Lineare Gemischte Modelle (6) Poweranalyse & Modelldiagnostik

Power im LMM

- Angenommen, die Testung einer Schüler*in kostet 10€, und man hat 10.000€ zur Verfügung ...
  - 40 Klassen mit je 25 Schüler*innen?
  - 100 Klassen mit je 10 Schüler*innen?
- 3 Levels:
  - Jeweils 10 Klassen aus 10 Schulen mit je 10 Schüler*innen?
  - Jeweils 20 Klassen aus 5 Schulen mit je 10 Schüler*innen?
  - Jeweils 5 Klassen aus 10 Schulen mit je 20 Schüler*innen?
- Trade-off: Jede Schule kostet 500€ Zusatzgebühr (Anfahrt). Soll ich 50 Schüler*innen weniger erheben, dafür aber eine Schule extra auf L3 aufnehmen?

Power im LMM: Schwierigkeiten

- Stichprobengrößen auf mehreren Ebenen
- Man benötigt a-priori Abschätzungen für Varianzkomponenten („Wie viel Varianz in den Intercepts erwarte ich?“) → schwer abzuschätzen, meist unbekannt
- Möglicherweise liegen Pilotdaten zur Abschätzung vor, allerdings sind diese meist zu klein um eine sinnvolle Abschätzung liefern zu können.
- Das Einführen von Kovariaten kann die optimalen Stichprobengrößen auf jedem Level ändern, je nachdem wie viel Varianz eine Kovariate innerhalb oder zwischen Gruppen erklärt.

Power im LMM: Einfluss ICC

- Zusätzlich: Abhängigkeit vom ICC
  - Wenn ICC gegen 1 geht, unterscheiden sich die Einheiten innerhalb einer Gruppe kaum

    → jede einzelne L1-Einheit bietet kaum zusätzliche Information über die bereits vorhandenen L1-Einheiten hinaus

    → mehr L1-Einheiten bringen keinen Gewinn an zusätzlicher Information; es hilft hier nur, mehr L2-Einheiten zu messen
- „ICC values typically range between .05 and .20“ (Snijders & Bosker, 1999; Bliese, 2000)
- „values greater than .30 will be rare“ (Bliese, 2000)
- „median value of .12; this is likely an overestimate“ (James, 1982)

→ „a value between .10 and .15 will provide a conservative estimate of the ICC when it cannot be precisely computed“

Scherbaum, C. A., & Ferreter, J. M. (2008). Estimating Statistical Power and Required Sample Sizes for Organizational Research Using Multilevel Modeling. Organizational Research Methods, 12(2), 347–367. http://doi.org/10.1177/1094428107308906

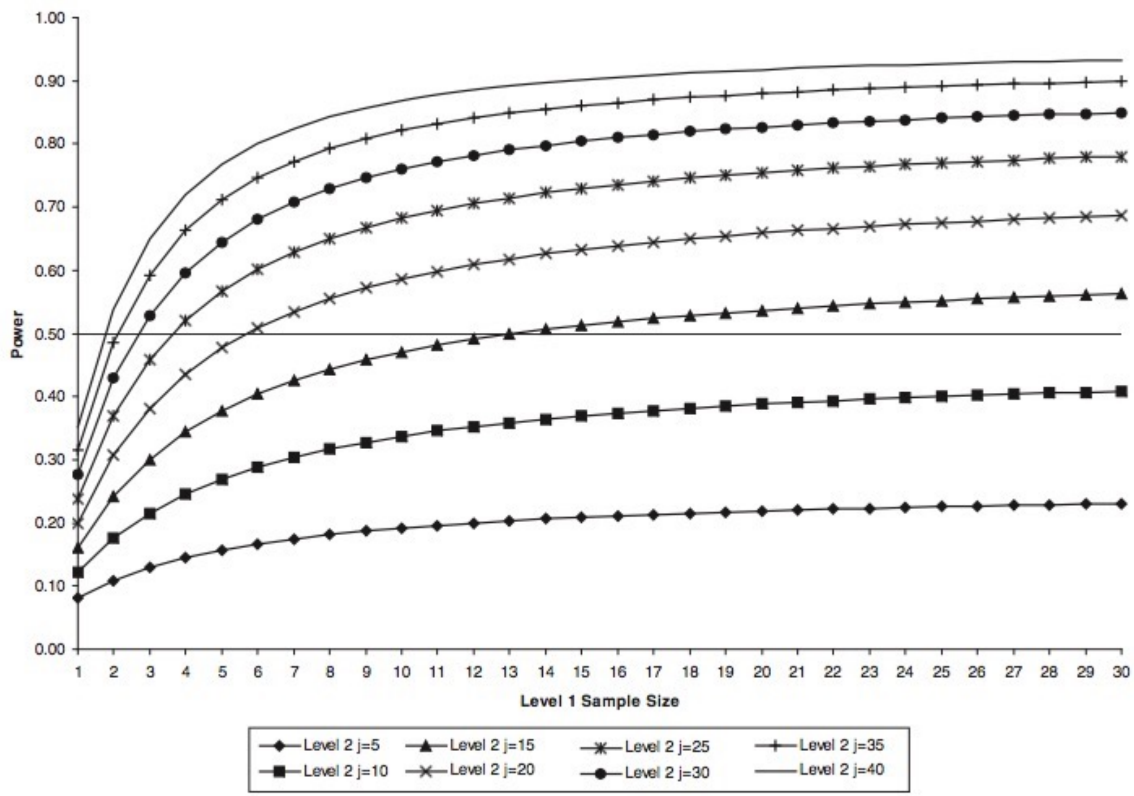
Power im LMM

- Level-Abhängigkeit (wenn man 2 Level annimmt):
  - Für die Power von **L1-Prädiktoren** ist die Anzahl der L1-Einheiten (also z.B. Schüler*innen) relevant
  - Für die Power von **L2-Prädiktoren** ist die Anzahl der L2-Einheiten (also z.B. Schulen) sowie die Gruppengröße (also z.B. die Klassengröße) relevant.

Ganz grobe Faustregeln (werden aber oft kritisiert!)

- Generell: *n* auf L2 (bzw. auf dem höchsten Level) ist wichtiger als das *n* auf L1 (bzw: auf den niedrigeren Leveln)

Figure 1
Estimates of Statistical Power at a Medium Effect Size for γ_{01} under Varying Sample Sizes at Level 1 and Level 2



Scherbaum, C. A., & Ferreter, J. M. (2008). Estimating Statistical Power and Required Sample Sizes for Organizational Research Using Multilevel Modeling. Organizational Research Methods, 12(2), 347–367. http://doi.org/10.1177/1094428107308906

Ganz grobe Faustregeln (werden aber oft kritisiert!)

- Generell: *n* auf L2 (bzw: auf dem höchsten Level) ist wichtiger als das *n* auf L1 (bzw. auf den niedrigeren Leveln)
- Entscheidend ist v.a. das Level, auf dem der fokale Effekt liegt.
- Kreft (1996): 30/30 Regel
    - >= 30 Gruppen auf L2
    - >= 30 L1-Einheiten pro L2-Gruppe

  → 30 * 30 = 900 Einheiten auf L1
- Hox (1998): 50/20 Regel
    - >= 50 Gruppen auf L2
    - >= 20 L1-Einheiten pro L2-Gruppe

  → 50 * 20 = 1000 Einheiten auf L1

Bsp: 3-Level-Therapiestudie

- L1: Messzeitpunkte (5, 11, oder 21 pro Patient*in), geschachtelt in ...
- L2: Patient*innen (2, 4, oder 8 pro Therapeut*in), geschachtelt in...
- L3: Therapeut*innen (Fokus der Power-Analyse: Wie viele Therapeut*innen akquiriert man?)
- AV auf L1: Therapiefortschritt (Selbstbericht der Patient*innen)
- Fokaler Effekt: Therapeut*innen bekommen Feedback über Therapiefortschritt oder nicht (d.h., between-group Experimentalfaktor auf L3) → Effekt auf L1-Slope.
  - Ist der Therapiefortschritt über die Zeit (rate of change) schneller wenn die Therapeut*innen Feedback über den aktuellen Status bekommen?
  - Also: Ein L1-L3-Interaktionseffekt

de Jong, K., Moerbeek, M., & van der Leeden, R. (2010). A priori power analysis in longitudinal three-level multilevel models: An example with therapist effects. Psychotherapy Research, 20(3), 273–284. http://doi.org/10.1080/10503300903376320

Bsp: 3-Level-Therapiestudie

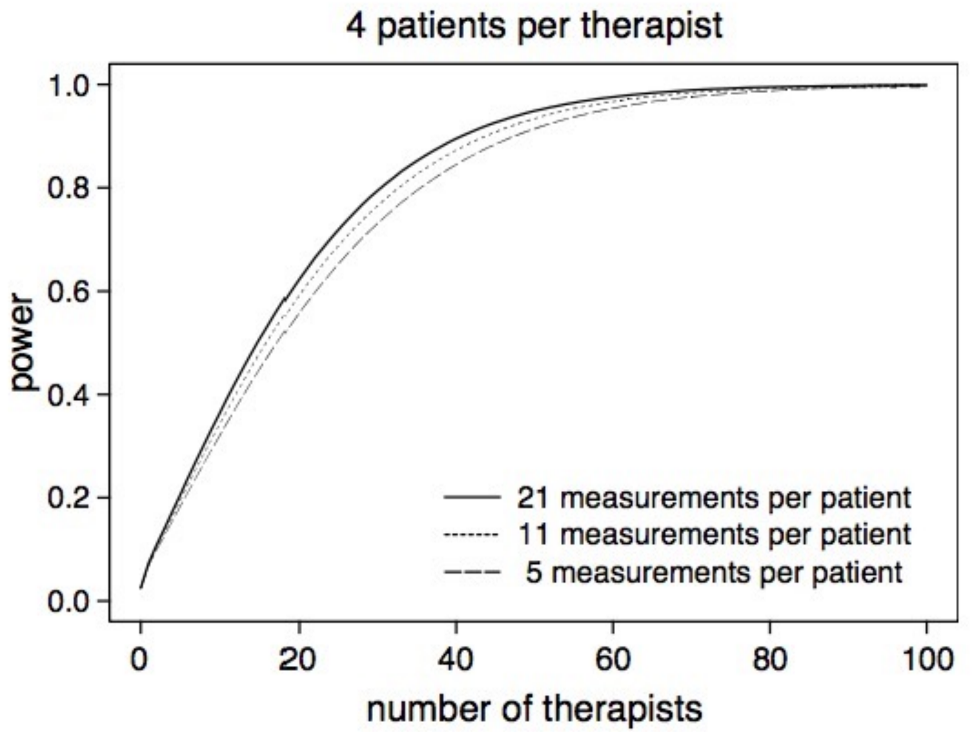


Figure III. Estimated power for a varying number of measurements per patient, randomization at the patient level.

→ Sättigung ab ca. 60 Therapeut*innen:
mehr braucht es nicht, wenn jede Therapeut*in 4 Patient*innen hat

→ Anzahl des Messzeitpunkte auf L1 relativ irrelevant

de Jong, K., Moerbeek, M., & van der Leeden, R. (2010). A priori power analysis in longitudinal three-level multilevel models: An example with therapist effects. Psychotherapy Research, 20(3), 273–284. http://doi.org/10.1080/10503300903376320

Bsp: 3-Level-Therapiestudie

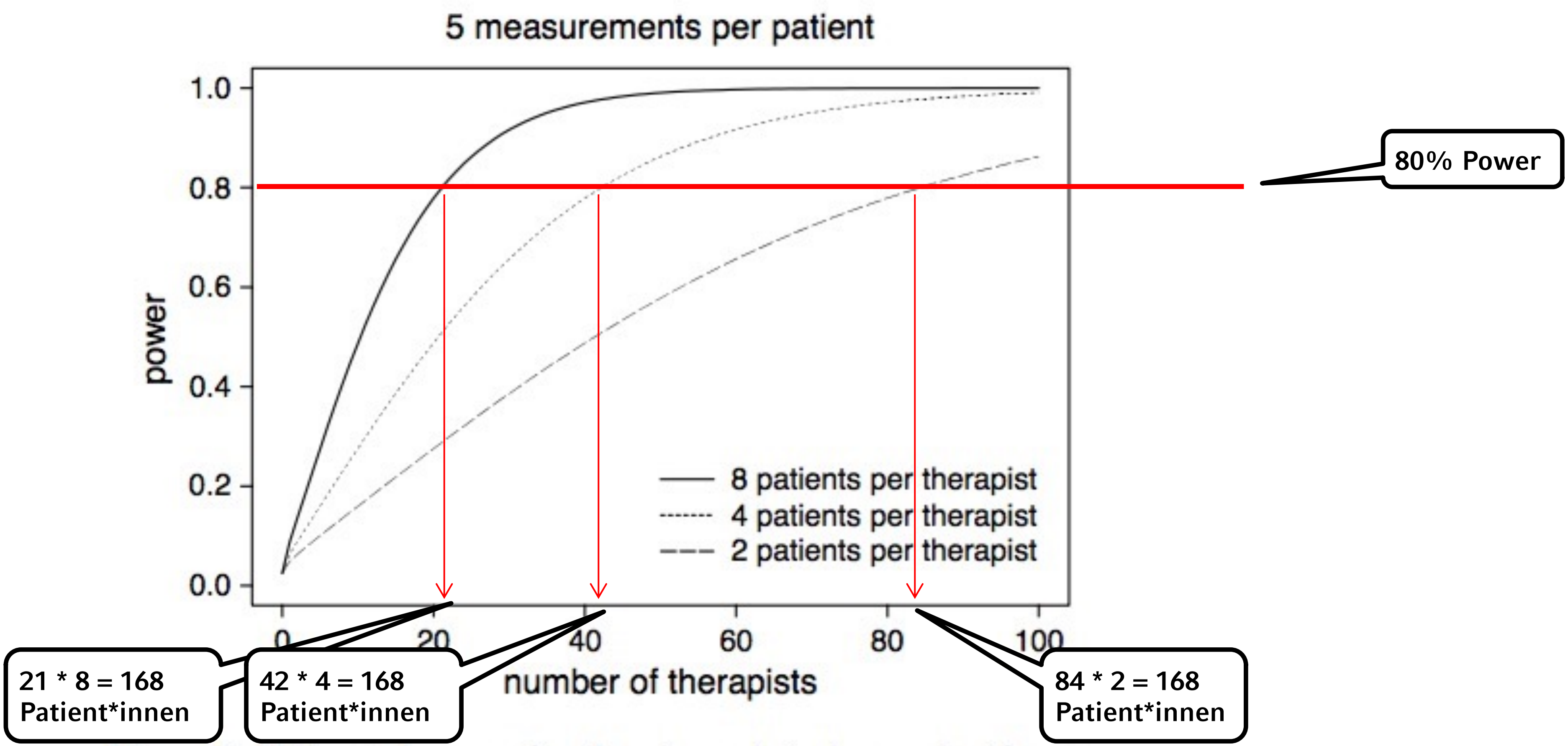
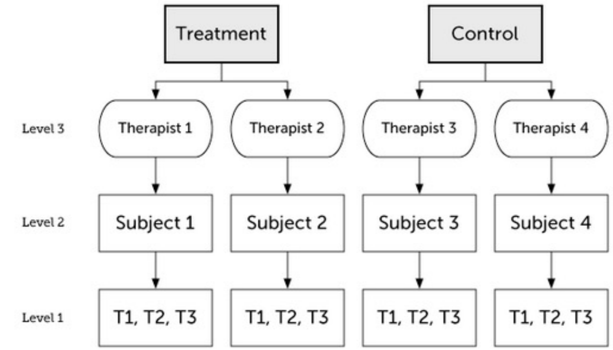


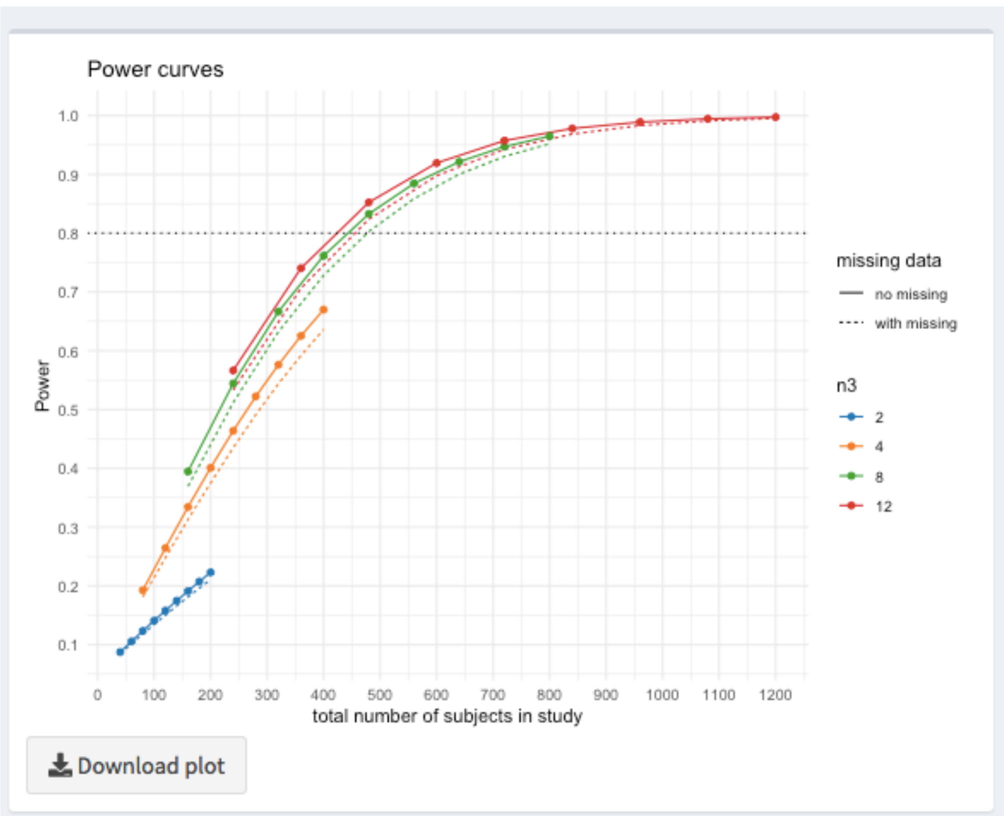
Figure I. Estimated power for the planned study, no significant slope variance at the therapist level.

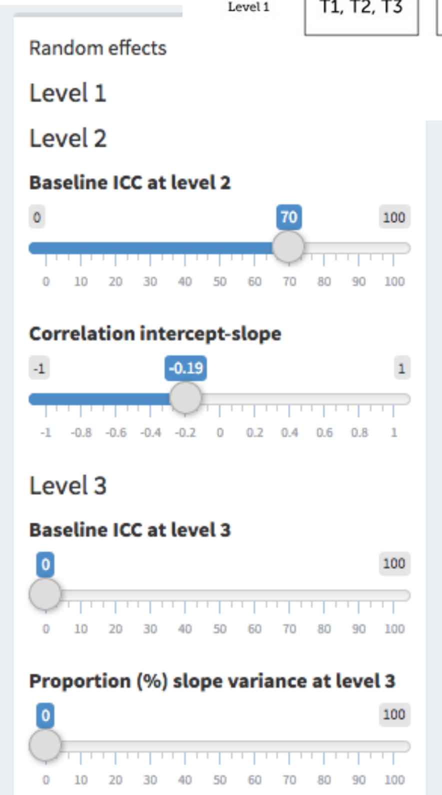
de Jong, K., Moerbeek, M., & van der Leeden, R. (2010). A priori power analysis in longitudinal three-level multilevel models: An example with therapist effects. Psychotherapy Research, 20(3), 273–284. http://doi.org/10.1080/10503300903376320

Erste Abschätzungen mit R

- Paket *powerlmm*
- Inklusive Shiny-Application
- Allerdings nur für eine bestimmte Klasse an Standarddesigns







Maßgeschneiderte Simulationen in R

- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
  Kapitel 20.5

- Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd edition). Boston: Academic Press.
  Kap. 13

- McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan (2nd edition).* New York, CRC Press Taylor & Francis Group.
  Keine Powerberechnungen, aber Daten Simulationen sehr ausführlich behandelt.

- Link zum „Self-paced tutorial: Advanced Power Analysis by Simulation“ von 2023:
  https://lmu-osc.github.io/Simulations-for-Advanced-Power-Analyses/
  Ch. 4: Linear Mixed Models / Multilevel models

Modelldiagnostik &
ein paar praktische Tipps

Residualdiagnostik: Überprüfung der Annahmen

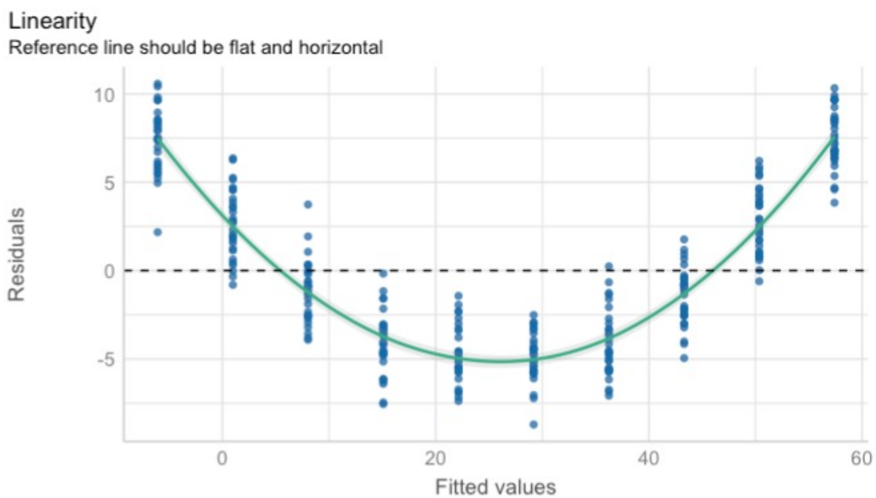
- Annahme 1 (Verteilung der Residuen): Die Residuen sind normalverteilt
- Annahme 2 (Homoskedastizität): Die Fehlervarianz σ^2 ist konstant
- Annahme 3 (Linearität): Die Prädiktoren hängen linear mit dem Kriterium zusammen

```
Formula: y ~ 1 + x + (1 + x | pid)

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)    
(Intercept) -13.04927    0.58144 298.00000  -22.44   <2e-16 ***
x             7.03927    0.09371 298.00000   75.12   <2e-16 ***
```

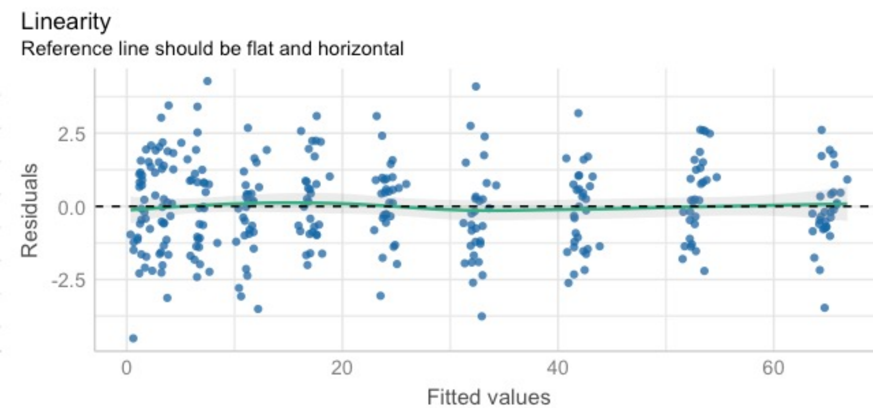
```
Formula: y ~ 1 + x + x2 + (1 + x | pid)

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)    
(Intercept) 5.513e-01  2.235e-01 1.141e+02   2.466  0.01513 *  
x           2.390e-01  8.407e-02 2.621e+02   2.843  0.00482 ** 
x2          6.182e-01  7.404e-03 2.390e+02  83.498  < 2e-16 ***
```

Annahme: Linearität

y ~ 1 + x



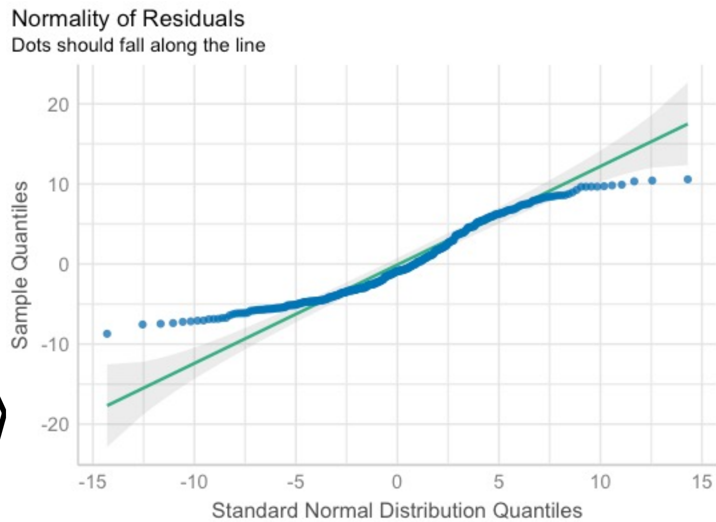
y ~ 1 + x + x2



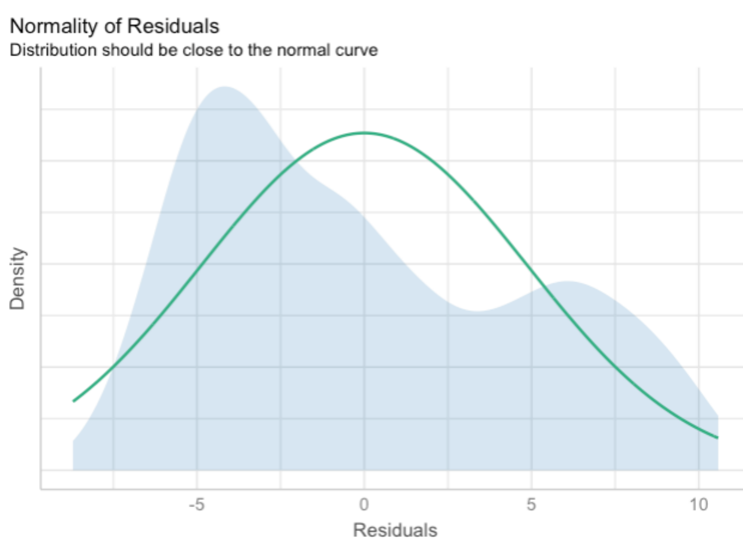
Die Residuen sollten unsystematisch um die gestrichelte 0-Linie herum schwanken

Annahme: Normalverteilung der Residuen

y ~ 1 + x

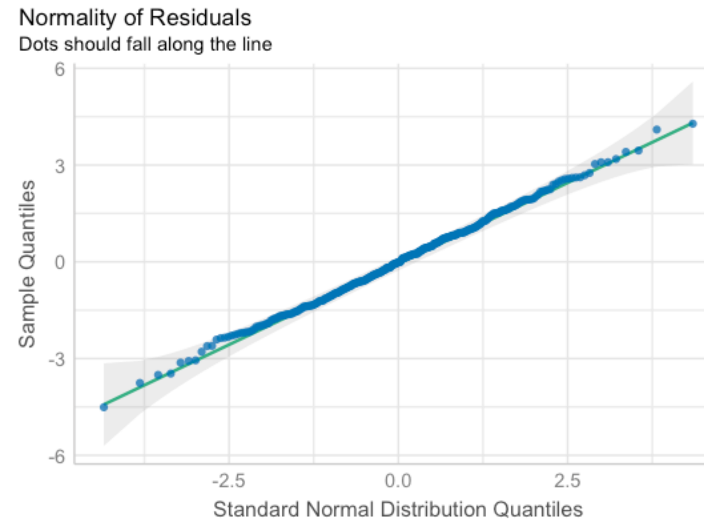


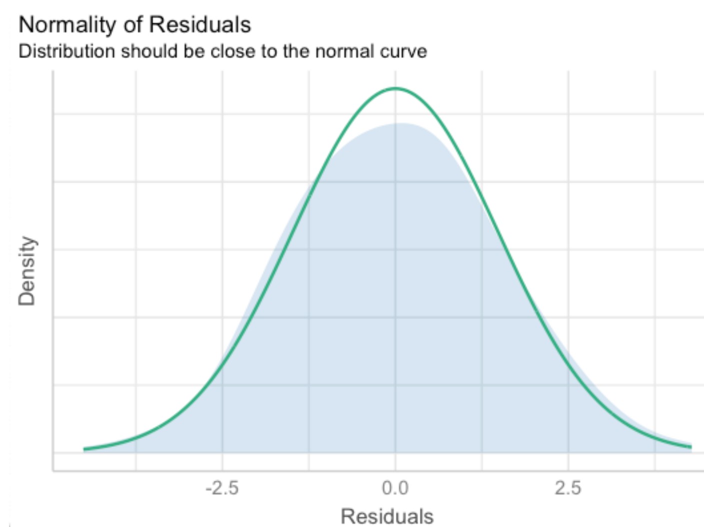
Q-Q Plot = Quantile-quantile-plot: Quantile der Normalverteilung vs. Quantile der empirischen Verteilung der Residuen. <u>Sollten auf der grünen Linie liegen, innerhalb der grauen Konfidenzbänder</u>



Alternative Darstellung als Verteilung

y ~ 1 + x + x2

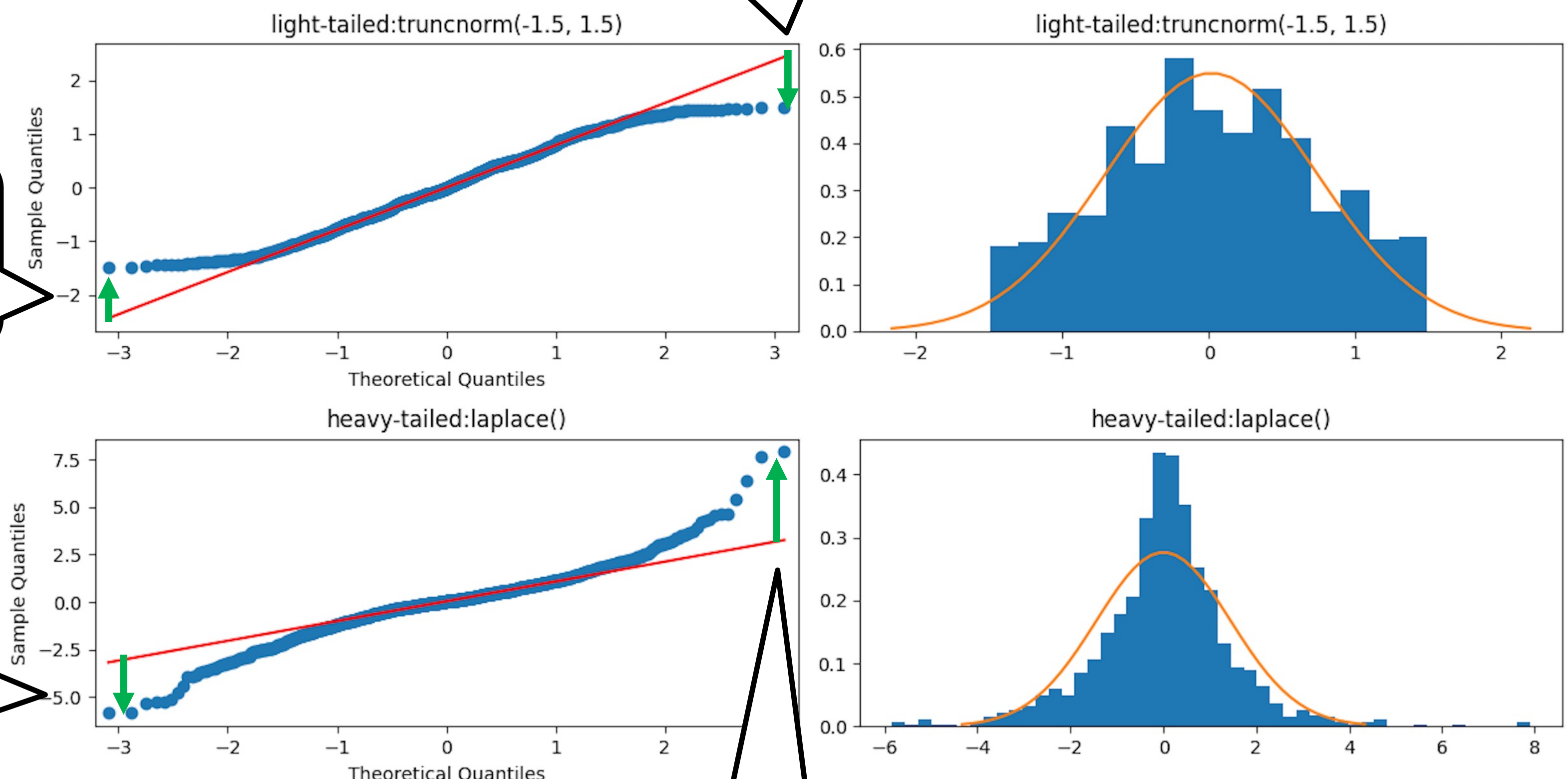




Interpretation QQ-plots

… und die oberen Quantile nicht positiv genug.

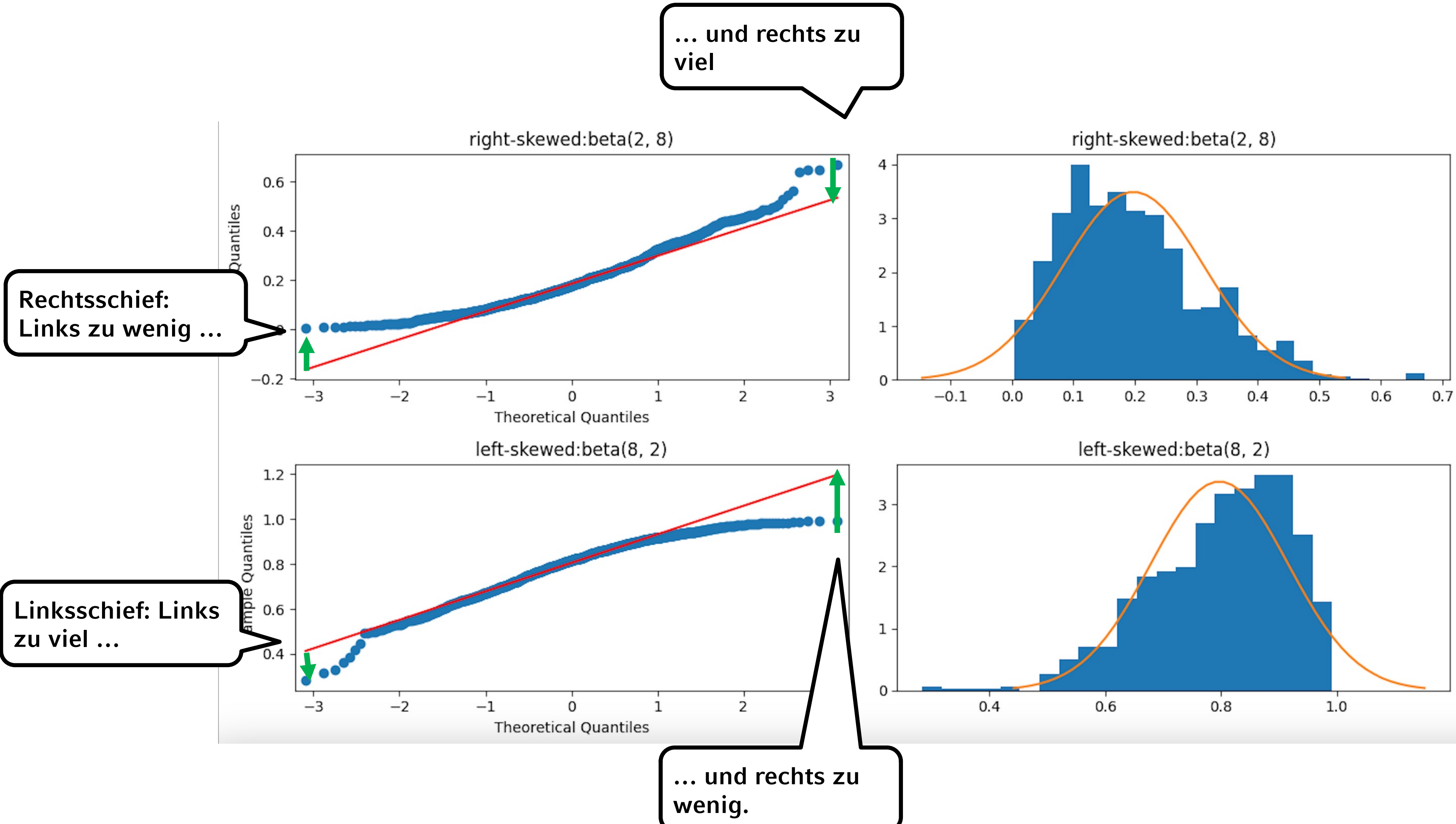
Empirisch sind die unteren Quantile nicht negativ genug …



Empirisch sind die unteren Quantile zu negativ …

… und die oberen Quantile zu positiv.

Interpretation QQ-plots: Im Vgl. zur Standardnormlaverteilung

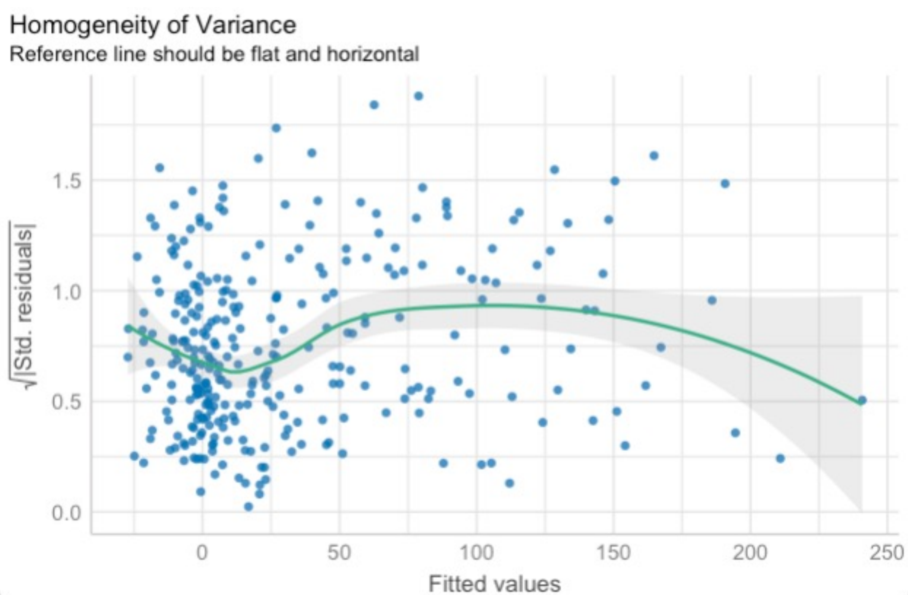


Plots von https://stats.stackexchange.com/a/644971,

Annahme: Homoskedastizität

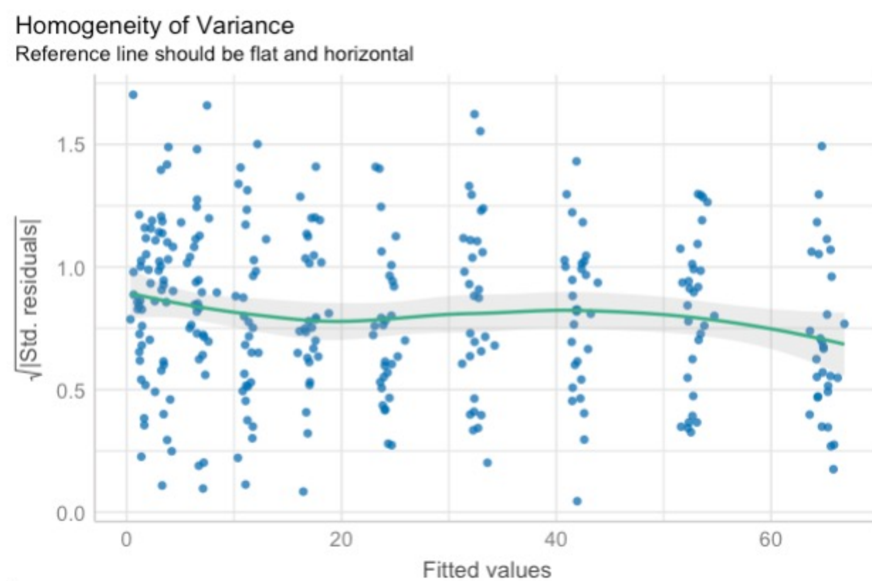
y ~ 1 + x + x2, heterogene Simulation

Warning: Heteroscedasticity (non-constant error variance) detected ($p < .001$)



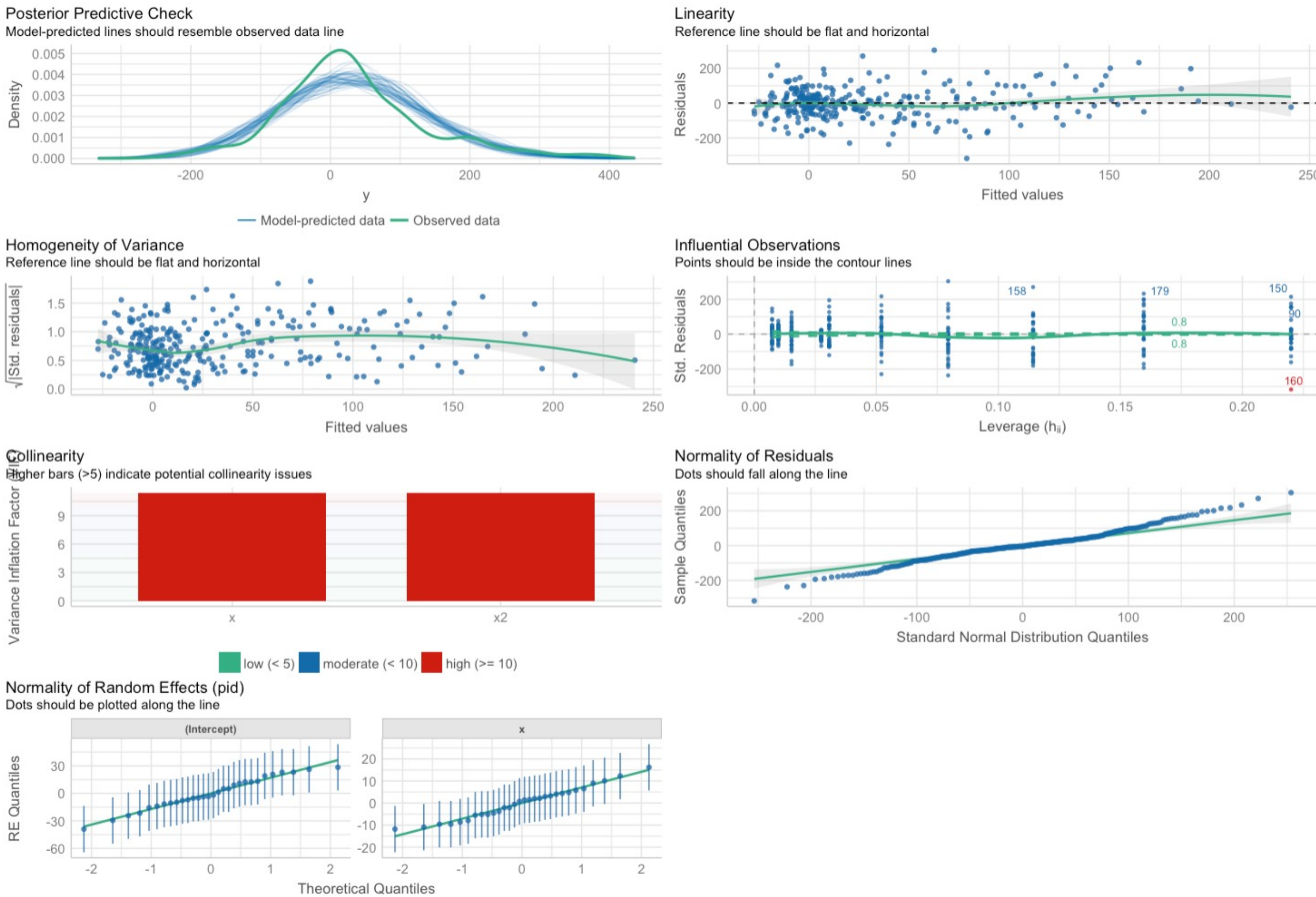
y ~ 1 + x + x2, homogene Simulation

OK: Error variance appears to be homoscedastic ($p = 0.381$).




```
library(performance)
l2 <- lmer(…)
check_heteroscedasticity(l2)
plot(check_heteroscedasticity(l2))
```

Gesamtüberblick




```
library(performance)
l2 <- lmer(…)
check_model(l2)
```

Praktische Tipps

- „model failed to converge“ Warnung
  - lme4 konnte auf keine Lösung konvergieren. **Selbst wenn ein Ergebnis zurückgegeben wird, sollten Sie dieses <u>nicht</u> interpretieren**! Die Werte können völlig daneben liegen.
  - Lösungsansätze:
    - Modell vereinfachen (siehe „parsimonious models“ in VL-Sitzung 6)
    - Variablen zentrieren/standardisieren
    - Weitere Lösungsansätze: siehe `?convergence` im lme4-Paket in R
- „singular fit“ Warnung
  - Bedeutet, dass mindestens eine random variance Komponente Null ist, oder dass eine Korrelation zwischen random effects -1 oder +1 ist.
  - Ist nicht unbedingt ein Problem, aber ein Hinweis darauf, dass man das Modell noch mal genau ansehen sollte, und evtl. im Sinne von Ockhams Rasiermesser das Modell vereinfachen könnte, indem man den nicht benötigten random term aus dem Modell herausnimmt.
  - Könnte auch an hohen Kollinearitäten von Prädiktoren liegen.
  - siehe auch: https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html#singular-models-random-effect-variances-estimated-as-zero-or-correlations-estimated-as---1

Das war‘s mit den linearen gemischten Modellen!