

Eine Einführung in R: Statistische Tests

Bernd Klaus, Verena Zuber

Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE),
Universität Leipzig

25. November 2010

I. Einführungsbeispiel

II. Theorie: Statistische Tests

Hypothesen aufstellen

Betrachtung der Daten

Aufstellen der Prüfgröße

Durchführen des Tests

Testentscheidung

III. Zwei Klassiker: t -Test und Wilcoxon-Rangsummen - Test

t -Test - EW einer Population

t -Test - EW zweier Populationen

t -Test - EW gepaarter Populationen

Der Wilcoxon-Rangsummen - Test

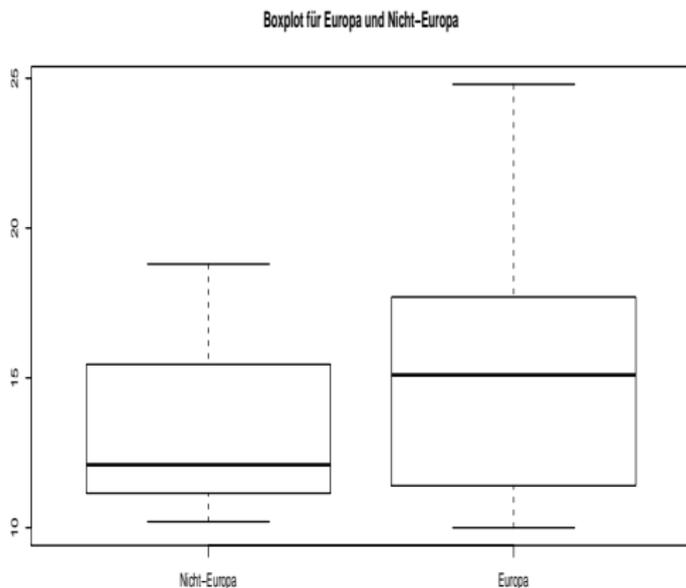
IV. t -Test und Wilcoxon-Rangsummen - Test in R

I. Einführungsbeispiel: Trinkt die Jugend in Europa mehr Alkohol als im Rest der Welt?

Untersucht wird die Variable *Alkohol* im oecd-Datensatz:
Der Anteil an 13-15 jährigen Jugendlichen, die mindestens zweimal betrunken waren

Erster Schritt: Deskriptive Analyse

1. Graphisch mit dem Boxplot
`boxplot(Alkohol~Geo)`



2. Kennzahlen, wie

- ▶ Mittelwert

```
mu<-tapply(Alkohol, Geo, FUN=mean, na.rm=TRUE)
  Nicht-Europa  Europa
  13.700       15.443
```

- ▶ Standardabweichung

```
sigma<-tapply(Alkohol, Geo, FUN=sd, na.rm=TRUE)
  Nicht-Europa  Europa
  4.518        4.341
```

Es ist zu erkennen, dass in Europa im Mittel ein höherer Anteil an Jugendlichen schon mindestens zweimal betrunken war als in nicht-europäischen Staaten.

Doch dies könnte auch ein Zufall sein!
Denn die Beobachtungen beruhen auf Stichproben,
sie sind Realisierungen einer Zufallsvariable.

Eigentliches Ziel:

Überprüfung von Annahmen über das Verhalten des interessierenden Merkmals in der Grundgesamtheit mittels Stichproben.

- ▶ Annahme: Jugendliche in Europa trinken mehr Alkohol als im Rest der Welt
- ▶ Merkmal: Alkoholkonsum der Jugend
- ▶ Grundgesamtheit: Jugendliche in Europa und im Rest der Welt
- ▶ Stichprobe: Die *oecd*-Daten

Für solche Fragestellungen mit gleichzeitiger Kontrolle der Fehlerwahrscheinlichkeit sind statistische Tests geeignet!

Statistisches Testen I

1. Aufstellen von zwei komplementären Hypothesen:
 - ▶ Testhypothese (H_0): Der Anteil in Europa ist kleiner dem im Rest der Welt $\mu_E \leq \mu_{NE}$
 - ▶ Alternativhypothese (H_1): Der Anteil in Europa größer als der im Rest der Welt $\mu_E > \mu_{NE}$
2. Fehlerwahrscheinlichkeit festlegen:
 H_0 soll mit einer W'keit von weniger als 5% abgelehnt werden, wenn H_0 wahr ist.

Also: wenn der Anteil in Wahrheit kleiner oder gleich ist, soll der Test nur mit einer W'keit von weniger als 5% zu dem (falschen) Ergebnis kommen, dass der Anteil größer ist.

Statistisches Testen II

3. Beobachtete Daten: 2 Gruppen

	$\hat{\mu}$	$\hat{\sigma}$	n
Nicht-Europa	13.700	4.518	3
Europa	15.443	4.341	21

4. (Weitere Annahmen: Normalverteilung, Varianzgleichheit)
5. Berechnen der Prüfgröße T , einer Kennzahl, die zeigt, wie stark die Gruppenmittel voneinander abweichen:

- (a) Mittelwertsdifferenz der beiden Gruppen
(b) Standardisieren mit der entsprechenden Standardabweichung

$$T = (\hat{\mu}_E - \hat{\mu}_{NE}) / \sqrt{\left(\frac{1}{n_E} + \frac{1}{n_{NE}}\right) \frac{(n_E - 1)\hat{\sigma}_E^2 + (n_{NE} - 1)\hat{\sigma}_{NE}^2}{n_E + n_{NE} - 2}}$$

- (c) (Hypothetische Verteilung der Prüfgröße festlegen, hier t-Verteilung mit $3 + 21 - 2 = 22$ Freiheitsgraden)

Statistisches Testen III

6. Berechnung der Prüfgröße T in R:

(a) Mittelwertsdifferenz der beiden Gruppen

```
m.diff<-mu[2]-mu[1]
```

(b) Standardisieren mit der entsprechenden Standardabweichung

```
diff.std2 <- sqrt((1/21+1/3)*  
(20*sigma[2]^2+2*sigma[1]^2)/(21+3-2))
```

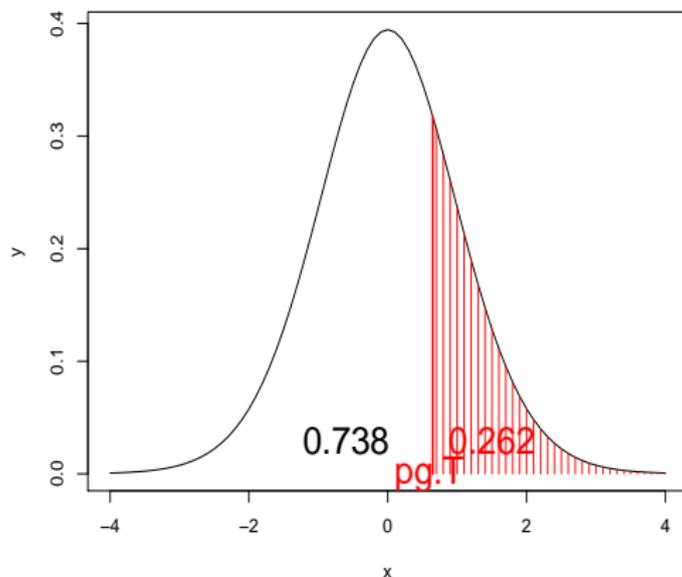
(c) Prüfgröße:

```
pg.T <-m.diff/diff.std  
0.648
```

7. Wie wahrscheinlich ist es (unter der Nullhypothese), eine Prüfgröße T zu beobachten, die größer oder gleich 0.648 ist?

```
1-pt(pg.T, df=22)  
0.262
```

Statistisches Testen IV



Mit hoher W'keit (26.2%) kann eine solche Prüfgröße $pg.T$ beobachtet werden, wenn der Mittelwert in Europa und kleiner als der in Nicht-Europa ist.

Statistisches Testen V

8. Entscheidung: Aus diesen Daten kann nicht geschlossen werden, dass in Europa Jugendliche mehr Alkohol trinken als im Rest der Welt.

9. Grund: Zu geringe Fallzahl!

Mit $n_E = n_{NE} = 101$ ergibt sich

(b) Standardisieren mit der entsprechenden Standardabweichung

```
diff.std <- sqrt((1/101+1/101)*  
(100*sigma[2]^2+100*sigma[1]^2)/(101+101-2))
```

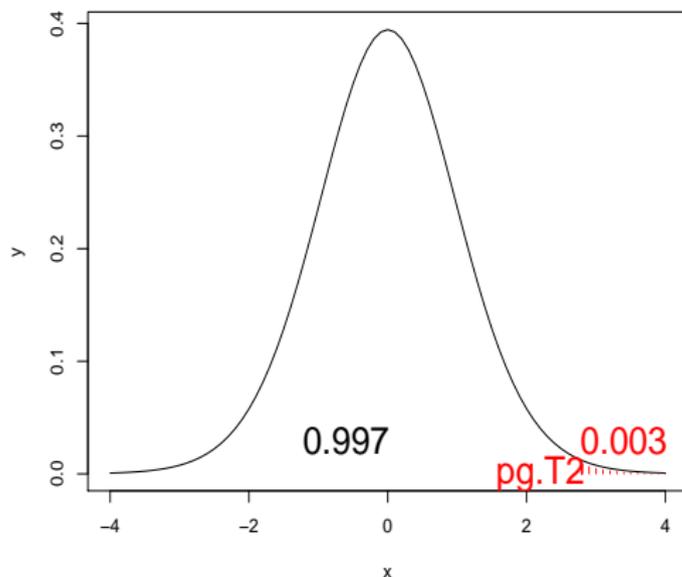
(c) Prüfgröße:

```
pg.T2 <- m.diff/diff.std2  
2.796
```

(d) Vergleich mit der t -Verteilung:

```
1-pt(pg.T2, df=200)  
0.003
```

Statistisches Testen IV



Mit nur sehr geringer W'keit (0.003%) kann eine solche Prüfgröße $pg.T2$ beobachtet werden, wenn wenn der Mittelwert in Europa und kleiner als der in Nicht-Europa ist.

II. Der Baukasten für statistische Test: Wie geht man vor?

Fünf Schritte zum Testergebniss

- I. Hypothesen aufstellen
- II. Betrachtung der Daten
- III. Aufstellen der Prüfgröße
- IV. Durchführen des Tests
- V. Testentscheidung

I. Hypothesen aufstellen

- ▶ Was soll verglichen werden?
 - ▶ Gegen einen festen Wert
 - ▶ Zwei Gruppen (t -Test)
 - ▶ Messwiederholungen
- ▶ Einseitige oder zweiseitige Fragestellung? Beispiel:
 - ▶ Einseitige Fragestellung :
 $H_0 : \mu_1 \leq \mu_2$ gegen $H_1 : \mu_1 > \mu_2$
 - ▶ Zweiseitige Fragestellung :
 $H_0 : \mu_1 = \mu_2$ gegen $H_1 : \mu_1 \neq \mu_2$
- ▶ Aufstellen der eigentlich interessierenden Alternativhypothese H_1 und der Nullhypothese H_0 .
- ▶ Signifikanzniveau α festlegen.

Welche Fehler kann man beim Testen machen?

	Entscheidung: H_0	Entscheidung: H_1
H_0 wahr	richtig	Fehler 1. Art (α)
H_1 wahr	Fehler 2. Art (β)	richtig

- ▶ Fehler erster Art (α -Fehler):
Obwohl H_0 wahr ist, entscheidet man sich für H_1
(*False Positiv*)
- ▶ Fehler zweiter Art (β -Fehler):
Obwohl H_1 wahr ist, entscheidet man sich für H_0
(*False Negativ*)

II. Betrachtung der Daten

- ▶ Können Verteilungsannahmen getroffen werden?
 - ▶ Ja: Parametrische Tests
 - ▶ Nein: Nonparametrische Tests
- ▶ Weitere Annahmen wie z.B. Varianzgleichheit in den Gruppen...

Aus Schritt I. und II. folgen alle weiteren Schritte!

III. Aufstellen der Prüfgröße

- ▶ Aus den Hypothesen ergibt sich die Form der Prüfgröße, z.B. die Mittelwertsdifferenz
- ▶ Standardisieren mit
 - ▶ unter H_0 gültigen Erwartungswert
 - ▶ unter H_0 gültigen Standardabweichung
- ▶ Festlegen der Verteilung, die unter H_0 gültig ist.

IV. Durchführen des Tests und V. Testentscheidung

Hier sind zwei Werte entscheidend:

- ▶ Kritischer Wert κ : Welchen Wert darf die Prüfgröße maximal annehmen, wenn H_0 tatsächlich gültig ist.
- ▶ p -Wert: Wahrscheinlichkeit, die vorliegenden Daten zu beobachten, wenn H_0 gültig ist.

Entscheidung: H_0 ablehnen, falls

- ▶ die Prüfgröße größer als der kritische Wert ist (Vorsicht bei nonparametrischen Tests: hier kleiner als der kritische Wert).
- ▶ falls der p -Wert kleiner dem vorher festgelegten Signifikanzniveau α ist.

III. Zwei Klassiker t -Test und Wilcoxon-Rangsummen - Test ... ein bisschen Theorie

1. Ziel, Hypothesen und Voraussetzungen

- ▶ Vergleich das emp. Populationsmittel \bar{x} mit einem hypothetischen Mittelwert μ_0 .
- ▶ benötigt eine Normalverteilung der Stichprobe
- ▶ Varianz wird als unbekannt angenommen

Varianten für die Hypothesen:

- (a) Einseitige Fragestellung 1 :
 $H_0 : \bar{x} \leq \mu_0$ gegen $H_1 : \bar{x} > \mu_0$
- (b) Einseitige Fragestellung 2 :
 $H_0 : \bar{x} \geq \mu_0$ gegen $H_1 : \bar{x} < \mu_0$
- (c) Zweiseitige Fragestellung :
 $H_0 : \bar{x} = \mu_0$ gegen $H_1 : \bar{x} \neq \mu_0$

2. Teststatistik

► Teststatistik

$$T = \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n}$$

► Schätzung der Standardabweichung σ durch:

$$s = \left[\frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n - 1} \right]^{0.5}$$

3. Kritische Bereiche

Kritische Bereiche:

(a) Einseitige Fragestellung 1 :

$$T > t_{1-\alpha}(n-1)$$

(b) Einseitige Fragestellung 2 :

$$T < t_{\alpha}(n-1)$$

(c) Zweiseitige Fragestellung :

$$|T| > t_{1-\alpha/2}(n-1)$$

1. Ziel, Hypothesen und Voraussetzungen

- ▶ Vergleiche die emp. Populationsmittel \bar{x}_1 und \bar{x}_2 miteinander
- ▶ benötigt eine Normalverteilung der Stichprobe
- ▶ Varianz der Populationen unbekannt
- ▶ 2 Varianten: Varianzen der Populationen gleich oder ungleich

Varianten für die Hypothesen:

- (a) Einseitige Fragestellung 1 :
 $H_0 : \bar{x}_1 \leq \bar{x}_2$ gegen $H_1 : \bar{x}_1 > \bar{x}_2$
- (b) Einseitige Fragestellung 2 :
 $H_0 : \bar{x}_1 \geq \bar{x}_2$ gegen $H_1 : \bar{x}_1 < \bar{x}_2$
- (c) Zweiseitige Fragestellung :
 $H_0 : \bar{x}_1 = \bar{x}_2$ gegen $H_1 : \bar{x}_1 \neq \bar{x}_2$

2. Teststatistik

- ▶ Teststatistik

$$T = \frac{\bar{x}_1 - \bar{x}_2}{s} \cdot \sqrt{n}$$

- ▶ Schätzung der Standardabweichung σ durch:

$$s = \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \cdot \frac{(n_1 - 1)s_1 + (n_2 - 1)s_2}{n_1 + n_2 - 1} \right]^{0.5}$$

wobei s_1 und s_2 die Standardvarianzschätzer für die Populationen sind

3. Kritische Bereiche

Kritische Bereiche:

(a) Einseitige Fragestellung 1 :

$$T > t_{1-\alpha}(n_1 + n_2 - 2)$$

(b) Einseitige Fragestellung 2 :

$$T < t_{\alpha}(n_1 + n_2 - 2)$$

(c) Zweiseitige Fragestellung :

$$|T| > t_{1-\alpha/2}(n_1 + n_2 - 2)$$

1. Ziel, Hypothesen und Voraussetzungen

- ▶ Teste die Differenz $\bar{d} := \sum_{i=1}^n d_i$ miteinander gepaarter Stichproben (x_{1i}, x_{2i})
- ▶ typisches Bsp.: Messen eines Blutwertes vor und nach einer med. Behandlung
- ▶ benötigt eine Normalverteilung der Stichprobe

Varianten für die Hypothesen:

- (a) Einseitige Fragestellung 1 :
 $H_0 : d \leq 0$ gegen $H_1 : d > 0$
- (b) Einseitige Fragestellung 2 :
 $H_0 : d \geq 0$ gegen $H_1 : d < 0$
- (c) Zweiseitige Fragestellung :
 $H_0 : d = 0$ gegen $H_1 : d \neq 0$

2. Teststatistik

► Teststatistik

$$T = \frac{\bar{d}}{s} \cdot \sqrt{n}$$

► Schätzung der Standardabweichung σ durch:

$$s = \left[\frac{\sum_{i=1}^n (\bar{d} - d_i)^2}{n - 1} \right]^{0.5}$$

3. Kritische Bereiche

Kritische Bereiche:

(a) Einseitige Fragestellung 1 :

$$T > t_{1-\alpha}(n-1)$$

(b) Einseitige Fragestellung 2 :

$$T < t_{\alpha}(n-1)$$

(c) Zweiseitige Fragestellung :

$$|T| > t_{1-\alpha/2}(n-1)$$

1. Ziel, Hypothesen und Voraussetzungen

- ▶ Teste nicht-parametrisch, ob zwei Population den gleichen Median besitzen.
- ▶ Zu verwenden, wenn Vor. für den t -Test nicht erfüllt sind
- ▶ benötigt KEINE konkrete Verteilungsannahme
- ▶ “ t -Test-Ersatz”

Varianten für die Hypothesen:

(a) Einseitige Fragestellung 1 :

$$H_0 : x_{1,\text{med}} \leq x_{2,\text{med}} \text{ gegen } H_1 : x_{1,\text{med}} > x_{2,\text{med}}$$

(b) Einseitige Fragestellung 2 :

$$H_0 : x_{1,\text{med}} \geq x_{2,\text{med}} \text{ gegen } H_1 : x_{1,\text{med}} < x_{2,\text{med}}$$

(c) Zweiseitige Fragestellung :

$$H_0 : x_{1,\text{med}} = x_{2,\text{med}} \text{ gegen } H_1 : x_{1,\text{med}} \neq x_{2,\text{med}}$$

2. Teststatistik

- ▶ Bilde für sämtlichen Beobachtungen $x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}$
Ränge $rg(x_{11}), \dots, rg(x_{1n_1}), rg(x_{21}), \dots, rg(x_{2n_2})$
- ▶ Teststatistik

$$R = \sum_{i=1}^{n_1} rg(x_{1i})$$

- ▶ Wertebereich: $\frac{n_1(n_1+1)}{2} < R < \frac{(n_1+n_2)(n_1+n_2+1)}{2} - \frac{n_1(n_1+1)}{2}$
- ▶ Nullverteilung von R liegt tabelliert vor
- ▶ Approximation durch die Normalverteilung ab einer Stichprobengröße von ca. 20 möglich

3. Kritische Bereiche

Kritische Bereiche:

(a) Einseitige Fragestellung 1 :

$$R > w_{1-\alpha}(n_1, n_2)$$

(b) Einseitige Fragestellung 2 :

$$R < w_{\alpha}(n_1, n_2)$$

(c) Zweiseitige Fragestellung :

$$R > w_{1-\alpha/2}(n_1, n_2) \text{ oder } R < w_{\alpha/2}(n_1, n_2)$$

IV. t -Test und Wilcoxon-Rangsummen - Test in R - Praktische Durchführung

t-Test in R

```
t.test(x, y, alternative, paired, var.equal)
```

Erklärung der Parameter

- ▶ `x, y = NULL`: Die Daten, beim t-Test für eine Population genügt es, `x` anzugeben.
- ▶ `alternative = c("two.sided", "less", "greater")`: Varianten für die Alternativhypothese
- ▶ `var.equal = TRUE`: Gibt an, ob Varianzgleichheit bei den Populationen vorliegt
- ▶ `paired`: Gibt an, ob `x` und `y` als gepaarte Stichprobe anzusehen sind

Wilcoxon-Rangsummen - Test in R

```
wilcox.test(x, y, alternative, paired, exact)
```

Erklärung der Parameter

- ▶ Parameter fast wie beim t -Test ...
- ▶ **exact** : Soll die Teststatistik exakt bestimmt werden, oder per Approximation an die Normalverteilung?

Beispiel

- ▶ Nettokaltmieten pro m^2 für 1 (X) und 2-Raum (Y) Wohnungen
- ▶ Gibt es einen Unterschied zwischen beiden Gruppen?
- ▶ Wir untersuchen diese Frage per Wilcoxon- und t-Test.

	1	2	3	4	5
X	8.70	11.28	13.24	8.37	12.16
Y	3.36	18.35	5.19	8.35	13.10

	6	7	8	9	10
X	11.04	10.47	11.16	4.28	19.54
Y	15.65	4.29	11.36	9.09	

t-Test

```
miete <- read.csv("Miete.csv")  
attach(miete)  
t.test(X,Y, var.equal = FALSE, paired = FALSE)
```

R-Ausgabe:

```
Welch Two Sample t-test
```

```
data: X and Y  
t = 0.5471, df = 14.788, p-value = 0.5925  
alternative hypothesis: true difference in means is  
not equal to 0  
=>  $p > 0.05$ , kein signifikanter Unterschied
```

Wilcoxon-Rangsummen - Test in R

```
wilcox.test(X,Y, exact = TRUE)
```

R-Ausgabe:

```
Wilcoxon rank sum test
```

```
data: X and Y
```

```
W = 51, p-value = 0.6607
```

```
alternative hypothesis: true location shift is not  
equal to 0
```

```
=>  $p > 0.05$ , kein signifikanter Unterschied
```