

Die t-Verteilung und die Prüfstatistik

Jonathan Harrington

Standard error of the mean (SE)

ist die Standardabweichung von Mittelwerten

Ich werfe 5 Würfel und berechne den Mittelwert der Zahlen

$m = 3.5$ der wahrscheinlichste Wert

$$SE = \frac{\sigma}{\sqrt{5}}$$

Die Verteilung der Mittelwerte.
Bedeutung: ich werde nicht jedes Mal einen Mittelwert $m = 3.5$ bekommen, sondern davon abweichende Mittelwerte. Der SE ist eine numerische Verschlüsselung dieser Abweichung.

Standard error of the mean (SE)

$$SE = \frac{\sigma}{\sqrt{5}}$$

`sigma()/sqrt(5)`

0.7637626

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \mu^2}$$

`sigma <- function(unten=1, oben=6)
{
 x = unten:oben
 n = length(x)
 m = mean(x)
 sqrt((sum(x^2)/n - m^2))
}`

Standard error of the mean (SE) und Konfidenzintervall

95% Vertrauensintervall

$3.5 - 1.96 * \text{sigma()}/\text{sqrt}(5)$

$$\mu \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

oder

$\text{qnorm}(0.025, 3.5, \text{sigma()}/\text{sqrt}(5))$

2.003025

$\text{qnorm}(0.025)$

$\text{qnorm}(0.975, 3.5, \text{sigma()}/\text{sqrt}(5))$

4.996975

Bedeutung:

Wenn ich 5 Würfel werfe, dann liegt der Stichproben-Mittelwert, m , dieser 5 Zahlen zwischen 2.00 und 5.00 mit einer Wahrscheinlichkeit von 95% (0.95).

Probieren!

$a = \text{proben}(1, 6, 5, 100)$

$\text{sum}(a < 2 \mid a > 5)$

Standard error of the mean (SE) und Konfidenzintervall

SE wird kleiner, umso größer n .

$$SE = \frac{\sigma}{\sqrt{n}} \quad \text{umso größer } n, \text{ umso weniger weicht } m \text{ von } \mu \text{ ab.}$$

Oder: Je mehr Würfel wir werfen, umso wahrscheinlicher ist es/sicherer wird es sein, dass m nah an μ ist. Im unendlichen Fall – wir werfen unendlich viele Würfel und berechnen deren Zahlenmittelwert – ist SE 0 (NULL) und $m = \mu = 3.5$.

Standard error of the mean (SE) **wenn σ unbekannt ist.**

Lenneberg behauptet, dass wir im Durchschnitt mit einer Geschwindigkeit von 6 Silben pro Sekunde sprechen.

Hier sind 12 Werte (Silben/Sekunde) von einem Sprecher.

werte

[1] 6 5 6 9 6 5 6 8 5 6 10 9

Frage: sind die Werte überraschend?
(angenommen $\mu = 6$?).

Präzisere/bessere Frage: ist der Unterschied zwischen μ und m **signifikant**? (Oder: fällt m außerhalb des 95% Vertrauensintervalls von μ ?).

Das Verfahren: **a one-sampled t-test**

Präzisere/bessere Frage: fällt m außerhalb des 95% Vertrauensintervalls von μ ?

A. Um das Vertrauensintervall um μ zu berechnen, benötigen wir den SE.

B. Damit lässt sich ein Vertrauensintervall

$\mu - k SE$ bis $\mu + k SE$ setzen

(k ist eine gewisse Anzahl von SEs).

C. Wenn m (in diesem Fall 6.75) innerhalb dieses Intervalls fällt, ist das Ergebnis 'nicht signifikant' (konsistent mit der Hypothese, dass wir im Durchschnitt mit 6 Silben pro Sekunde sprechen).

A. shut: die Einschätzung von σ

$$SE = \frac{\sigma}{\sqrt{n}} \quad \hat{\sigma} = \sqrt{\frac{\sum x^2}{n-1} - m^2}$$

Aber das können wir nicht berechnen, weil wir σ nicht wissen! Wir können aber $\hat{\sigma}$ oder **unsere beste Einschätzung** von σ berechnen

In R kann $\hat{\sigma}$ ganz einfach mit **sd()** berechnet werden.

Für diesen Fall:

werte

[1] 6 5 6 9 6 5 6 8 5 6 10 9

shut = sd(werte)

A. SEhut: die Einschätzung von SE

werte

[1] 6 5 6 9 6 5 6 8 5 6 10 9

shut = sd(werte)

Einschätzung des Standard-Errors

$$\hat{SE} = \frac{\hat{\sigma}}{\sqrt{n}}$$

SEhut = shut/sqrt(12)

0.5093817

B. Vertrauensintervall: die t-Verteilung

Wenn die Bevölkerungs-Standardabweichung **eingeschätzt** werden muss, dann wird das Vertrauensintervall nicht mit der Normal- sondern der **t-Verteilung** mit einer gewissen Anzahl von **Freiheitsgraden** berechnet.

Die t-Verteilung ist der Normalverteilung recht ähnlich, aber die 'Glocke' und daher das Vertrauensintervall sind etwas breiter (dies berücksichtigt, die zusätzliche Unsicherheit die wegen $\hat{\sigma}$ entsteht).

Bei diesem one-sample t-test ist die Anzahl der Freiheitsgrade, df (degrees of freedom), von der **Anzahl der Werte in der Stichprobe** abhängig: **$df = n - 1$**

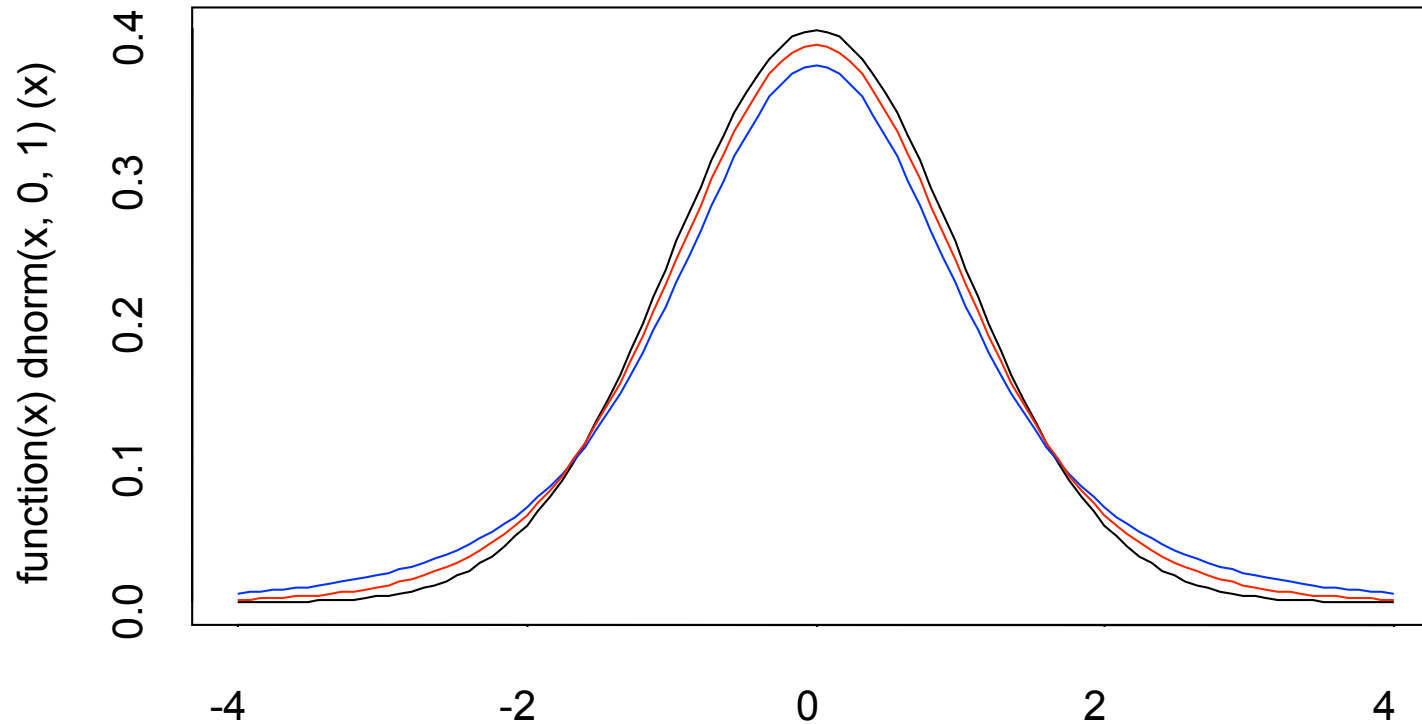
Je höher df, umso sicherer können wir sein, dass $\hat{\sigma} = \sigma$ und umso mehr nähert sich die t-Verteilung der Normalverteilung

Normalverteilung, $\mu = 0$, $\sigma = 1$.

```
curve(dnorm(x, 0, 1), -4, 4)
```

t-Verteilung, $\mu = 0$, $\sigma = 1$, $df = 3$

```
curve(dt(x, 3), -4, 4, add=T, col="blue")
```



```
curve(dt(x, 10), -4, 4, add=T, col="red")
```

B. Vertrauensintervall um $\mu = 6$

`mu = 6`

`n = length(swerte)`

`SEhut = sd(swerte)/sqrt(n) # eingeschätzter SE`

`frei = n - 1 # Freiheitsgrade`

`mu + SEhut * qt(0.025, frei) # untere Grenze`

4.878858

`mu + SEhut * qt(0.975, frei) # obere Grenze`

7.121142

C. Signifikant?

Auf der Basis dieser Stichprobe liegt μ zwischen 4.878858 und 7.121142 mit einer Wahrscheinlichkeit von 95%.

Frage: angenommen $\mu = 6$ sind die Werte überraschend?

`mean(werte)`

`[1] 6.75`

Nein.

The two-sampled t-test

Meistens werden wir **2 Stichprobenmittelwerte** miteinander vergleichen wollen (und wesentlich seltener wie im vorigen Fall einen Stichprobenmittelwert, m , mit einem Bevölkerungsmittelwert, μ).

Zwei Händler, X und Y, verkaufen Äpfel am Markt.

Die Äpfel von Y sind teurer, weil seine Äpfel mehr wiegen (behauptet Y).

Ich kaufe 20 Äpfel von X, 35 von Y. Ich wiege jeden Apfel und berechne:

	X	Y
Gewicht-Mittelwert	$m_x = 200$	$m_y = 220$
Gewicht S-abweichung*	$s_x = 20$	$s_y = 30$
Anzahl	$n_x = 20$	$n_y = 35$

Ist dieser Unterschied $m_x - m_y = 200 - 220 = -20$ g **signifikant**?

* Standardabweichung der Stichprobe, shut wird mit `sd()` in R gemessen

Hypothesen

H0: Es gibt keinen signifikanten Unterschied zwischen den Mittelwerten.

= die Wahrscheinlichkeit, dass der Unterschied zwischen diesen Mittelwerten 0 sein könnte ist mehr als 0.05 (kommt öfter als 5 Mal pro Hundert vor).

H1: Es gibt einen signifikanten Unterschied zwischen den Mittelwerten

= die Wahrscheinlichkeit, dass der Unterschied zwischen diesen Mittelwerten 0 sein könnte ist weniger als 0.05 (kommt seltener als 5 Mal pro Hundert vor).

Vorgang

Wir nehmen an, dass $m_x - m_y = -20$ g **einen Stichprobenmittelwert ist, die einer Normalverteilung folgt.**

1. Wir müssen die Parameter μ , σ (und dann SE) dieser Normalverteilung **einschätzen.**
2. Wir erstellen ein 95% Vertrauensintervall fuer die t-Verteilung.
3. Wenn dieses Vertrauenintervall 0 einschließt, ist H_0 akzeptiert (= kein signifikanter Unterschied zwischen m_x und m_y) sonst H_1 (= der Unterschied ist signifikant).

μ , SE einschätzen

Die beste Einschätzung von μ (μ) ist der Mittelwertunterschied unserer Stichprobe

Fuer diesen Fall $\mu = m_x - m_y = -20$

SE einschätzen

SEhut der Unterschiede zwischen Stichprobenmittelwerten

$$\hat{SE} = \sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} \times \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

	X	Y
Gewicht S-abweichung	sx = 20	sy = 30
Anzahl	nx = 20	ny = 35

```
z = ((nx - 1) * sx^2) + ((ny - 1) * sy^2)
```

```
nenn = nx + ny - 2
```

```
SEhut = sqrt(z/nenn) * sqrt(1/nx + 1/ny)
```

```
[1] 7.525339
```

95% Vertrauensintervall

$$df = nx + ny - 2$$

$$-20 - qt(0.025, df) * SEhut \quad -4.906081$$

$$-20 + qt(0.025, df) * SEhut \quad -35.09392$$

$$\mu = -20$$

$$SEhut = 7.525339$$

Der Unterschied zwischen den Mittelwerten liegt zwischen -35.09392g und -4.906081g mit einer Wahrscheinlichkeit von 95%

Der Unterschied zwischen den Mittelwerten liegt zwischen -35.09392g und -4.906081g mit einer Wahrscheinlichkeit von 95%

Die Wahrscheinlichkeit, dass der Unterschied zwischen den Mittelwerten 0 sein könnte ist daher weniger als 5% (kommt weniger als 5 Mal pro 100 Stichproben vor).

Daher akzeptieren wir H1:

H1: Es gibt einen signifikanten Unterschied zwischen den Mittelwerten

Die benötigten Dauern (Minuten) an 9 Tagen im Winter in die Arbeit zu fahren sind:

20 15 19 22 17 16 23 18 20

Die entsprechenden Dauern an 11 Tagen im Sommer sind:

18 15 17 24 15 12 14 11 13 17 18

Ist der Unterschied zwischen den durchschnittlichen Sommer- und Winterzeiten signifikant ($p < 0.05$)?

$x = c(20, 15, 19, 22, 17, 16, 23, 18, 20)$

$y = c(18, 15, 17, 24, 15, 12, 14, 11, 13, 17, 18)$

Eine R-Funktion **SE2(x,y)** um SEhut zu berechnen

$$\sqrt{\frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}} \times \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}$$

```
SE2 <- function(x, y)
{
  nx = length(x)
  ny = length(y)
  sx = sd(x)
  sy = sd(y)
  num = ((nx - 1) * sx^2) + ((ny - 1) * sy^2)
  den = nx + ny - 2
  sqrt(num/den) * sqrt(1/nx + 1/ny)
}
```

SE2(x, y)

[1] 1.446872

```
x = c(20, 15, 19, 22, 17, 16, 23, 18, 20)
y = c(18, 15, 17, 24, 15, 12, 14, 11, 13, 17, 18)
```

```
# SE
```

```
SEhut =          = SE2(x,y)
```

```
#  $\mu$ 
```

```
d =          mean(x) - mean(y)
```

```
# Anzahl der Freiheitsgrade
```

```
df =          length(x) + length(y) - 2
```

```
# Vertrauensintervall
```

```
d - qt(0.025, df) * SEhut          d + qt(0.025, df) * SEhut
```

```
[1] 6.110471
```

```
[1] 0.03094282
```

Die t-test() Funktion

```
> t.test(x, y, var.equal=T)
```

Die Wahrscheinlichkeit, dass
 $mx - my = 0$ (Null)

data: x and y

$t = 2.1223$, $df = 18$, $p\text{-value} = 0.04794$

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

95% Vertrauensintervall

0.03094282 6.11047132

sample estimates:

mean of x mean of y

18.88889 15.81818

$t=2.1223$ bedeutet: $(mx - my)$ und 0 sind 2.1223

Standardabweichungen voneinander entfernt

$d = \text{mean}(x) - \text{mean}(y)$ SEhut

d/SEhut

$d - 0$

[1] 3.070707

1.446872

[1] 2.122308

Die t-test() Funktion: Formel-Methode

```
xlab = rep("winter", length(x))
```

```
ylab = rep("sommer", length(y))
```

```
d.df = data.frame(Dauer=c(x,y), Saison = factor(c(xlab, ylab)))
```

```
t.test(Dauer ~ Saison, var.equal=T, data=d.df)
```

Kriterien für eine t-test Durchführung

zwei Stichproben, x und y. Sind die Mittelwerte von x und y voneinander signifikant unterschiedlich?

mfdur.txt in "Tabellarische Daten" der Webseite

```
pfad = "etwas"
```

```
mfdur = read.table(file.path(pfad, "mfdur.txt"))
```

```
head mfdur
```

```
duration Gender
1  115.250      F
2   74.687      F
3  124.813      F
```

Kriterien für eine t-test Durchführung

Sind die Verteilungen pro Stufe normalverteilt?

`shapiro.test()`

ja

nein

`wilcox.test()`

Haben die Stufen des Faktors eine ähnliche Varianz?

`var.test()`

nein (Default)

`t.test()`

ja

`t.test(..., var.equal=T)`

```
with(mfdur, tapply(duration, Gender, shapiro.test))
```

Stufe F des Faktors Gender

\$F

Shapiro-Wilk normality test

data: X[[1L]]

W = 0.9866, p-value = 0.9037

Die Wahrscheinlichkeit,
dass die Werte
normalverteilt sind.

\$M

Shapiro-Wilk normality test

data: X[[2L]]

W = 0.9528, p-value = 0.08804

Wenn $p < 0.05$ dann weicht die Stichprobe signifikant von einer Normalverteilung ab, und der t-test sollte nicht eingesetzt werden.

var.test()

prüft ob die Varianzen der beiden Stichproben voneinander signifikant abweichen.

Um signifikante Unterschiede zwischen Varianzen festzustellen, wird ein **F-test** und die **F-Verteilung** verwendet – diese Verteilung ist das gleiche wie die t-Verteilung hoch 2.

die Varianzen der beiden Stufen

`with(mfdur, tapply(duration, Gender, var))`

F	M
428.9193	516.3584

```
var.test(duration ~ Gender, data = mfdur)
```

F test to compare two variances

data: x and y

$F = 0.8307$, num df = 40, denom df = 40, p-value = 0.5601

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.4429731 1.5576553

sample estimates:

ratio of variances

0.830662

Der Unterschied zwischen den Varianzen ist nicht signifikant (genauer: das Verhältnis zwischen den Varianzen weicht nicht signifikant ab von 1).

$F(40, 40) = 0.83, p > 0.05$

Wenn keine Normalverteilung

Wilcoxon Rank Sum and Signed Rank Tests (Mann-Whitney test)

```
wilcox.test(duration ~ Gender, data = mfdur)
```

Wilcoxon rank sum test with continuity correction

data: x and y

W = 1246, p-value = 0.0001727

alternative hypothesis: true location shift is not equal to 0

Der Unterschied zwischen x und y ist signifikant. (Wilcoxon rank sum test, $W = 1246$, $p < 0.001$)

Normalverteilung, Varianzen sind unterschiedlich

```
t.test(duration ~ Gender, data = mfdur)
```

Welch Two Sample t-test

data: x and y

t = 3.6947, df = 79.321, p-value = 0.0004031

alternative hypothesis: true difference in means is not equal to
0

95 percent confidence interval:

8.183973 27.297539

sample estimates:

mean of x mean of y

97.95751 80.21676

Der Unterschied zwischen x und y ist signifikant ($t[79.3] = 3.7$, $p < 0.001$). Oder $t_{79.3} = 3.7$, $p < 0.001$

...sonst

```
t.test(duration ~ Gender, var.equal=T, data = mfdur)
```

Beispiel. t-test Fragen, Frage 3(a)

```
tv = read.table(file.path(pfad, "tv.txt"))  
head(tv)  
with(tv, table(V))
```

```
# boxplot  
with(tv, boxplot(d ~ V, data=tv))
```

```
# Prüfen, ob sie einer Normalverteilung folgen  
with(tv, tapply(d, V, shapiro.test))  
# alles OK
```

```
# Prüfen, ob sich die Varianzen unterscheiden  
var.test(d ~ V, data=tv)  
# oder  
with(tv, var.test(d ~ V))  
# Die Varianzen unterscheiden sich signifikant. Daher:
```

```
t.test(d ~ V, data = tv)
```

Die Vokalkategorie hat einen signifikanten Einfluss auf die Dauer
($t[12.5] = 4.3$, $p < 0.001$)