# Biostatistik

## Wilcoxons Rangsummen-Test

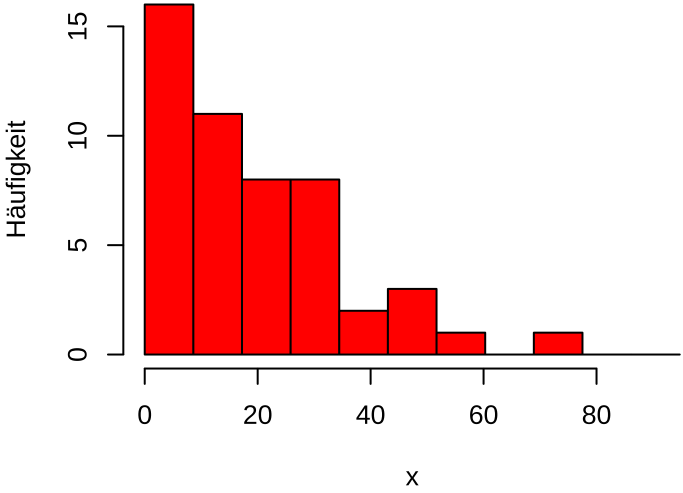
Matthias Birkner

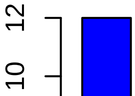
JOHANNES GUTENBERG UNIVERSITÄT MAINZ

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
- Genexpressionsniveaus

Gesucht:

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

(Eine Aufgabe der *nicht-parametrischen Statistik*.)

Beobachtungen: Zwei Stichproben

$$X : x_1, x_2, \dots, x_m$$
$$Y : y_1, y_2, \dots, y_n$$

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:

$$X : x_1, x_2, \ldots, x_m$$
$$Y : y_1, y_2, \ldots, y_n$$

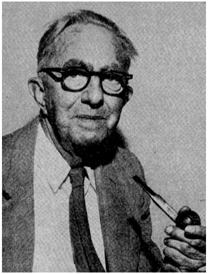
- Sortiere alle Beobachtungen der Größe nach.
- Bestimme die Ränge der $m$ $X$-Werte unter allen $m + n$ Beobachtungen.
- Wenn die Nullhypothese zutrifft, sind die $m$ $X$-Ränge eine rein zufällige Wahl aus $\{1, 2, \ldots, m + n\}$.
- Berechne die Summe der $X$-Ränge, prüfe, ob dieser Wert untypisch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

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

heißt

Wilcoxons Rangsummenstatistik

(Der minimal mögliche Wert ist 0, der maximal mögliche Wert ist $m \cdot n$.)

# Wilcoxons Rangsummenstatistik

Bemerkung:

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

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

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

Bemerkung

Der Wilcoxon-Test heißt auch Mann-Whitney-Test.

In der Literatur sind verschiedene Zentrierungen der Rangsumme gebräuchlich, ggfs. prüfen, bevor Sie eine Formel aus einem Buch verwenden.

# 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: $U = 1 + 2 + 4 - (1 + 2 + 3) = 1$

# Interpretation von $U$

$X$-Population kleiner $\Longrightarrow$ $U$ klein:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $\quad U = 0$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $\quad U = 1$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $\quad U = 2$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $\quad U = 2$

$X$-Population größer $\Longrightarrow$ $U$ groß:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $\quad U = 21$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $\quad U = 20$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $\quad U = 20$

1, 2, 3, 4, 5, 6, 7, 8, 9, 10 $\quad U = 19$

# Signifikanz

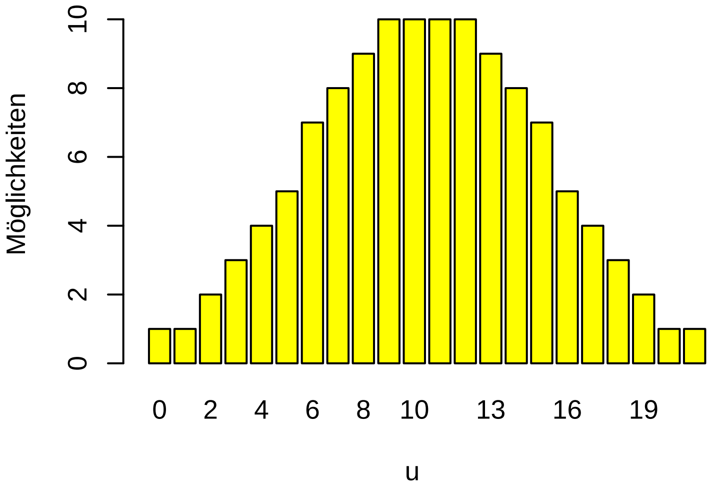
Nullhypothese:
$X$-Stichprobe und $Y$-Stichprobe
stammen aus
derselben Verteilung

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 ($m = 3, n = 7$)



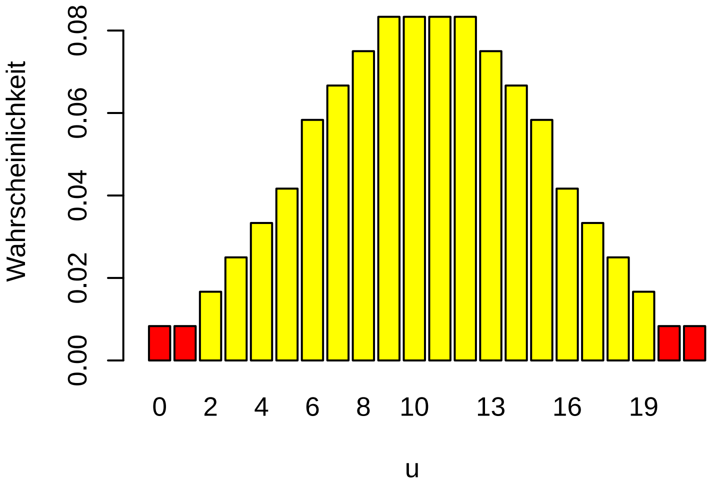
Unter der Nullhypothese sind alle Rangbelegungen gleich wahrscheinlich, also

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

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 $U = 1$

$$\mathbb{P}(U \leq 1) + \mathbb{P}(U \geq 20)$$
$$= \mathbb{P}(U = 0) + \mathbb{P}(U = 1) + \mathbb{P}(U = 20) + \mathbb{P}(U = 21)$$
$$= \frac{1+1+1+1}{120} \doteq 0{,}033$$

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



Für unser Beispiel ($U = 1$) also:

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

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

`R` kennt den Wilcoxon-Test mittels `wilcox.test`:

```
> x <- c(1.5,  5.6, 35.2)
> y <- c(7.9,  38.1,  41.0,  56.7, 112.1, 197.4, 381.8)

> wilcox.test(x,y)

Wilcoxon rank sum test

data:  x and y
W = 1, p-value = 0.03333
alternative hypothesis: true location shift is
not equal to 0
```

# Wilcoxon-Test: allgemeine Theorie

Gegeben zwei Stichproben $x_1, x_2, \ldots, x_m$ ($m$ $x$-Werte) und $y_1, \ldots, y_n$ ($n$ $y$-Werte)

bilde (normierte) Rangsumme

$$\begin{aligned} U &= \text{Summe aller } x\text{-Ränge} - (1 + 2 + \cdots + m) \\ &= \sum_{i=1}^{m} \#\{j \le n : x_i > y_j\} \;=: \sum_{i=1}^{m} U_i \end{aligned}$$

(mit möglichen Werten in $0, 1, \ldots, m \cdot n$).

Unter $H_0$ : die $x$-Werte und die $y$-Werte sind unabhängige Stichproben aus derselben Verteilung

hängt die Verteilung von $U$ nicht von der tatsächlichen Verteilung ab.

(Strenggenommen: Sofern die Verteilung „stetig“ ist, also kein Wert mehrmals vorkommen kann)

# Wilcoxon-Test: allgemeine Theorie, 2

normierte Rangsumme $U = U_{m,n} = \sum_{i=1}^{m} \#\{j \le n : x_i > y_j\}$

$H_0$ : $x$-Werte und $y$-Werte stammen aus derselben Verteilung

Die Verteilung von $U_{m,n}$ unter $H_0$ ist (prinzipiell) ein rein kombinatorisches Problem:

$$\mathbb{P}_{H_0}(U_{m,n} = u) = \frac{\#\left(\begin{array}{l}\text{Zuordnungen von } m \text{ Rängen an } x\text{-Werte,} \\ \text{die Rangsumme } u \text{ ergeben}\end{array}\right)}{\binom{m+n}{m}}$$

Für sehr kleine Stichproben kann man dies durch Auszählen bestimmen (wie vorhin), für größere Stichproben verwendet man eine Rekursionsformel (in `R` implementiert) oder eine Normalapproximation.

# Wilcoxon-Test: allgemeine Theorie, 3

$$\mathbb{P}_{H_0}(U_{m,n} = u) = \frac{\#\binom{\text{Zuordnungen von } m \text{ Rängen an } x\text{-Werte,}}{\text{die Rangsumme } u \text{ ergeben}}}{\binom{m+n}{m}}$$

(in `R`: `dwilcox(u,m,n)`, z.B. `dwilcox(1,3,7)=0.00833`, `R` kennt auch die Verteilungsfunktion `pwilcox` und die Quantilfunktion `qwilcox`.)

Quantile $u_{m,n,\alpha}$ mit

$$\mathbb{P}_{H_0}(U_{m,n} \leq u_{m,n,\alpha}) = \alpha$$

für $\alpha \in (0,1)$ findet man in Tabellen

(oder mit `R`: `qwilcox(`$\alpha$`,m,n)`)

# Zweiseitiger Wilcoxon-Rangsummen-Test

Gegeben $m$ $x$-Werte $x_1, \dots, x_m$, $n$ $y$-Werte $y_1, \dots, y_n$, Signifikanzniveau $\alpha \in (0, 1)$.

Bestimme (normierte) Rangsumme $U = \sum_{i=1}^{m} \#\{j \leq n : x_i > y_j\}$, $u_{m,n,1-\alpha/2}$ mit $\mathbb{P}_{H_0}(U_{m,n} \leq u_{m,n,1-\alpha/2}) = 1 - \alpha/2$ (aus Tabelle oder mit R)

Lehne $H_0$: die $x$-Werte und die $y$-Werte sind unabhängige Stichproben aus derselben Verteilung

ab zugunsten von

$H_1$: die Verteilung der $x$-Werte ist verschoben im Vergleich zu der der $y$-Werte

ab, falls

$$U > u_{m,n,1-\alpha/2} \quad \text{oder} \quad U < m \cdot n - u_{m,n,1-\alpha/2}$$

# Einseitiger Wilcoxon-Rangsummen-Test

Gegeben $m$ $x$-Werte $x_1, \ldots, x_m$, $n$ $y$-Werte $y_1, \ldots, y_n$, Signifikanzniveau $\alpha \in (0, 1)$.

Bestimme (normierte) Rangsumme $U = \sum_{i=1}^{m} \#\{j \leq n : x_i > y_j\}$,

$u_{m,n,1-\alpha}$ mit $\mathbb{P}_{H_0}(U_{m,n} \leq u_{m,n,1-\alpha}) = 1 - \alpha$ (aus Tabelle oder mit `R`)

Lehne $H_0$ ab zugunsten von

$H_1$: die Verteilung der $x$-Werte hat mehr Gewicht auf größen Werten im Vergleich zu der der $y$-Werte

ab, falls

$$U > u_{m,n,1-\alpha}$$

(analog für $H_1$ : „Verteilung der $x$-Werte nach links verschoben im Vgl. zu $y$-Werten“, falls $U < m \cdot n - u_{m,n,1-\alpha}$,
oder vertausche Rollen von $x$ und $y$.)

# Wilcoxon-Test: Normalapproximation

Unter $H_0$ : $x$-Werte und $y$-Werte stammen aus derselben Verteilung

ist

$$\frac{U_{m,n} - \frac{m \cdot n}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \approx \text{Standard-normalverteilt}$$

Damit lautet der (approximative) zweiseitige Test zum Signifikanzniveau $\alpha$ bei beobachter Rangsumme $U$:

$$\text{lehne } H_0 \text{ ab, wenn } \left| U - \frac{m \cdot n}{2} \right| > z_{1-\alpha/2} \sqrt{\frac{mn(m+n+1)}{12}}$$

wobei $z_{1-\alpha/2}$ = $(1 - \alpha/2)$-Quantil der Standard-Normalverteilung

Analog einseitig ($H_1$ = „Verteilung der $x$-Werte nach rechts verschoben im Vgl. zu $y$-Werten“):

Lehne $H_0$ ab, falls $U > \frac{m \cdot n}{2} + z_{1-\alpha} \sqrt{\frac{mn(m+n+1)}{12}}$.

# Wilcoxon-Test, Fall mit Bindungen

Bei diskreten Verteilungen (und in der Praxis!) kann es vorkommen, dass einzelne Werte in den Daten $x_i$ und $y_j$ mehrfach vorkommen. Dies nennt man Bindungen. In diesem Fall wird der Rang von $x_i$ in den $y_1, \dots, y_n$ berechnet als

$$U_i = \text{Anzahl der } j \text{ mit } y_j < x_i + \frac{1}{2}\text{Anzahl der } j \text{ mit } y_j = x_i$$

und die Rangsumme als

$$U = \sum_{i=1}^{m} U_i.$$

In diesem Fall hält der Wilcoxon-Test das geforderte Niveau nicht exakt ein, meistens aber doch recht gut. R gibt dann eine Warnmeldung.

# Beachte:

Wenn der Wilcoxon-Test Signifikanz anzeigt,
so kann das daran liegen, dass die zugrunde
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

Sowohl der (2-Stichproben-)*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. Sofern die zugrundeliegenden Verteilungen einigermaßen glockenförmig sind, ist der *t*-Test empfehlenswerter.

In besonderen Fällen

- Verteilungen sind asymmetrisch
- Stichprobenlänge ist klein

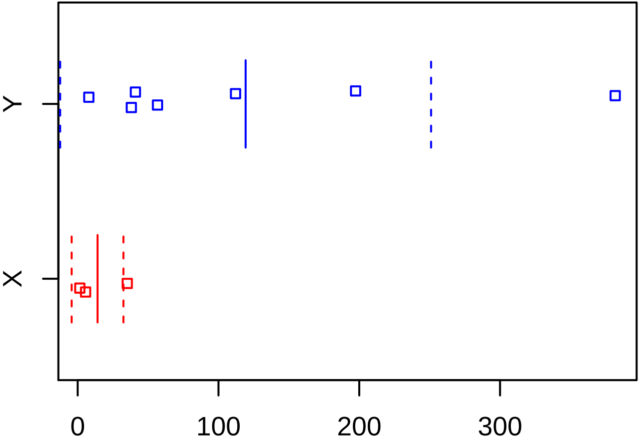
hat der Wilcoxon-Test eine höhere Testpower.

# 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 equal to 0
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