

# Evaluation Randomisiert-Kontrollierter Studien und Experimente mit R

## Deskriptive Statistik und Analyse von Dropouts

---

Prof. Dr. David Ebert & Mathias Harrer

Graduiertenseminar TUM-FGZ

Psychology & Digital Mental Health Care, Technische Universität München

# Deskriptive Statistik

---

Es ist sinnvoll, vor der eigentlichen Analyse der Daten deskriptive Statistiken zu “eyeballen,” um u.a. einen Eindruck von der **Verteilung der Daten** zu bekommen.

## Die skim Funktion aus dem package {skimr}

Die skim-Funktion bietet sich zur deskriptiven Statistik an, da sie einen guten Überblick über alle Variablen bietet:

- Zugehörigkeit zu Objektklassen (character oder numeric)
- Anzahl der Missings
- Bei characters: Anzahl unique Elemente
- Bei numerics: Mittelwert, Standardabweichung, Perzentile und Häufigkeitsverteilung

```
library(skimr)
skim(data)
```

---







— Data Summary —

	Values
Name	data
Number of rows	264
Number of columns	33

[...]

---

— Variable type: numeric —

	skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
[...]											
12	child	0	1	0.530	0.500	0	0	1	1	1	
13	sess	132	0.5	5.70	2.32	0	5	7	7	7	
14	cesd.0	0	1	23.6	8.03	4	18	23	28	56	
15	cesd.1	22	0.917	18.5	9.72	1	12	17	24	49	
16	cesd.2	38	0.856	17.6	9.33	1	11	16	24	44	
17	hadsa.0	0	1	11.0	3.36	4	8	11	13	21	
[...]											

**Deskriptive Statistik separat für Interventions- und Kontrollgruppe:**

```
library(skimr)
library(tidyverse)

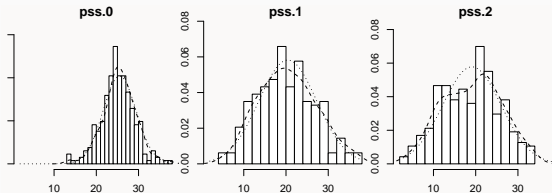
## Interventionsgruppe:
data %>%
  filter(group == 0) %>%
  skim()

## Kontrollgruppe:
data %>%
  filter(group == 1) %>%
  skim()
```

Es ist ratsam, sich die **Häufigkeitsverteilung des primären Outcomes** genauer anzusehen.

Mithilfe der Funktion `multi.hist` aus dem `{psych}` Package können mehrere Histogramme gleichzeitig erstellt werden:

```
library(psych)
library(tidyverse)
multi.hist(data %>% select(pss.0, pss.1, pss.2), ncol = 3)
```



## Analyse der Dropouts

---

**Wie klassischerweise in RCTs zu erwarten, gibt es Missings (fehlende Werte) in unserem Datensatz zur Postmessung und zum Follow-up.**

→ Empfehlung: **Quantifizierung der Missings**, gesamt und nach Gruppe



### Grundlagen

Wird eine Variable in die Funktion `is.na` eingespeist, gibt sie einen Vektor der Klasse `logical` aus, in dem alle Missings als `TRUE` klassifiziert werden. Über die Funktion `sum` werden alle als `TRUE` klassifizierten Werte gezählt, wodurch man die **Anzahl der Missings** erhält.

```
x <- c(1, 2, 3, 4, 5, NA, NA, 200)
sum(is.na(x))
## [1] 2
```

## Cave

Missings in unserem Datensatz fallen patientenweise an, d.h. wenn eine Person einen Fragebogen zu einem Zeitpunkt nicht ausgefüllt wurde, hat sie keinen Fragebogen zu diesem Zeitpunkt ausgefüllt. Deshalb genügt es, wenn wir uns auf die Dropoutanalyse des primary outcome beschränken.

## Missings in Interventions- und Kontrollgruppe:

```
with(data, {  
  c(sum(is.na(pss.0)),  
    sum(is.na(pss.1)),  
    sum(is.na(pss.2)))  
}) -> na.all  
  
## Werte in Prozent umrechnen  
na.all.p <- na.all/nrow(data)
```

## Missings in der Interventionsgruppe:

```
data %>%  
  filter(group == 1) %>%  
  with({  
    c(sum(is.na(pss.0)),  
      sum(is.na(pss.1)),  
      sum(is.na(pss.2)))  
  }) -> na.ig
```

## ## Werte in Prozent umrechnen

```
na.ig.p <- na.ig/nrow(data %>% filter(group == 1))
```

→ Analog für Missings in der Kontrollgruppe!

**Alle Werte in einen dataframe binden:**

```
na <- data.frame(na.all, na.all.p = na.all.p*100,  
                 na.ig, na.ig.p = na.ig.p*100,  
                 na.cg, na.cg.p = na.cg.p*100)
```

**## Zeilenamen des Dataframe ändern**

```
rownames(na) = c("t0", "t1", "t2")
```

```
na
```

```
##      na.all na.all.p na.ig na.ig.p na.cg na.cg.p  
## t0         0      0.00    0    0.00    0    0.00  
## t1        21      7.95   16   12.12    5    3.78  
## t2        28     10.60   17   12.87   11    8.33
```

→ Die Dropout-Rate ist zum Follow-up mit 10.6% am höchsten und es haben mehr Teilnehmende aus der IG als aus der KG die Studie vorzeitig abgebrochen.

## Referenzen

---

