Vertrauensintervall
Zweistichprobentest
Wilcoxon-Test

Peter Büchel

HSLU I

ASTAT: Block 10

Vertrauensintervall

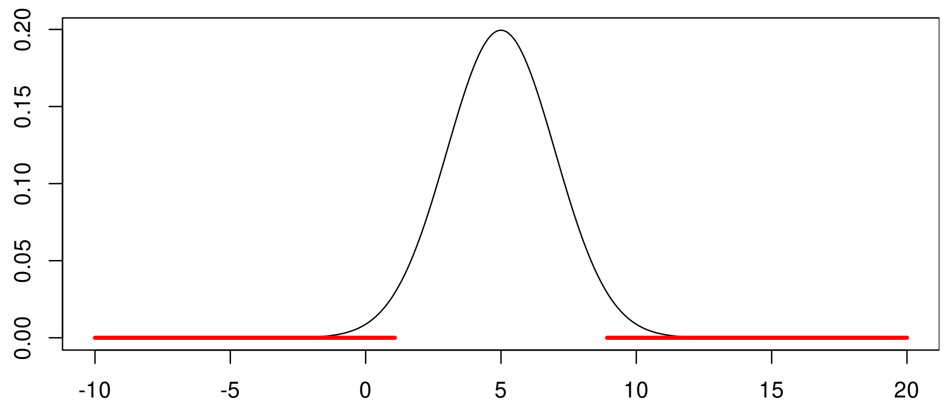
- Bei Punktschätzung für Mittelwert μ einer Messreihe $\rightarrow$ *Einziger* Zahlwert
- Wissen aber nicht, wie nahe dieser geschätzte Mittelwert beim wahren, aber meist unbekannten, Mittelwert der Verteilung der Messreihe liegt
- Vertrauensintervall: Intervall, das angibt , wo, grob gesagt, der wahre Mittelwert mit einer bestimmten vorgegebenen W'keit liegt
- Wollen mit Beispiel Vertrauensintervall verschaulichen

Beispiel

- Bestimmung Verwerfungsbereiches beim z-Test: Gehen vom einem wahren (aber unbekannten) Wert μ aus und einer bekannten Standardabweichung aus
- Bestimmen Quartile $q_{0.025}$ und $q_{0.975}$ (zweiseitiger Test, $\alpha = 0.05$)
- Normalverteilungskurve $X \sim \mathcal{N}(5, 2^2)$
- $q_{0.025}$- und $q_{0.975}$-Quantile

```
qnorm(p = c(0.025, 0.975), mean = 5, sd = 2)
## [1] 1.080072 8.919928
```

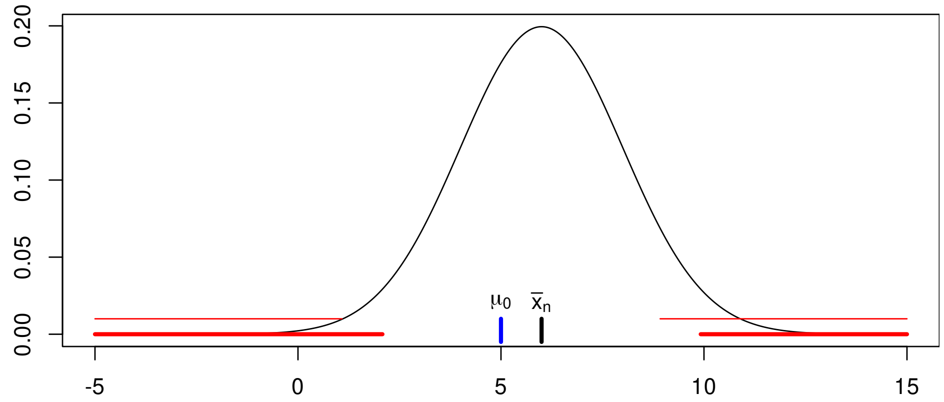

- Abbildung: Normalverteilungskurve mit dem Verwerfungsbereich rot:



- Liegt nun $\overline{x}_n$ im Verwerfungsbereich (roter Bereich), dann wird die Nullhypothese H_0 verworfen

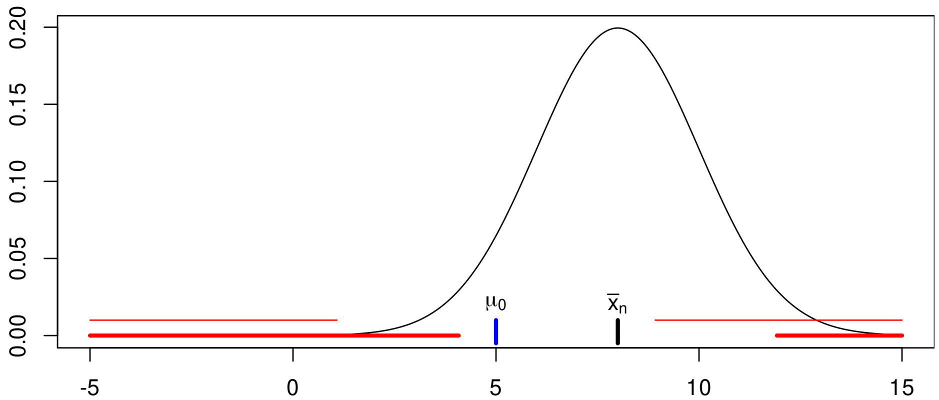
- Wahres μ praktisch immer unbekannt: Wert einfach angenommen
- Frage umkehren: Kennen $\overline{x}_n$ und fragen, für welche μ wird H_0 *nicht* verworfen
- Kann man rechnerisch herleiten, machen es aber nur graphisch
- Annahme $\mu_0 = 5$
- Gegeben $\overline{x}_n = 6$ und zeichnen Verwerfungsbereich für diesen Wert ein

- Abbildung:



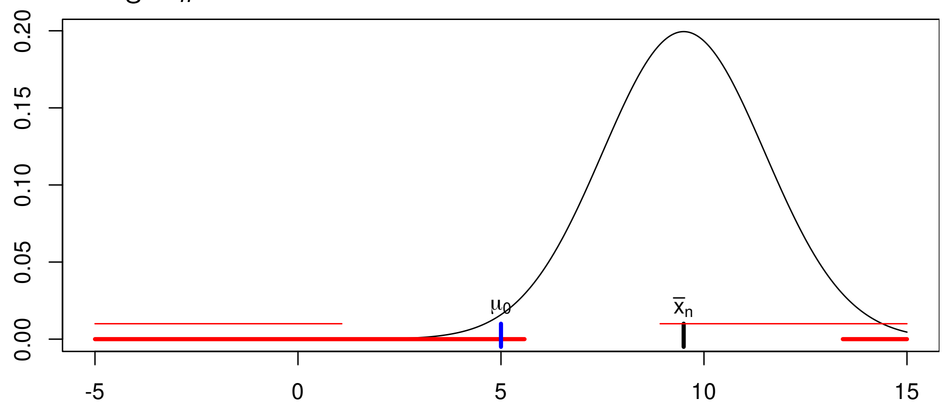
- Linien:
  - Dicke roten Linien: Verwerfungsbereich für $\overline{x}_n = 6$
  - Dünne roten Linien: Verwerfungsbereich für $\mu_0 = 5$
  - Vertikaler schwarze Strich: $\overline{x}_n = 6$
  - Vertikaler blauer Strich: $\mu_0 = 5$

- Stellen fest: $\overline{x}_n$ und μ_0 nicht innerhalb in einem der beiden Verwerfungsbereiche
- Idee: $\overline{x}_n$ vergrössern und $\mu_0 = 5$ konstant lassen
- Abbildung: $\overline{x}_n = 8$:



- Auch hier: $\overline{x}_n$ und μ_0 nicht innerhalb der beiden Verwerfungsbereiche

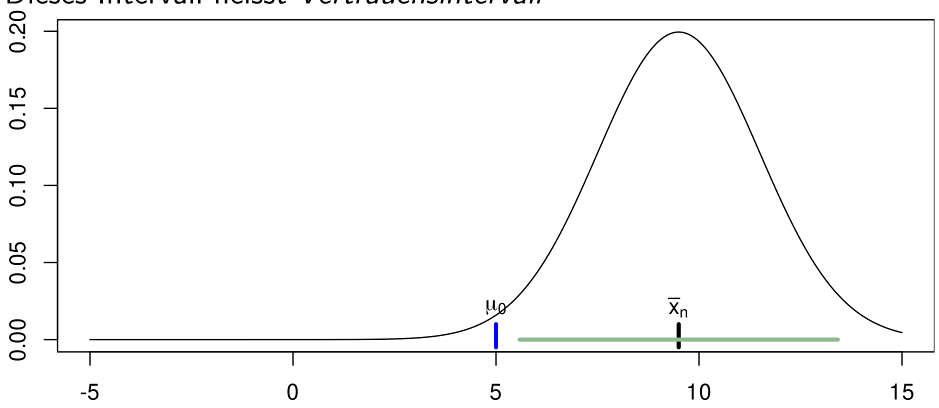
- Abbildung: $\overline{x}_n = 9.5$



- Andere Situation: $\overline{x}_n$ (schwarze Linie) im Verwerfungsbereich von $\mu_0 = 5$ (dünne blaue Linien)
- Nullhypothese H_0 nun verworfen
- Aber: $\mu_0 = 5$ im Verwerfungsbereich für $\overline{x}_n = 9.5$.

- Andere Darstellung: Bereich, der *nicht* zum Verwerfungsbereich gehört

- Dieses Intervall heisst *Vertrauensintervall*



- Wert 5 liegt nicht im Vertrauensintervall:

```
qnorm(p = c(0.025, 0.975), mean = 9.5, sd = 2)
## [1]  5.580072 13.419928
```


Interpretation: Vertrauensintervall

- Annahme: Kennen wahre Verteilung $\mathcal{N}(5, 2^2)$

- Es gilt also $\mu = \mu_0 = 5$

- Wählen Zufallszahl aus, die Normalverteilung $\mathcal{N}(5, 2^2)$ folgt

```
set.seed(1)
m <- rnorm(n = 1, mean = 5, sd = 2)
m
## [1] 3.747092
```

- Bestimmen Vertrauensintervall von 3.7470924:

```
qnorm(p = c(0.025, 0.975), mean = m, sd = 2)
## [1] -0.1728356  7.6670203
```

- Feststellung $\mu_0 = 5$ liegt im Vertrauensintervall

- Wählen andere Zufallszahl aus, die Normalverteilung $\mathcal{N}(5, 2^2)$ folgt

```
set.seed(7)
m <- rnorm(n = 1, mean = 5, sd = 2)
m
## [1] 9.574494
```

- Bestimmen Vertrauensintervall von 9.5744943:

```
qnorm(p = c(0.025, 0.975), mean = m, sd = 2)
## [1]  5.654566 13.494422
```

- Feststellung $\mu_0 = 5$ liegt nicht im Vertrauensintervall

- Frage: In wievielen Fällen liegt $\mu_0 = 5$ im Vertrauensintervall einer zufällig ausgewählten Zahl, die $\mathcal{N}(5, 2^2)$ folgt?

- Folgender Code: Bestimmt 1000 mal Zufallszahl und zählt wieviele Male $\mu_0 = 5$ im Vertrauensintervall der Zufallszahl liegt:

```
n <- 1000

r <- rnorm(n = n, mean = 5, sd = 2)
q_u <- qnorm(p = c(0.025), mean = r, sd = 2)
q_o <- qnorm(p = c(0.975), mean = r, sd = 2)

k <- 0

for (i in 1:n) {
    if ((q_u[i] <= 5 & 5 <= q_o[i]) == FALSE) {
        k <- k + 1
    }
}

print(k)
## [1] 47
```

- In 47 von 1000 Fällen liegt $\mu_0 = 0$ nicht im Vertrauensintervall der Zufallszahlen

- Folgender Code: Macht das 20 mal

```
vi2 <- function(n) {
    r <- rnorm(n = n, mean = 5, sd = 2)
    q_u <- qnorm(p = c(0.025), mean = r, sd = 2)
    q_o <- qnorm(p = c(0.975), mean = r, sd = 2)

    k <- 0

    for (i in 1:n) {
        if ((q_u[i] <= 5 & 5 <= q_o[i]) == FALSE) {
            k <- k + 1
        }
    }
    cat(k, " ")
}

for (i in 1:20) {
    vi2(1000)
}
## 52  52  50  58  46  55  60  53  48  47  56  48  52  50  66  61  47  51  57  42
```

- In etwa 50 von 1000 Fällen liegt $\mu_0 = 0$ nicht im Vertrauensintervall der Zufallszahlen

- Das sind etwa 5 %

Graphisch

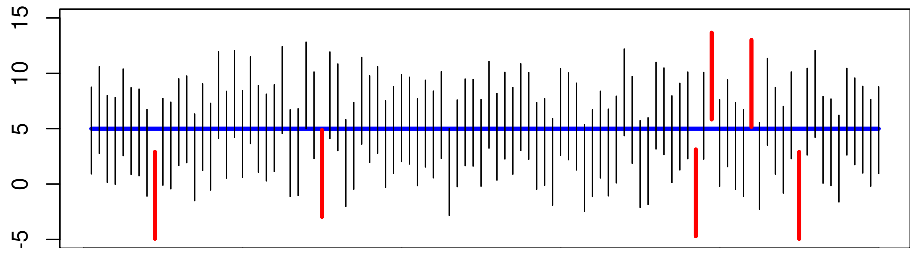
- Rote Linien zeigen an, wo $\mu_0 = 5$ nicht im Vertrauensintervall enthalten ist:

```
r <- rnorm(n = 100, mean = 5, sd = 2)
q_u <- qnorm(p = c(0.025), mean = r, sd = 2)
q_o <- qnorm(p = c(0.975), mean = r, sd = 2)

plot(NULL, xlim = c(1, 100), ylim = c(-5, 15), xlab = "", ylab = "")

lines(c(1, 100), c(5, 5), lwd = 3, col = "blue")
for (i in 1:100) {
    lines(c(i, i), c(q_u[i], q_o[i]))
    if ((q_u[i] <= 5 & 5 <= q_o[i]) == FALSE) {
        lines(c(i, i), c(q_u[i], q_o[i]), col = "red", lwd = 3)
    }

}
```



- Hier 100 Vertrauensintervalle

- In 6 Fällen ist $\mu_0 = 5$ nicht im Vertrauensintervall

- Führen dies oft durch: In etwa 5 % der Fälle liegt $\mu_0 = 5$ nicht im Vertrauensintervall

- Interpretation: Zu 95 % liegt der wahre Mittelwert im Vertrauensintervall

- Man spricht von einem 95 %-Vertrauensintervall

- Gesehen: Nullhypothese wird verworfen

- Fällt wahres μ also aus dem Vertrauensintervall, dann wird Nullhypothese verworfen

- Weiteren Interpretation des Vertrauensintervalls: Enthält alle μ_0's für die Nullhypothese *nicht* verworfen wird

- Es sagt uns also in welchem Intervall sich das wahre μ_0 befindet

- Gilt nicht absolut: Mit einer bestimmten W'keit liegt wahres μ_0 in diesem Intervall

- Hier: Wahres μ_0 liegt zu 95 % im Vertrauensintervall

- Sprechen deswegen auch von einem 95 %-Vertrauensintervall

- Weitere Möglichkeit für Testentscheid:
  - Liegt μ_0 der Nullhypothese im Vertrauensintervall, so wird die Nullhypothese *nicht* verworfen
  - Liegt μ_0 der Nullhypothese *nicht* im Vertrauensintervall, so wird die Nullhypothese verworfen

- `R`-Output: Gibt Vertrauensintervall (`confidence interval`) an
- Dieses besagt, dass bei einem Signifikanzniveau von 5 % das wahre μ zu 95 % in diesem Intervall liegt
- Mit Vertrauensintervall kann man ebenfalls Testentscheid durchführen

Beispiel: Waage A

- Nullhypothese

$$H_0: \quad \mu_0 = 80$$

- `R`-Output: Vertrauensintervall:

$$[80.00629, \; 80.03525]$$

- Zu 95 % liegt das wahre μ in diesem Intervall
- Aber $\mu_0 = 80$ *nicht* in diesem Intervall
- Zu 95 % Sicherheit ist das wahre μ *nicht* 80
- Nullhypothese wird verworfen und Alternativhypothese angenommen

Beispiel: Körpergrösse Frauen

- Nullhypothese:

$$H_0: \quad \mu_0 = 180$$

- `R`-Output: Vertrauensintervall:

$$(-\infty, \; 169.382]$$

- Zu 95 % liegt das wahre μ in diesem Intervall
- $\mu_0 = 180$ *nicht* in diesem Intervall
- Mit 95 % Sicherheit ist das wahre μ *nicht* 80
- Nullhypothese verwerfen; Alternativhypothese annehmen

Bemerkung

- Je schmaler das Vertrauensintervall ist, umso sicherer weiss man, wo sich der wahre Mittelwert befindet
- Ist das Vertrauensintervall schmal, wie

$$[105.12,\ 105.23]$$

so wissen wir sehr genau, wo der wahre Mittelwert mit 95 % Wahrscheinlichkeit liegt

- Ist das Vertrauensintervall breit, wie

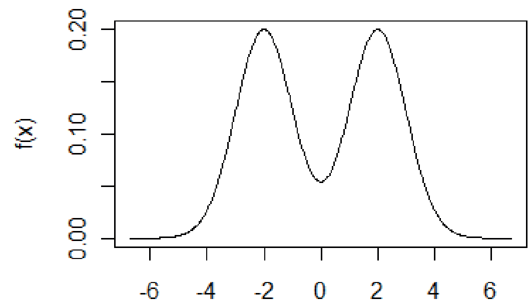
$$[10,\ 1000]$$

so besteht grosse Unsicherheit, wo das wahre μ liegt

Nicht-normalverteilte Daten: Wilcoxon-Test

- Alternative zu t-Test
- Wilcoxon-Test ist der weniger voraussetzt als der t-Test
- Voraussetzung: Verteilung unter Nullhypothese *symmetrisch* bez. μ_0
- Annahme:

$$X_i \sim F \text{ iid}, \quad F \text{ ist symmetrisch}$$



- Es wird ein V-Wert (*Rangsumme*) berechnet
- Grundidee dieselbe wie bei Hypothesentest bisher:
  - V-Wert„ weit" weg vom *Median*: Nullhypothese verwerfen
  - V-Wert „nahe" beim *Median*: Nullhypothese nicht verwerfen
  - R berechnet p-Wert

Beispiel: Waage A

- R Output:

```
x <- c(79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04, 79.97,
    80.05, 80.03, 80.02, 80, 80.02)

wilcox.test(x, mu = 80, alternative = "two.sided")
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  x
## V = 69, p-value = 0.0195
## alternative hypothesis: true location is not equal to 80
```

- Auf Signifikanzniveau von 5 % wird Nullhypothese verworfen, da p-Wert kleiner als 0.05 ist

Wilcoxon-Test versus t-Test

Wilcoxon-Test versus t-Test

DWilcoxon-Test ist in den allermeisten Fällen dem t-Test vorzuziehen: Er hat in vielen Situationen oftmals wesentlich grössere Macht (Wahrscheinlichkeit Nullhypothese richtigerweise zu verwerfen)

Selbst in den ungünstigsten Fällen ist er nie viel schlechter

Vergleich von zwei Stichproben: Mögliche Fragestellungen

- Vergleich von zwei Messverfahren (Messgerät A vs. Messgerät B): Gibt es einen signifikanten Unterschied?
- Vergleich von zwei Herstellungsverfahren (A vs. B): Welches hat die besseren Eigenschaften (z.B. bzgl. Brüchigkeit von Handy-Displays)?
- Werden männliche Dozenten von weiblichen Studierenden besser als von männlichen Studierenden bewertet?
- Sammeln also jeweils Daten von *zwei* Gruppen

Gepaarte Stichproben

- Beispiel Messgeräte: Jeder Prüfkörper wird mit *beiden* Messgeräten gemessen
- Pro *Versuchseinheit* (hier: Prüfkörper) zwei Beobachtungen (einmal Gerät A und einmal Gerät B)
- Man spricht auch von *gepaarten Stichproben*
- Beide Beobachtungen sind *nicht* unabhängig, da an *gleicher* Versuchseinheit zwei Mal gemessen wird!

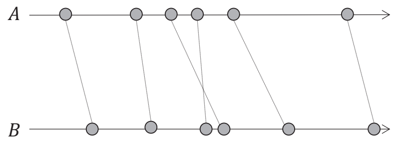
Ungepaarte (unabhängige) Stichproben

- Beispiel Herstellungsverfahren: Stichprobe von Verfahren A und eine andere Stichprobe von Verfahren B und messen jedes Objekt aus
- Beobachtungen sind hier *unabhängig*: „Es gibt *nichts*, was sie verbindet"
- Man spricht auch von *ungepaarten (oder unabhängigen) Stichproben*

Unterscheidung gepaart versus ungepaarte Stichproben

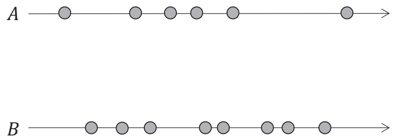
Gepaarte Stichproben

- Jede Beobachtung einer Gruppe kann eindeutig einer Beobachtung der anderen Gruppe zugeordnet werden
- Stichprobengrösse ist in beiden Gruppen zwangsläufig gleich



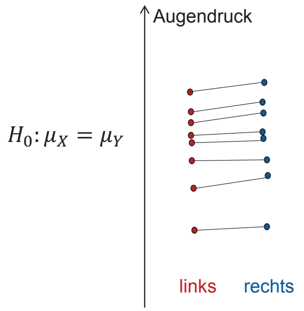
Ungepaarte Stichproben

- Keine Zuordnung von Beobachtungen möglich
- Stichprobengrössen können verschieden sein (müssen aber nicht!)
- Man kann die eine Gruppe vergrössern, ohne dass man die andere vergrössert



Gepaarte versus ungepaarte Stichproben

- Beispiel: Augeninnendruck; ein Auge behandelt, das andere nicht (gepaarter Test ist angebracht)
- Gemäss Vorraussetzungen dürfte auch ein ungepaarter Test angewendet werden



Ungepaart:
Intuition Teststatistik: $T = \frac{\overline{X} - \overline{Y}}{\widehat{\sigma_{\overline{X}}}}$

Gepaart:
Differenz $D_i = X_i - Y_i$
Teststatistik $T = \frac{\overline{D}}{\widehat{\sigma_{\overline{D}}}}$

Statistischer t-Test für gepaarte Stichproben mit

- *Gepaarte Stichproben:* Normalverteilte Daten

$$X_i \sim \mathcal{N}(\mu_X, \sigma_X^2) \qquad \text{und} \qquad Y_i \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

- Betrachten Differenzen

$$D_i = X_i - Y_i$$

- Führen einen t-Test durch
- Normalerweise für die Nullhypothese

$$\mathrm{E}(D) = \mu_D = 0$$

- *Kein* Unterschied!
- Falls Daten nicht normalverteilt $\rightarrow$ Wilcoxontest

- R-Output:

```
vorher <- c(25, 25, 27, 44, 30, 67, 53, 53, 52, 60, 28)
nachher <- c(27, 29, 37, 56, 46, 82, 57, 80, 61, 59, 43)

t.test(nachher, vorher, alternative = "two.sided", mu = 0, paired = TRUE,
    conf.level = 0.95)
##
##  Paired t-test
##
## data:  nachher and vorher
## t = 4.2716, df = 10, p-value = 0.001633
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   4.91431 15.63114
## sample estimates:
## mean of the differences
##              10.27273
```

- Nullhypothese wird auf Signifikanzniveau von 5 % verworfen, da p-Wert 0.001633 kleiner als 0.05

- Unterschied ist also auf dem 5 % Signifikanzniveau signifikant, weil der P-Wert kleiner als 5 % ist
- 95 %-Vertrauensintervall: Mittelwert der Unterschiede

$$[4.91431, 15.63114]$$

- Mit 95 % W'keit ist der Durchschnitt der Differenzen von `nachher` und `vorher` in diesem Bereich

Statistischer t-Test für ungepaarte Stichproben

- *Ungepaarte Stichproben:* Daten X_i und Y_j normalverteilt, aber ungepaart
- Beispiel: Waage A und B
- Zwei-Stichproben t-Test für ungepaarte Stichproben mit Nullhypothese

$$\mu_X = \mu_Y$$

- R-Output:

```
x <- c(79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04, 79.97,
    80.05, 80.03, 80.02, 80, 80.02)
y <- c(80.02, 79.94, 79.98, 79.97, 80.03, 79.95, 79.97)

t.test(x, y, alternative = "two.sided", mu = 0, paired = FALSE,
    conf.level = 0.95)
##
##  Welch Two Sample t-test
##
## data:  x and y
## t = 2.8399, df = 9.3725, p-value = 0.01866
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.008490037 0.073048425
## sample estimates:
## mean of x mean of y
##  80.02077  79.98000
```

- Auf Signifikanzniveau 5 % wird Nullhypothese verworfen, da p-Wert 0.01866 kleiner als 0.05 ist

- Unterschied ist also auf dem 5 % Signifikanzniveau signifikant, weil der P-Wert kleiner als 5 %
- 95 %-Vertrauensintervall: Unterschied in Gruppenmittelwerten:

$$[0.0167, 0.0673]$$

- Mit 95 % W'keit ist der Gruppenmittelwert von `x` um eine Zahl in diesem Bereich grösser als der Gruppenmittelwert von `y`

Mann-Whitney U-Test (aka Wilcoxon Rank-sum Test)

- Falls Daten nicht normalverteilt
- R-Output:

```
x <- c(79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04, 79.97,
    80.05, 80.03, 80.02, 80, 80.02)

y <- c(80.02, 79.94, 79.98, 79.97, 80.03, 79.95, 79.97)

wilcox.test(x, y, alternative = "two.sided", mu = 0, paired = FALSE,
    conf.level = 0.95)

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  x and y
## W = 76.5, p-value = 0.01454
## alternative hypothesis: true location shift is not equal to 0
```