

# Evaluation Randomisiert-Kontrollierter Studien und Experimente mit R

## Missing Data

---

Prof. Dr. David Ebert & Mathias Harrer

Graduiertenseminar TUM-FGZ

Psychology & Digital Mental Health Care, Technische Universität München

## Missing Data-Mechanismen

---

*“Obviously the best way to treat missing data is not to have them.”*

Orchard & Woodbury (1972)

Fehlende Werte sind aber, insbesondere in der medizinischen und psychologischen Forschung, häufig **unvermeidbar!**

## Taxonomie: Gründe für fehlende Werte in klinischen Studien

- **Instrumente:** Response Burden, zu langes Assessment, ...
- **Teilnehmende:** Überforderung, Privatsphäre-Bedenken, Motivationsprobleme, ...
- **Center:** Inadäquate Umsetzung des Studienmethodik, Personalmangel, ...
- **Personal:** Falsche Datenerfassung, Datenverlust, falsche Dateneingabe, ...
- **Studie:** Technische Fehler, Zeitverzögerungen bei Follow-Ups, ...

Palmer et al. (2018)

**Der Umgang mit fehlenden Werten in der (medizinisch-psychologischen) Forschung ist oftmals mangelhaft:** (Akl et al., 2015; Van Buuren, 2018, Kapitel 1.1.2; Wood et al., 2004)

- Verteilung fehlender Werte nicht transparent berichtet
- Unpassendes missing data handling (z.B. **listwise deletion**)
- Adäquate Imputationsmethoden (z.B. MI, FIML) häufig nicht benutzt, inadäquat angewendet, oder unzureichend berichtet

**R macht es nicht (automatisch) “richtig!”**

```
y <- 1:10
x <- c(1, NA, NA, NA, 3, 5, 8, 10, -1, 10)
summary(lm(y ~ x))
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.9403      1.7738   2.785   0.0387 *
## x             0.3172      0.2709   1.171   0.2945
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.904 on 5 degrees of freedom
## (3 observations deleted due to missingness)
```

## Was dagegen tun?

- Umsetzung von **Reporting-Standards**; z.B. STROBE (Von Elm et al., 2007) & CONSORT (Schulz et al., 2010), s.a. Sterne et al. (2009).
- Adäquates **Missing Data Handling**
  - Basierend auf plausiblen Annahmen, warum fehlende Werte entstanden sind (“Missing Data Mechanism”)
  - ggf. Schätzung (Imputation) fehlender Werte unter Einbezug von deren Unsicherheit



Donald B. Rubin

**Grundannahme:** Das Fehlen oder Vorhandensein von Daten ist Resultat eines probabilistischen Prozesses.

Diesem Prozess versucht man sich durch ein Modell (**missing data model**) anzunähern.

(Rubin, 1976)





(Harrer et al., 2021)

Nach Rubin (1976) können Missing Data-Mechanismen in 3 Untertypen klassifiziert werden:

## MCAR

Missing Completely At Random: rein zufällig fehlende Werte einer Variable.

## MAR

Missing At Random: das Fehlen von Werten einer Variable ist abhängig von anderen (observierten) Variablen  $\psi$ .

## MNAR

Missing Not At Random / "Nonignorable Missing Data": das Fehlen von Werten einer Variable ist (u.A.) abhängig von der Variable selbst.

→ Für jede Annahme ergeben sich unterschiedliche Auswirkungen bei der Datenauswertung!

## Beispiele:

### MCAR

Missing Completely At Random: rein zufällig fehlende Werte einer Variable.

### MAR

Missing At Random: das Fehlen von Werten einer Variable ist abhängig von anderen (observierten) Variablen  $\psi$ .

### MNAR

Missing Not At Random / "Nonignorable Missing Data": das Fehlen von Werten einer Variable ist (u.A.) abhängig von der Variable selbst.

→ Für jede Annahme ergeben sich unterschiedliche Auswirkungen bei der Datenauswertung!

## Second Section

---

### Bulleted Lists

- Element A
- Element B
  - B.1
  - B.2
- Element C

## Elements

---

The theme provides sensible defaults to `\emph{emphasize}` text, `\alert{accent}` parts or show `\textbf{bold}` results.

In Markdown, you can also use `_emphasize_` and `**bold**`.

becomes

The theme provides sensible defaults to *emphasize* text, **accent** parts or show **bold** results (Arendt, 1989; Bandersson & Cuijpers, 2009; Smit et al., 2006)

.

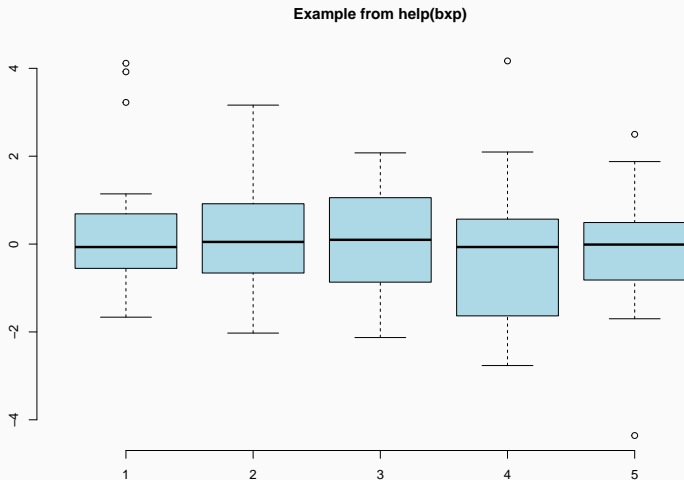
In Markdown, you can also use *emphasize* and **bold**.

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$



The following code generates the plot on the next slide (taken from `help(bxp)` and modified slightly):

```
library(stats)
set.seed(753)
bx.p <- boxplot(split(rt(100, 4),
                      gl(5, 20)), plot=FALSE)
bxp(bx.p, notch = FALSE, boxfill = "lightblue",
     frame = FALSE, outl = TRUE,
     main = "Example from help(bxp)")
```



A simple `knitr::kable` example:

```
knitr::kable(mtcars[1:5, 1:8],  
             caption="(Parts of) the mtcars dataset")
```

**Table 1:** (Parts of) the mtcars dataset

	mpg	cyl	disp	hp	drat	wt	qsec	vs
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0

As one example of falling back into  $\LaTeX$ , consider the example of three different block environments are pre-defined and may be styled with an optional background color.

**Default**

Block content.

**Alert**

Block content.

**Example**

Block content.

## For more information:

- See the [Metropolis repository](#) for more on Metropolis
- See the [RMarkdown repository](#) for more on RMarkdown
- See the [binb repository](#) for more on binb
- See the [binb vignettes](#) for more examples.

- Akl, E. A., Shawwa, K., Kahale, L. A., Agoritsas, T., Brignardello-Petersen, R., Busse, J. W., Carrasco-Labra, A., Ebrahim, S., Johnston, B. C., Neumann, I., & others. (2015). Reporting missing participant data in randomised trials: Systematic survey of the methodological literature and a proposed guide. *BMJ Open*, 5(12), e008431.
- Arendt, H. (1989). *Vom leben des geistes, band 2, das wollen. Aus dem amerikanischen von hermann vetter*. Piper.
- Bandersson, G., & Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: A meta-analysis. *Cognitive Behaviour Therapy*, 38(4), 196–205.
- Harrer, M., Cuijpers, P., A, F. T., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide* (1st ed.). Chapman & Hall/CRC Press.
- Orchard, T., & Woodbury, M. A. (1972). A missing information principle: Theory and applications. In *Theory of statistics* (pp. 697–716). University of California Press.
- Palmer, M. J., Mercieca-Bebber, R., King, M., Calvert, M., Richardson, H., & Brundage, M. (2018). A systematic review and development of a classification framework for factors associated with missing patient-reported outcome data. *Clinical Trials*, 15(1), 95–106.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592.

- Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Trials*, 11(1), 1–8.
- Smit, H., Cuijpers, P., Oostenbrink, J., Batelaan, N., Graaf, R. de, & Beekman, A. (2006). Costs of nine common mental disorders: Implications for curative and preventive psychiatry. *The Journal of Mental Health Policy and Economics*, 9(4), 193–200.
- Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *Bmj*, 338.
- Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC press.
- Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Bulletin of the World Health Organization*, 85, 867–872.
- Wood, A. M., White, I. R., & Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials*, 1(4), 368–376.