

# 1 Regression

Wie erstellt man eine Regressionsgleichung in R? Gibt es verschiedene Wege?

Eine Regressionsgleichung:

$$\hat{y}_m = b_0 + b_1 x_m$$

- beachte den Unterschied zwischen Regressionsgleichung und Regressionsmodell (Modellgleichung)!

Regressionsmodell:

$$\hat{y}_m = b_0 + b_1 x_m + e_m$$

In R:

```
lm(y_m ~ x_m)
```

Generell:

```
formel <- formula(y_m ~ x_m)
lm(formel)
```

In welcher Reihenfolge müssen die unabhängigen bzw. abhängigen Variablen R bei der Partialkorrelation übergeben werden?

Für die Berechnung der Partialkorrelationen spielt die Reihenfolge keine Rolle. Sie übergeben `pcor` einen Data Frame mit den entsprechenden Variablen. Es wird im Kontext von Korrelationen auch nicht von unabhängigen und abhängigen Variablen gesprochen. Bei Korrelationen geht man nicht von einem gerichteten Zusammenhang aus.

Partialkorrelation:

```
library(ppcor)
pcor(sub[, c("lz.1", "gs.1", "wm.1")])$estimate
```

	lz.1	gs.1	wm.1
lz.1	1.0000000	0.6248441	-0.1540567
gs.1	0.6248441	1.0000000	0.4260447
wm.1	-0.1540567	0.4260447	1.0000000

```
pcor(sub[, c("gs.1", "wm.1", "lz.1")])$estimate
```

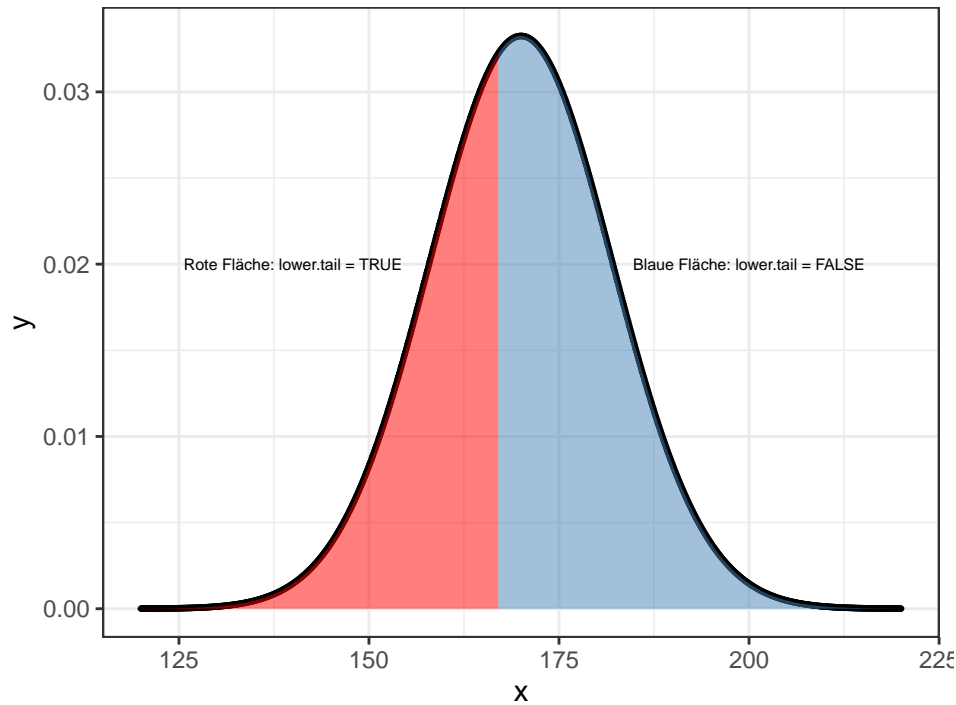
	gs.1	wm.1	lz.1
gs.1	1.0000000	0.4260447	0.6248441
wm.1	0.4260447	1.0000000	-0.1540567
lz.1	0.6248441	-0.1540567	1.0000000

```
# ?ppcor
```

- Output abhängig von der Reihenfolge der Spalten

# 2 Wahrscheinlichkeitstheorie

Warum benutzt man das `lower.tail` Argument? Was sagt dieses Argument in seinen beiden Varianten `TRUE` und `FALSE` genau aus und wofür werden beide Versionen des `lower.tail` Arguments verwendet?



```
pnorm(q = 167, mean = 170, sd = 12, lower.tail = TRUE) # das ist der Default
```

```
[1] 0.4012937
```

```
pnorm(q = 167, mean = 170, sd = 12, lower.tail = FALSE)
```

```
[1] 0.5987063
```

- `lower.tail = TRUE` =  $P(X \leq x)$
- `lower.tail = FALSE` =  $P(X > x)$

**Können `dbinom` und `dnorm` synonymartig verwendet werden? Beziehungsweise, was ist der Unterschied zwischen den beiden Funktionen? (und auch `pbinom`, `rbinom`, `pnorm`, `rnorm`)**

`dbinom` und `dnorm` können nicht synonym verwendet werden. Die zugrundeliegenden Verteilungen sind respektive die Binomialverteilung und die Normalverteilung, die sich fundamental unterscheiden (siehe Vorlesungsfolien).

