   
$$P(R = r) = \binom{6}{r} 0.5^r (1 - 0.5)^{6-r}$$

**Annahmebereich**  $[1, 5]$

**Ablehnungsbereich**  $[0] \cup [6]$

**Typ I Fehler (Test-Niveau)**

$$P(R \notin [1, 5] | H_0) = 0.03125$$

**$p$ -Wert von  $R = 5$**   $P(R \notin [2, 4] | H_0) = 0.2188$

**Alternativhypothese**  $p = p_1 = 0.8$

**Typ II Fehler**  $P(R \in [1, 5] | H_1) = 0.74$

**Gütefunktion**  $g(p_1) = P(R \in [1, 5] | p = p_1)$

## Binomial Test (zwei Stichproben)

**Nullhypothese**  $p_1 = p_2$

**Teststatistik**  $X/n_1 - Y/n_2$

**Verteilung unter  $H_0$**

$$E[T] = 0, V[T] = p(1 - p)(1/n_1 + 1/n_2)$$

**Annahmebereich** ?

**Ablehnungsbereich** ?

**Typ I Fehler (Test-Niveau)** ?

**Alternativhypothese**  $p_1 \neq p_2$

**Typ II Fehler** ?

**Gütefunktion** ?

Oder wir nehmen an, dass  $n_1, n_2$  groß genug sind, dass die Normalapproximation gilt. Dann haben wir:

$$X/n_1 \sim N(p, p(1-p)/n_1) \text{ und } Y/n_2 \sim N(p, p(1-p)/n_2)$$

$$\Rightarrow T = X/n_1 - Y/n_2 \sim N(0, p(1 - p)(1/n_1 + 1/n_2))$$