Biostatistik, WS 2013/2014

Wilcoxons Rangsummen-Test

Matthias Birkner

http://www.mathematik.uni-mainz.de/~birkner/Biostatistik1314/

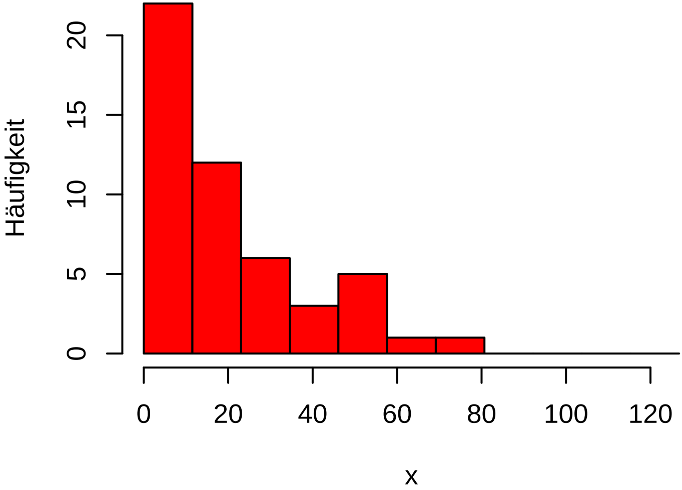
20.12.2013

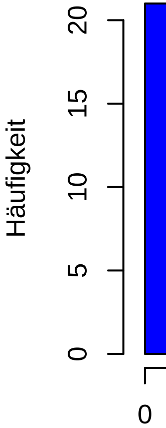
JOHANNES GUTENBERG UNIVERSITÄT MAINZ

Bei (ungefähr) glockenförmigen und
symmetrisch verteilten Beobachtungen
oder wenn die Stichprobenumfänge genügend groß sind
können wir den t-Test benutzen,
um die Nullhypothese $\mu_1 = \mu_2$ zu testen:
Die t-Statistik ist (annährend) Student-verteilt.

Besonders bei sehr asymmetrischen und
langschwänzigen Verteilungen
kann das anders sein

Nehmen wir an, wir sollten folgende Verteilungen vergleichen:





Beispiele

- Wartezeiten
- Ausbreitungsentfernungen
- Zelltypenhäufigkeiten

Gesucht:

ein „verteilungsfreier“ Test,
mit dem man die Lage zweier Verteilungen
zueinander testen kann

Beobachtungen: Zwei Stichproben

$$X : x_1, x_2, \ldots, x_m$$

$$Y : y_1, y_2, \ldots, y_n$$

Wir möchten die Nullhypothese:
X und Y aus derselben Population
(X und Y haben diesselbe Verteilung)
testen

gegen die Alternative:
Die beiden Verteilungen sind gegeneinander verschoben.

Wir sind also in einer Situation, die wir schon beim t-Test getroffen haben: Die zwei Verteilungen sind möglicherweise gegeneinander verschoben (haben insbesondere möglicherweise unterschiedliche Mittelwerte), aber wir möchten *nicht* die implizite Annahme treffen, dass es sich dabei (wenigstens ungefähr) um Normalverteilungen handelt.

Idee

Beobachtungen:

$$X : x_1, x_2, \ldots, x_m$$

$$Y : y_1, y_2, \ldots, y_n$$

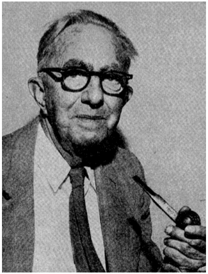
- Sortiere alle Beobachtungen der Größe nach.
- Bestimme die Ränge der m X-Werte unter allen $m + n$ Beobachtungen.
- Wenn die Nullhypothese zutrifft, sind die m X-Ränge eine rein zufällige Wahl aus $\{1, 2, \ldots, m + n\}$.
- Berechne die Summe der X-Ränge, prüfe, ob dieser Wert untypisch groß oder klein.

Wilcoxons Rangsummenstatistik

Beobachtungen:

$X : x_1, x_2, \ldots, x_m$

$Y : y_1, y_2, \ldots, y_n$



Frank Wilcoxon, 1892–1965

$W =$ Summe der X-Ränge $- (1 + 2 + \cdots + m)$

heißt

Wilcoxons Rangsummenstatistik

Wilcoxons Rangsummenstatistik

Bemerkung:

$$W = \text{Summe der } X\text{-Ränge} - (1 + 2 + \cdots + m)$$

Wir könnten auch die Summe der Y-Ränge benutzen, denn

$$\begin{aligned} &\text{Summe der } X\text{-Ränge} + \text{Summe der } Y\text{-Ränge} \\ &\quad = \text{Summe aller Ränge} \\ &\quad = 1 + 2 + \cdots + (m+n) = \frac{(m+n)(m+n+1)}{2} \end{aligned}$$

Bemerkung

Der Wilcoxon Test heißt auch Mann-Whitney-Test.
Die Mann-Whitney Statistik $U = W +$ Konstante.

Ein kleines Beispiel

- Beobachtungen:

$$X : 1{,}5;\ 5{,}6;\ 35{,}2$$
$$Y : 7{,}9;\ 38{,}1;\ 41{,}0;\ 56{,}7;\ 112{,}1;\ 197{,}4;\ 381{,}8$$

- Lege Beobachtungen zusammen und sortiere:
  1,5; 5,6; 7,9; 35,2; 38,1; 41,0; 56,7; 112,1; 197,4; 381,8
- Bestimme Ränge:
  1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- Rangsumme: $W = 1 + 2 + 4 - (1 + 2 + 3) = 1$

Interpretation von W

X-Population kleiner $\Longrightarrow$ W klein:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $W = 0$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $W = 1$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $W = 2$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $W = 2$

X-Population größer $\Longrightarrow$ W groß:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $W = 21$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $W = 20$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $W = 20$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $W = 19$

Signifikanz

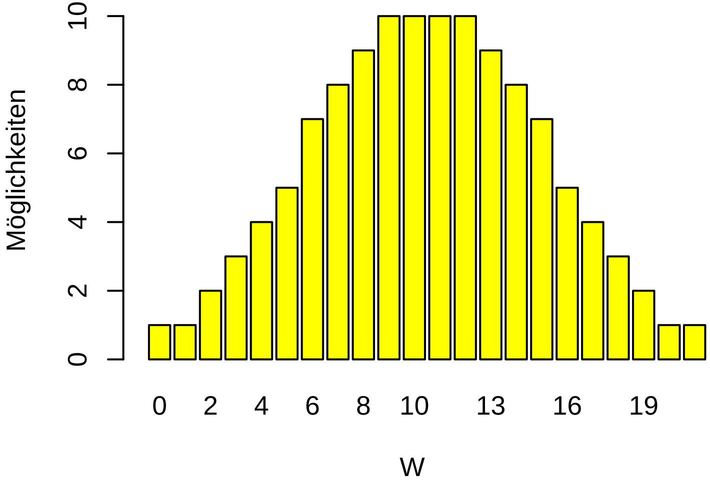
Nullhypothese:
X-Stichprobe und Y-Stichprobe
stammen aus
derselben Verteilung

Die 3 Ränge der X-Stichprobe
1 2 3 4 5 6 7 8 9 10
hätten genausogut irgendwelche 3 Ränge
1 2 3 4 5 6 7 8 9 10
sein können.

Es gibt $\frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$ Möglichkeiten.

(Allgemein: $\frac{(m+n)(m+n-1)\cdots(n+1)}{m(m-1)\cdots 1}) = \frac{(m+n)!}{n!m!} = \binom{m+n}{m}$ Möglichkeiten)

Verteilung der Wilcoxon-Statistik ($m = 3, n = 7$)



Unter der Nullhypothese sind alle Rangbelegungen gleich wahrscheinlich, also

$$\mathbb{P}(W = w) = \frac{\text{Anz. Möglichkeiten mit Rangsummenstatistik } w}{120}$$

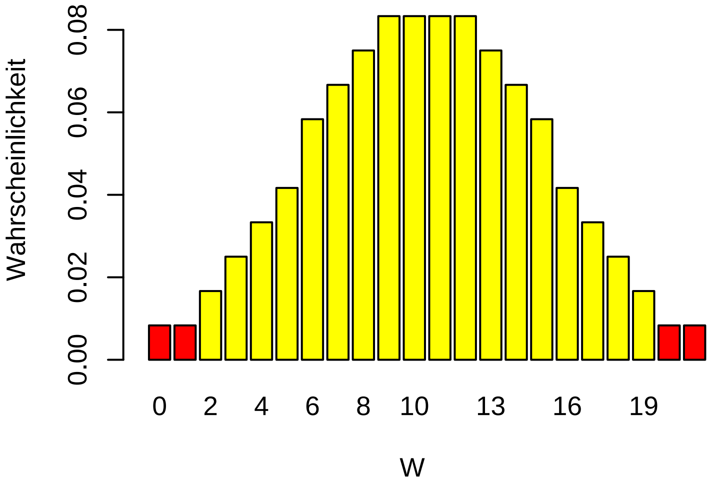
Wir beobachten in unserem Beispiel:

1,5, 5,6; 7,9; 35,2; 38,1; 41,0; 56,7; 112,1; 197,4; 381,8

somit $W = 1$

$$\begin{aligned} &\mathbb{P}(W \leq 1) + \mathbb{P}(W \geq 20) \\ &= \mathbb{P}(W = 0) + \mathbb{P}(W = 1) + \mathbb{P}(W = 20) + \mathbb{P}(W = 21) \\ &= \tfrac{1+1+1+1}{120} \doteq 0{,}033 \end{aligned}$$

Verteilung der Wilcoxon-Statistik ($m = 3, n = 7$)



Für unser Beispiel ($W = 1$) also:

p-Wert = $\mathbb{P}(\text{ein so extremes } W) = 4/120 = 0{,}033$

Wir lehnen die Nullhypothese,
dass die Verteilungen
von X und Y
identisch sind,
auf dem 5%-Niveau ab.

Achtung

Achtung!!!

Wenn der Wilcoxon-Test Signifikanz anzeigt,
so kann das daran liegen, dass die zu grunde
liegenden Verteilungen verschiedene Formen haben.

Der Wilcoxon-Test kann beispielsweise Signifikanz anzeigen,
selbst wenn die Stichproben-Mittelwerte übereinstimmen!

Vergleich von *t*-Test und Wilcoxon-Test

Beachte:

Sowohl der *t*-Test als auch
der Wilcoxon-Test können verwendet werden,
um eine vermutete Verschiebung der Verteilung zu stützen.

Der *t*-Test testet „nur“ auf Gleichheit der Erwartungswerte.
Der Wilcoxon-Test dagegen testet auf Gleichheit
der gesamten Verteilungen.

In den meisten Fällen liefern beide Tests dasselbe Ergebnis.
Im Allgemeinen empfehlen wir den *t*-Test, da er robuster ist.

In besonderen Fällen

- Verteilungen sind asymmetrisch
- Stichprobenlänge ist klein

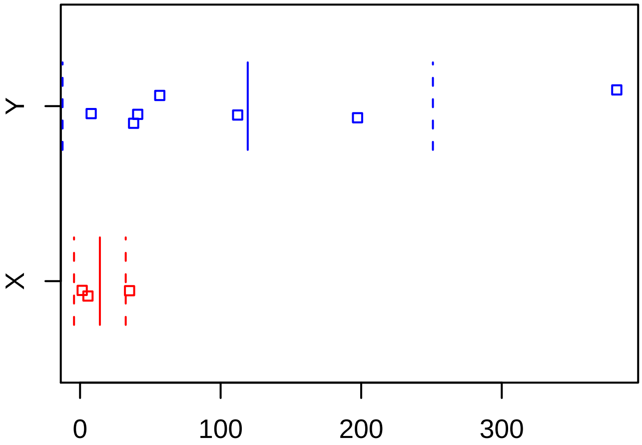
hat der Wilcoxon-Test eine höhere Testpower.

Vergleichen wir (spaßeshalber) mit dem *t*-Test:

```
> x
[1]  1.5  5.6 35.2
> y
[1]   7.9  38.1  41.0  56.7 112.1 197.4 381.8
> t.test(x,y)

Welch Two Sample t-test

data:  x and y
t = -2.0662, df = 6.518, p-value = 0.08061
alternative hypothesis: true difference in means is not e
95 percent confidence interval:
 -227.39182   17.02039
sample estimates:
mean of x mean of y
  14.1000  119.2857
```


Y
X
0
100
200
300