Visual Inference for Graphcial Diagnostic of Linear Mixed Models

Supervisor: Dr Emi Tanaka & Prof Di Cook

Kaiwen Jin

Monash University

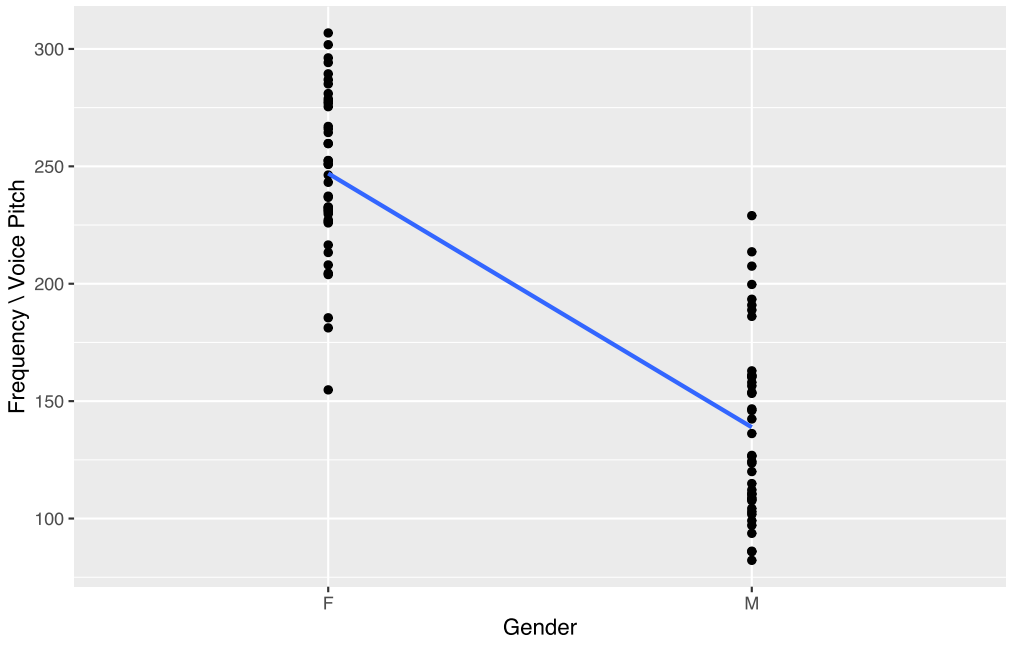
2020-09-18

Linguistic Case

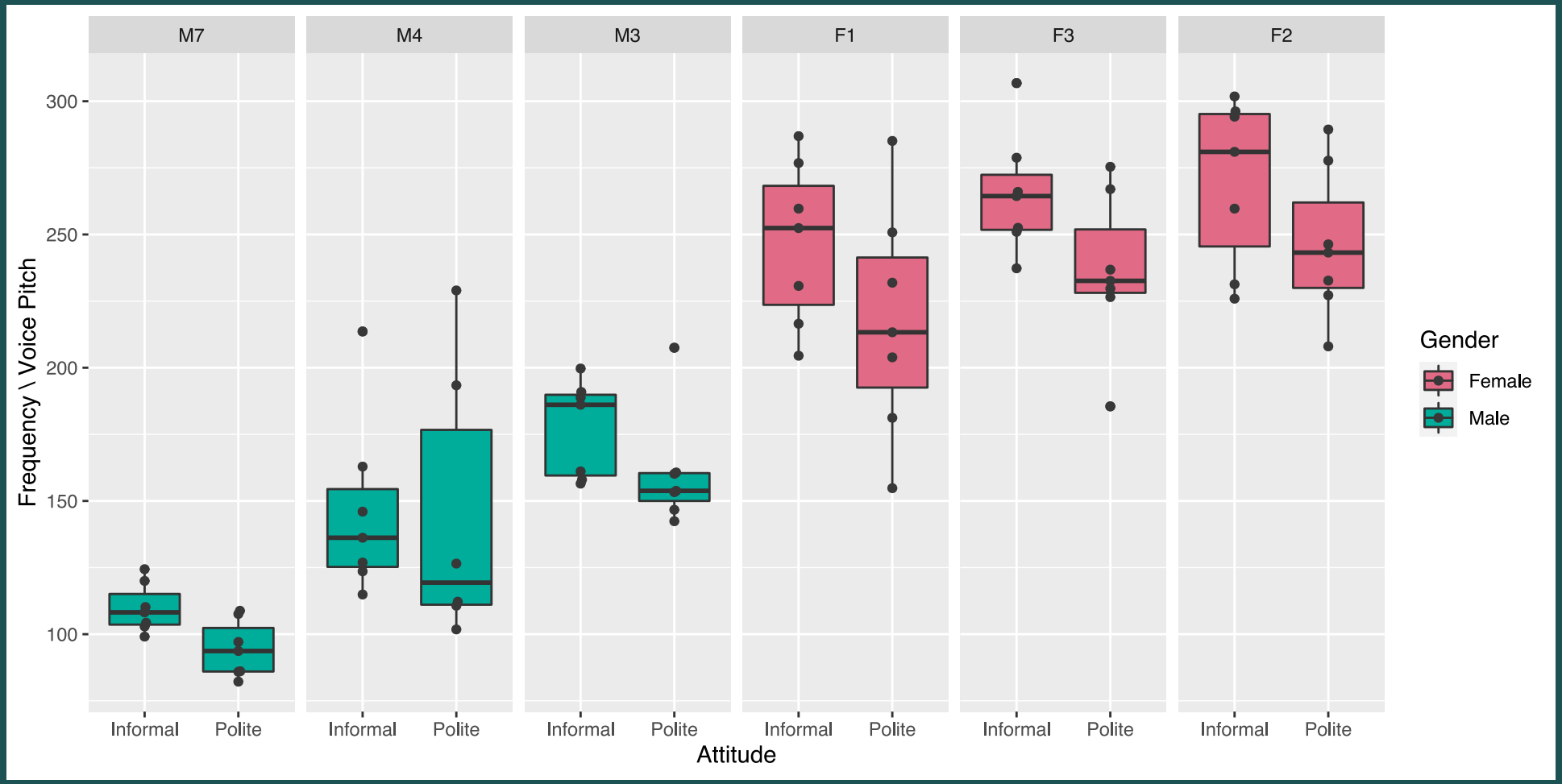
Data

- Gender (female or male)
- Attitude (informal or polite)
- 6 subjects (3 male and 3 female)
- 7 scenarios (such as excusing for coming too late)
- Frequency (also called voice pitch)

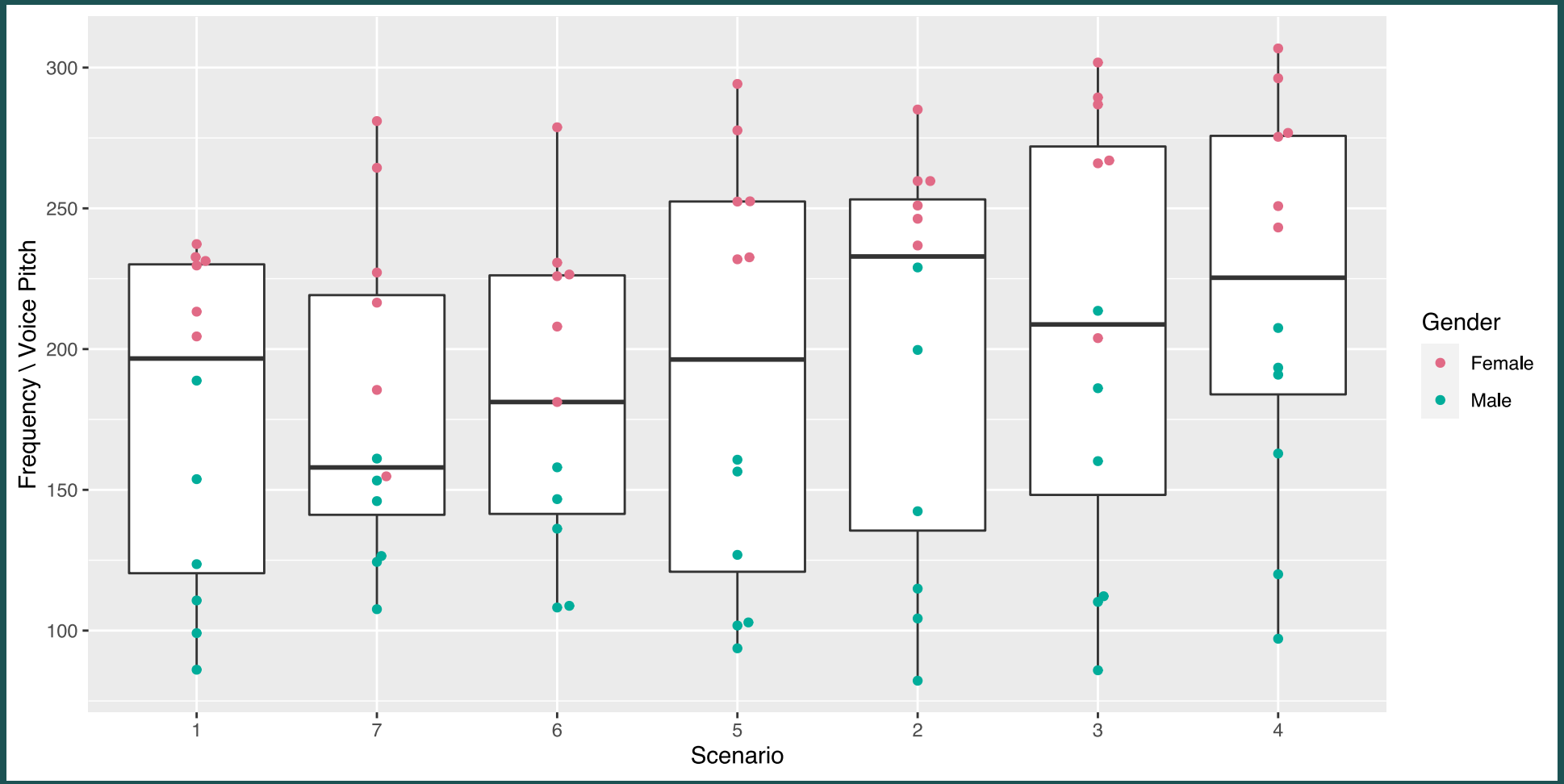
frequency ~ gender + ε



With different subjects



With different scenarios



Linear Mixed Model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{e}$$

where

- $\mathbf{y}$ is a $\mathbf{N} \times \mathbf{1}$ vector of observations, outcome variable
- $\mathbf{X}$ is a $\mathbf{N} \times \mathbf{p}$ matrix
- $\boldsymbol{\beta}$ is a $\mathbf{p} \times \mathbf{1}$ vector of the fixed effect
- $\mathbf{Z}$ is a $\mathbf{N} \times \mathbf{q}$ matrix
- $\mathbf{b}$ is a $\mathbf{q} \times \mathbf{1}$ vector of the random effect

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{e} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Omega} = \boldsymbol{Z}\boldsymbol{\Gamma}\boldsymbol{Z}^{\top} + \boldsymbol{R})$$

How can we implement the LME?

- `lmer` function from `lme4` package
- `mmer` function from `sommer` package

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Graphical diagnostic on residual analysis

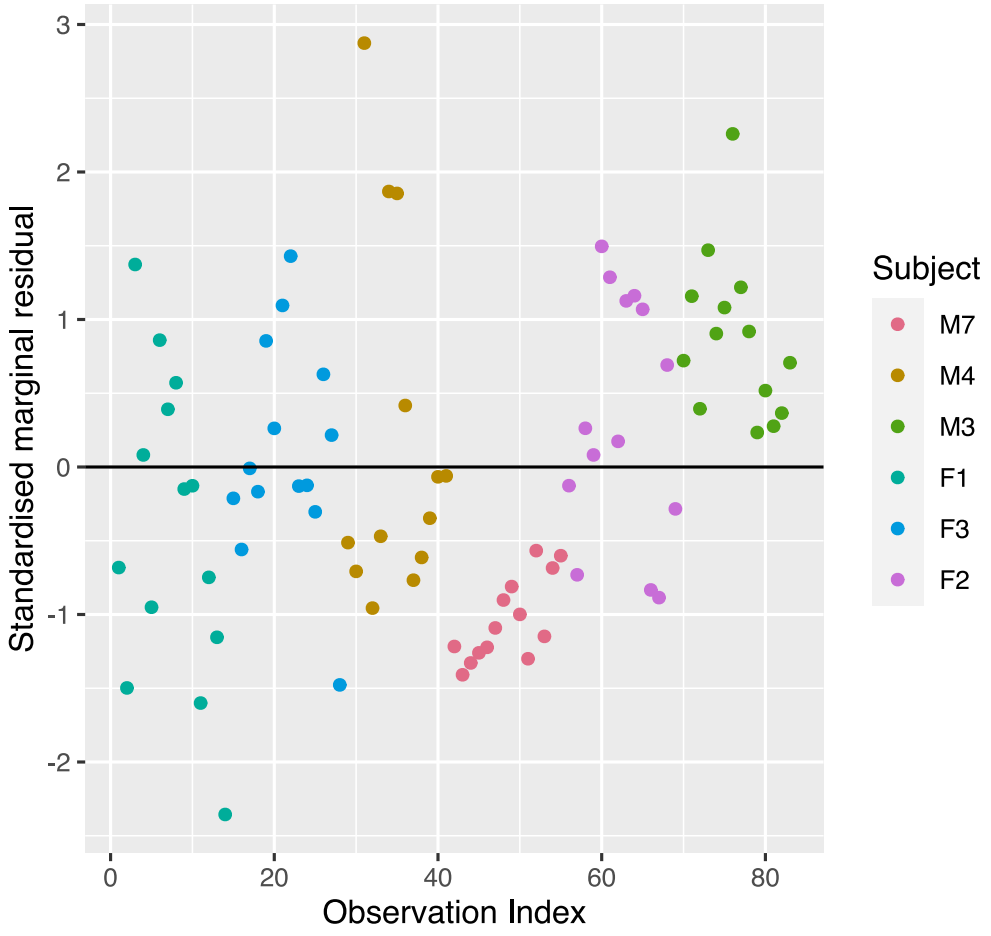
Types of residuals and corresponding residual diagnostic purpose:

- Marginal residuals, $\hat{\boldsymbol{\xi}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$
  - Linear of the effects fixed
  - Presence of outlying observations
  - Within-units covariance matrix
- Random effect residuals, $\mathbf{Z}\hat{\mathbf{b}}$
  - Presence of outlying subjects
  - Normality of the random effects
- Conditional residuals, $\hat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{Z}\hat{\boldsymbol{b}}$
  - Presence of outlying observations
  - Homoskedasticity of conditional errors
  - Normality of conditional errors

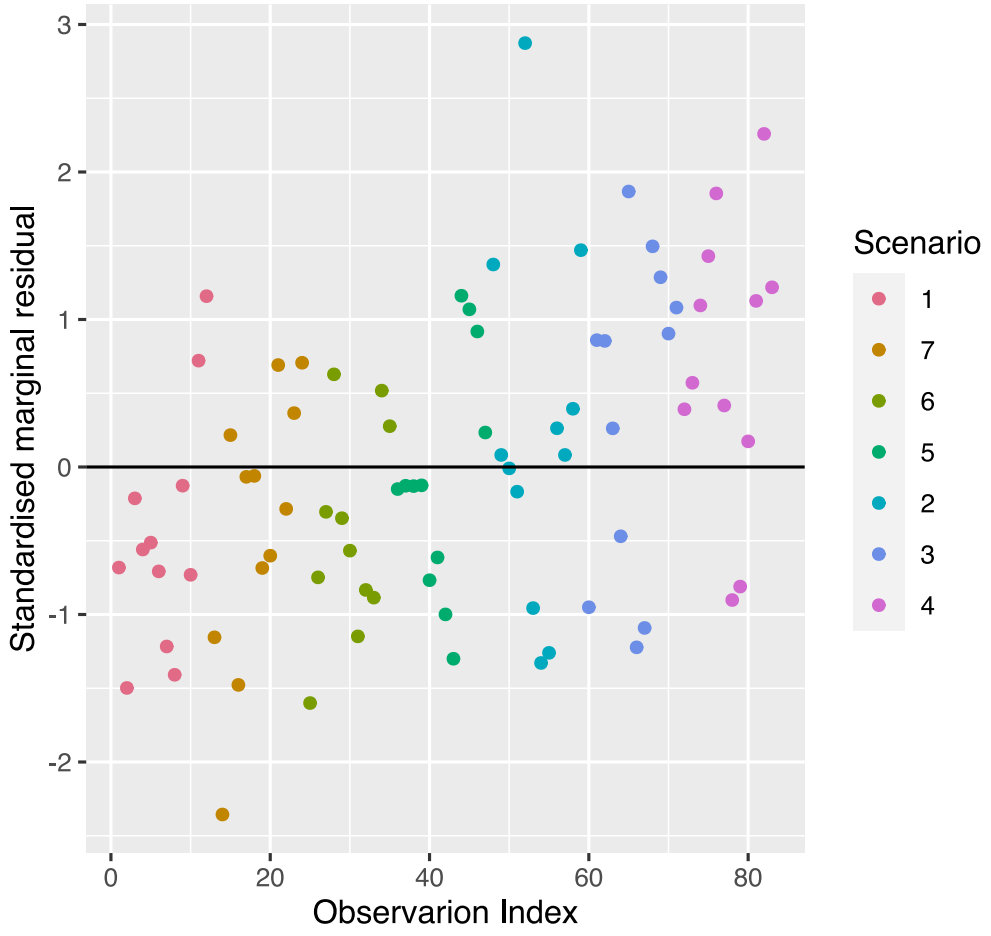
[1.]Haslett, J., & Haslett, S. J. (2007). The three basic types of residuals for a linear model. International Statistical Review, 75(1), 1-24. Chicago
[2.]Singer, J. M., Rocha, F. M., & Nobre, J. S. (2017). Graphical tools for detecting departures from linear mixed model assumptions and some remedial measures. International Statistical Review, 85(2), 290-324.
[3.]Loy, A., Hofmann, H., & Cook, D. (2017). Model choice and diagnostics for linear mixed-effects models using statistics on street corners. Journal of Computational and Graphical Statistics, 26(3), 478-492.

Presence of outlying observations

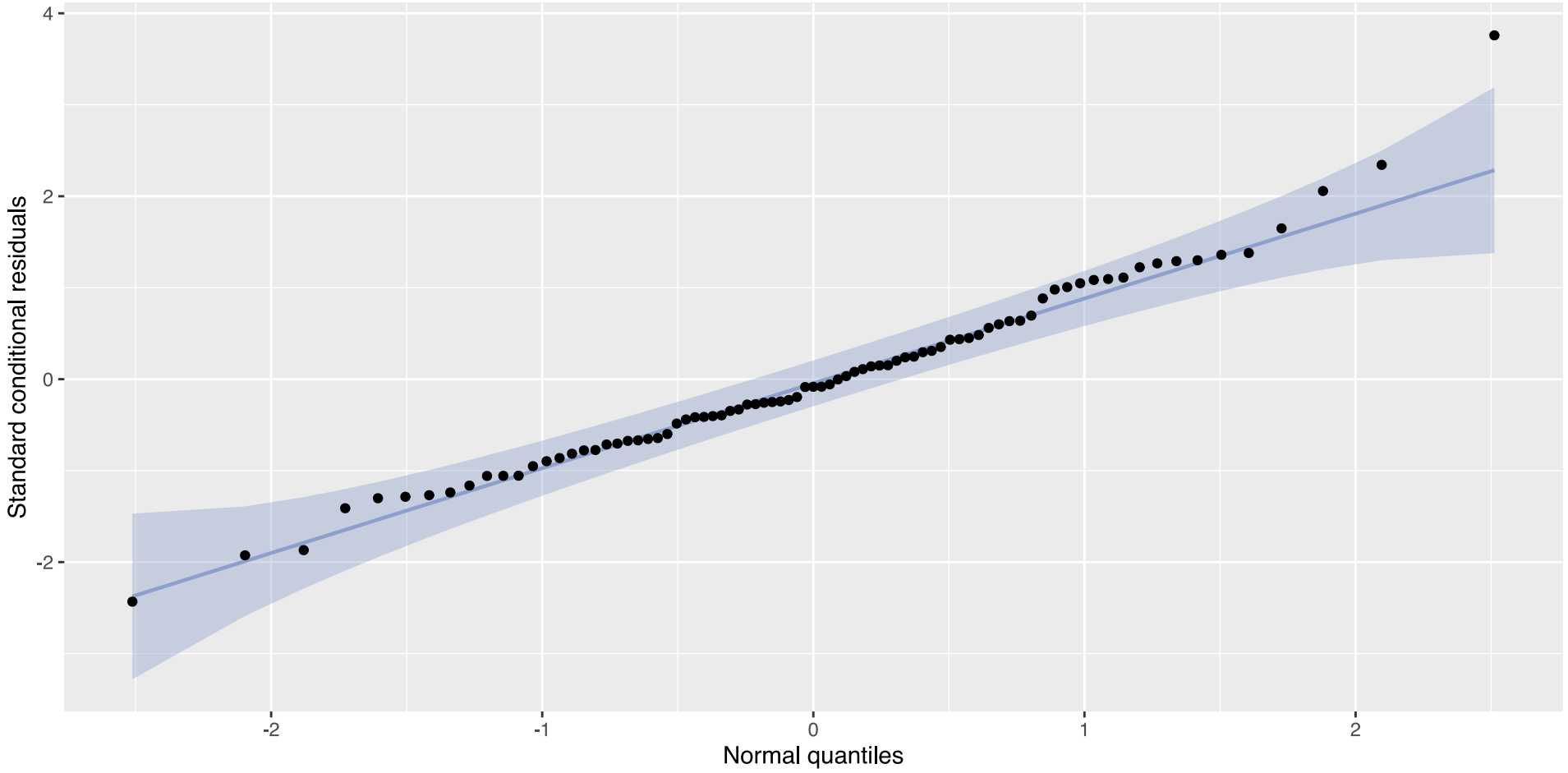
- Subject



- Scenario



Normality of conditional errors



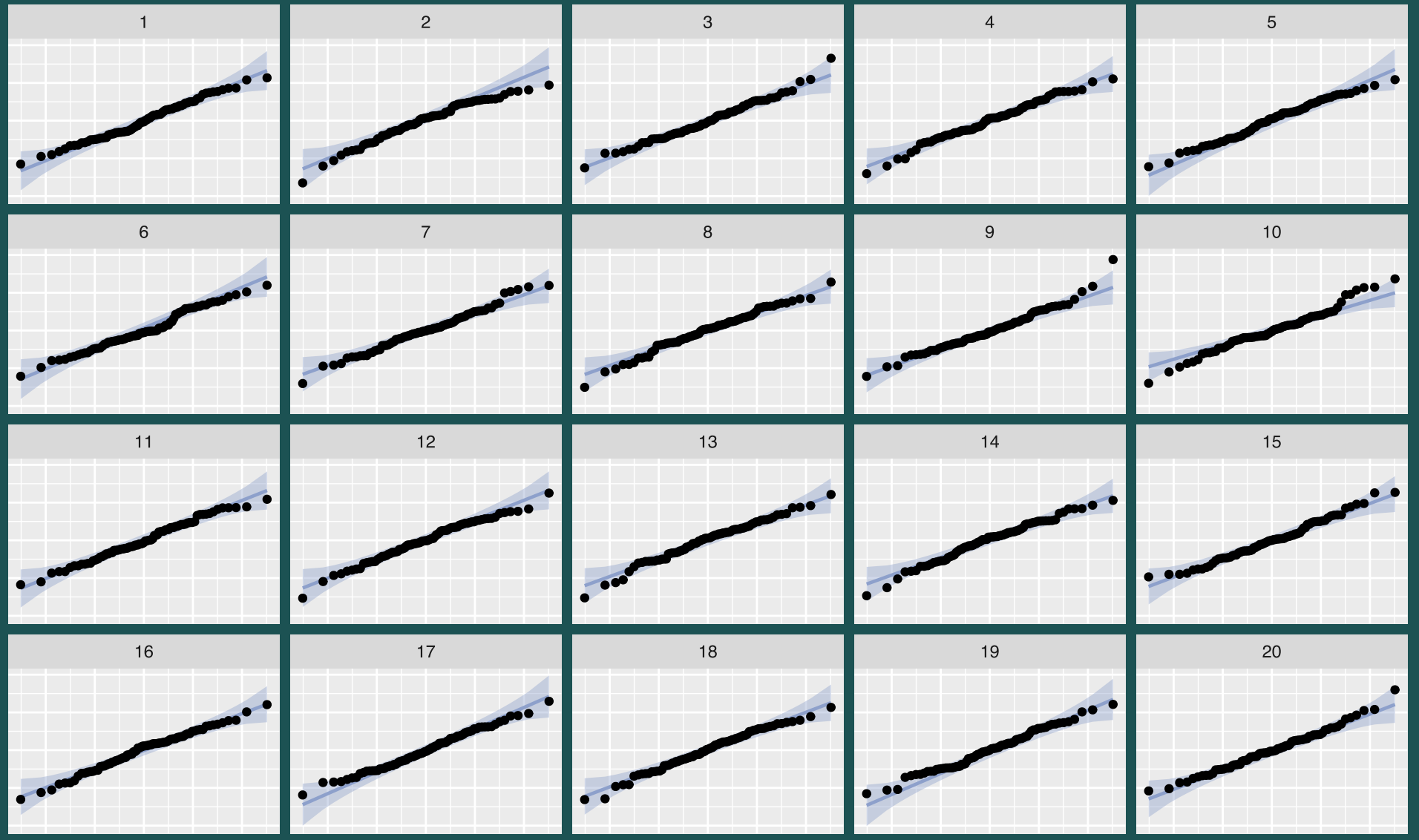
Lineup Protocol

1. Simulate the new responses
2. Refit the model to these simulated responses
3. Extract the residuals from the proposed model
4. Construct the lineup
5. **Which plot is most different?**

Visual Inference

- Hypothesis
- Test Statistic: Observed plot
- Sampling Distribution: Lineup
- Desicion Rule: If the observed plot is identifiable, then we can reject the null hypothesis.

[1.]Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E.-K., Swayne, D. F, Wickham, H. (2009) Statistical Inference for Exploratory Data Analysis and Model Diagnostics Royal Society Philosophical Transactions A, 367(1906):4361-4383. http://rsta.royalsocietypublishing.org/content/367/1906/43611.
[2.]Mahbubul Majumder , Heike Hofmann & Dianne Cook (2013) Validation of Visual Statistical Inference, Applied to Linear Models, Journal of the American Statistical Association, 108:503, 942-956, DOI: 10.1080/01621459.2013.808157


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Research Objective

1. Least confounded conditional residuals $\mathbf{c}_k^\top \hat{\mathbf{e}}$ v.s Standardised conditional residuals
2. Presence of outlying observation:
    - Marginal residual v.s Conditional residuals

Research Plan

Do a **user study**

[1.]Loy, A., Hofmann, H., & Cook, D. (2017). Model choice and diagnostics for linear mixed-effects models using statistics on street corners. Journal of Computational and Graphical Statistics, 26(3), 478-492.
[2.]Hilden-Minton, J.A. (1995). Multilevel diagnostics for mixed and hierarchical linear models, Unpublished PhD Thesis, University of California, Los Angeles.
[3.]Schützenmeister, A. & Piepho, H.P. (2012). Residual analysis of linear mixed models using a simulation approach. Comput. Stat. Data Anal., 56, 1405–1416.