

# Mathematik für Biologen

Prof. Dr. Rüdiger W. Braun

Heinrich-Heine-Universität Düsseldorf

13. Januar 2011

## 1 Nichtparametrische Tests

- Ränge
- Der U-Test
- Bindungen

# Ränge

- Zwei Gruppen von Zufallsvariablen mit unbekannten Verteilungen
- Die erste Gruppe hat Erwartungswert  $\mu_1$ , die zweite  $\mu_2$
- $\mu_1$  und  $\mu_2$  sollen verglichen werden
- Es gibt keine Verteilungsannahme, daher sind die absoluten Werte der Unterschiede wertlos
- Die einzige Information von Bedeutung ist, welches von zwei Elementen das größere ist
- Daher werden alle Ergebnisse der Größe nach angeordnet
- Das kleinste Element bekommt Rang 1, das zweitkleinste Rang 2 usw.
- Wenn  $\mu_1 < \mu_2$ , dann haben tendenziell die Elemente der ersten Gruppe kleinere Ränge als die der zweiten

# Testablauf des $U$ -Tests

- $R_1$  sei die Summe sämtlicher Ränge in der ersten Gruppe und  $R_2$  sei die Summe sämtlicher Ränge in der zweiten Gruppe
- Die Zufallsvariable  $U$  wird definiert als

$$U = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - R_1$$

$U$  gibt die Anzahl der Rangplatzüberschreitungen an

- Aus diesem  $U$  wird die Teststatistik berechnet

# Rangplatzüberschreitungen

- $U$  gibt die Anzahl der Paare  $(i, j)$  an, für welche der Rang des  $i$ -ten Elements der ersten Gruppe kleiner als der Rang des  $j$ -ten Elements der zweiten Gruppe ist
- Man kann sich überlegen, dass

$$U = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - R_1$$

# Rangplatzunterschreitungen

- Analog definiert man die Anzahl  $U'$  der Rangplatzunterschreitungen
- Es gilt

$$U + U' = n_1 \cdot n_2$$

- Wenn die Messwerte der ersten und der zweiten Gruppe derselben Verteilung genügen, dann im Mittel

$$U = U' = \frac{n_1 \cdot n_2}{2}$$

- Auf dieser Beobachtung beruht der  $U$ -Test

# Beispiel zur Berechnung der Rangplatzüberschreitungen

- Werte der ersten Gruppe: 11, 22, 14
- Werte der zweiten Gruppe: 12, 17, 20, 15

|               |    |    |    |    |    |    |    |
|---------------|----|----|----|----|----|----|----|
| erste Gruppe  | 11 |    | 14 |    |    |    | 22 |
| zweite Gruppe |    | 12 |    | 15 | 17 | 20 |    |
| Rang          | 1  | 2  | 3  | 4  | 5  | 6  | 7  |

$$n_1 = 3$$

$$n_2 = 4$$

$$U = 4 + 3 = 7$$

$$R_1 = 11$$

# U-Test von Mann-Whitney

- $X_1, \dots, X_{n_1}$  bezeichnen die Messwerte der Gruppe 1 und  $Y_1, \dots, Y_{n_2}$  die Messwerte der Gruppe 2
- Verteilungsvoraussetzungen:
  - alle  $X_j$  besitzen dieselbe Verteilung mit  $E(X_j) = \mu_1$
  - alle  $Y_j$  besitzen dieselbe Verteilung mit  $E(Y_j) = \mu_2$
  - die Verteilung der  $Y_j$  hat dieselbe Form wie die der  $X_j$ , ist aber möglicherweise verschoben
- Ziel:  $\mu_1$  und  $\mu_2$  sollen verglichen werden

# U-Test, Fortsetzung

- Für  $n_1 < 10$  oder  $n_2 < 10$  gibt es Tabellen der kritischen Werte des  $U$ -Tests
- Für  $n_1 \geq 10$  und  $n_2 \geq 10$  verwendet man die Quantile  $q_{1-\alpha}$  der Standardnormalverteilung
- Realisierungen (also Datenwerte) werden wieder mit Kleinbuchstaben geschrieben.
- Die Teststatistik ist

$$t = \frac{u - \mu}{\sigma}$$

wobei  $u$  die Anzahl der Rangplatzüberschreitungen ist und

$$\mu = \frac{n_1 \cdot n_2}{2}$$

$$\sigma = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}$$

# U-Test, Fortsetzung

- Das Signifikanzniveau sei  $\alpha$
- Entscheidung:

$H_0 = \{\mu_1 = \mu_2\}$ : Die Nullhypothese  $H_0$  wird abgelehnt, wenn  
 $|t| > q_{1-\alpha/2}$

$H_0 = \{\mu_1 \leq \mu_2\}$ : Die Nullhypothese  $H_0$  wird abgelehnt, wenn  
 $t < -q_{1-\alpha}$

$H_0 = \{\mu_1 \geq \mu_2\}$ : Die Nullhypothese  $H_0$  wird abgelehnt, wenn  
 $t > q_{1-\alpha}$

## Beispiel: Reaktionszeit unter Alkoholeinfluss

- Unter Alkoholeinfluss steigt die Reaktionszeit
- Der Hersteller von Präparat A behauptet, dass sein Mittel den Einfluss von Alkohol auf die Reaktionszeit verringert
- Diese Behauptung soll zum Signifikanzniveau  $\alpha$  überprüft werden
- Bestimme dazu die Reaktionszeit von 27 Personen unter Alkoholeinfluss, von denen 15 zusätzlich noch das Präparat genommen haben.
- Seien  $x_1, \dots, x_{12}$  die Reaktionszeiten ohne und  $y_1, \dots, y_{15}$  mit Präparat A

## Beispieldaten

| ohne Präparat A | mit Präparat A |
|-----------------|----------------|
| Zeit in [ms]    | Zeit in [ms]   |
| 85              | 96             |
| 106             | 105            |
| 118             | 104            |
| 81              | 108            |
| 138             | 86             |
| 90              | 84             |
| 112             | 99             |
| 119             | 101            |
| 107             | 78             |
| 95              | 124            |
| 88              | 121            |
| 103             | 97             |
|                 | 129            |
|                 | 87             |
|                 | 109            |

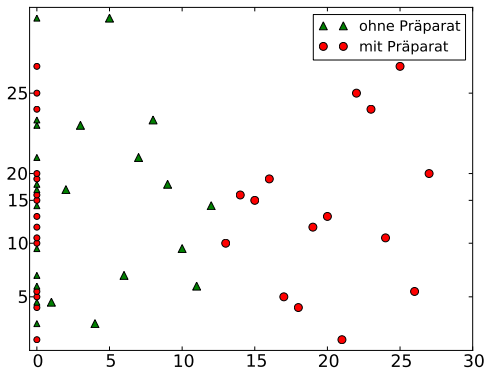
# Reaktionszeiten, Fortsetzung

- Alle 27 Werte werden angeordnet, der kleinste bekommt den Rang 1, der größte den Rang 27
- Da alle Werte verschieden sind, gelingt die Anordnung
- Bei gleichen Werten spricht man von "Bindungen"

## Ränge der Beispieldaten

| ohne Präparat A |      | mit Präparat A |      |
|-----------------|------|----------------|------|
| Zeit in [ms]    | Rang | Zeit in [ms]   | Rang |
| 85              | 4    | 96             | 10   |
| 106             | 17   | 105            | 16   |
| 118             | 22   | 104            | 15   |
| 81              | 2    | 108            | 19   |
| 138             | 27   | 86             | 5    |
| 90              | 8    | 84             | 3    |
| 112             | 21   | 99             | 12   |
| 119             | 23   | 101            | 13   |
| 107             | 18   | 78             | 1    |
| 95              | 9    | 124            | 25   |
| 88              | 7    | 121            | 24   |
| 103             | 14   | 97             | 11   |
|                 |      | 129            | 26   |
|                 |      | 87             | 6    |
|                 |      | 109            | 20   |

# Grafik



Daten des Beispiels “Reaktionszeit”. Auf der y-Achse kann man die Ränge ablesen

## Zurück zum Beispiel

- Sei  $\mu_1$  der Erwartungswert ohne Präparat,  $\mu_2$  mit Präparat
- $H_0 : \mu_1 \leq \mu_2$
- Signifikanzniveau  $\alpha = 0.05$
- $q_{0.95} = 1.645$

- Die Anzahl der Rangüberschreitungen ist 86

$$u = 86$$

$$n_1 = 12$$

$$n_2 = 15$$

$$\mu = \frac{n_1 \cdot n_2}{2} = 90$$

$$\sigma = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}} = \sqrt{420} = 20.49$$

$$t = \frac{u - \mu}{\sigma} = -0.1952$$

- $H_0$  wird abgelehnt für  $t < -q_{1-\alpha}$
- Das ist nicht der Fall

# U-Test mit Bindungen

- Im Fall von Bindungen wird jedem Messwert als Rangplatz das arithmetische Mittel der fortlaufend vergebenen Rangplätze zugewiesen.
- Der einzige Unterschied zum Fall ohne Bindungen ist, dass  $\sigma$  durch die korrigierte Größe

$$\sigma_{\text{kor}} = \sqrt{\frac{n_1 \cdot n_2}{12n \cdot (n-1)} \left( n^3 - n - \sum_{k=1}^K (t_k^3 - t_k) \right)}$$

ersetzt wird.

- Dabei

$$n = n_1 + n_2$$

$K$  = Anzahl der Messwertgruppen mit gleichen Werten

$t_k$  = Anzahl der Bindungen in der  $k$ -ten Gruppe