Visual Inference for Graphcial Diagnostic of Linear Mixed Models

Supervisor: Dr Emi Tanaka & Prof Di Cook

Kaiwen Jin

Monash University

2020-11-13

Linear Mixed Model 👩‍💻

Linear Mixed Model(con't)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{e}$$

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

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

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

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.

Covarrubias-Pazaran G. 2016. Genome assistedprediction of quantitative traits using the R package sommer. PLoS ONE 11(6):1-15.

CONVENTIONAL INFERENCE

NULL HYPOTHESIS

↓

TEST STATISTICS:

$$T^{(i)}(y)(i \in I)$$

↓

JUDGEMENT BY NUMERICAL VALUES

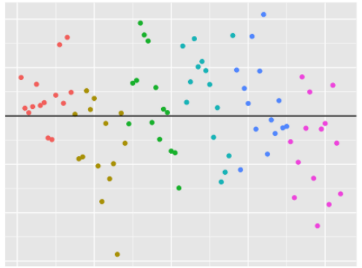
LINEUP PROTOCOL

NULL HYPOTHESIS

↓

TEST STATISTICS:

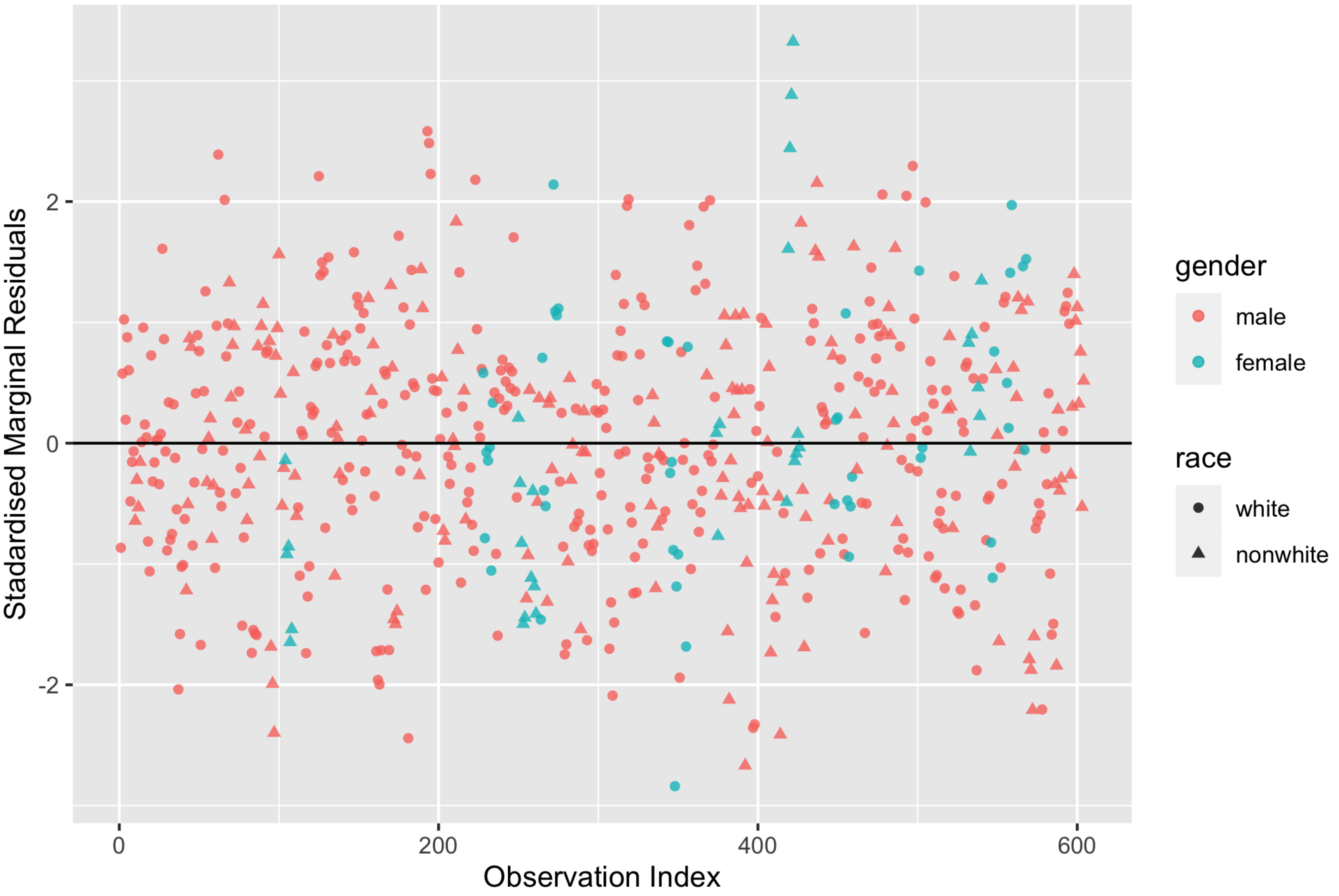
(DATA PLOT)



↓

JUDGEMENT BY HUMAN PRECEPTIONS

Buja, A., Cook, D., Hofmann, H., Law ... ee, E.-K., Swayne, D. F., and ... for Exploratory Data Analy ... Model Diagnostics," Royal Society



Stadardised Marginal Residuals
Observation Index
2
0
-2
0
200
400
600
gender
male
female
race
white
nonwhite

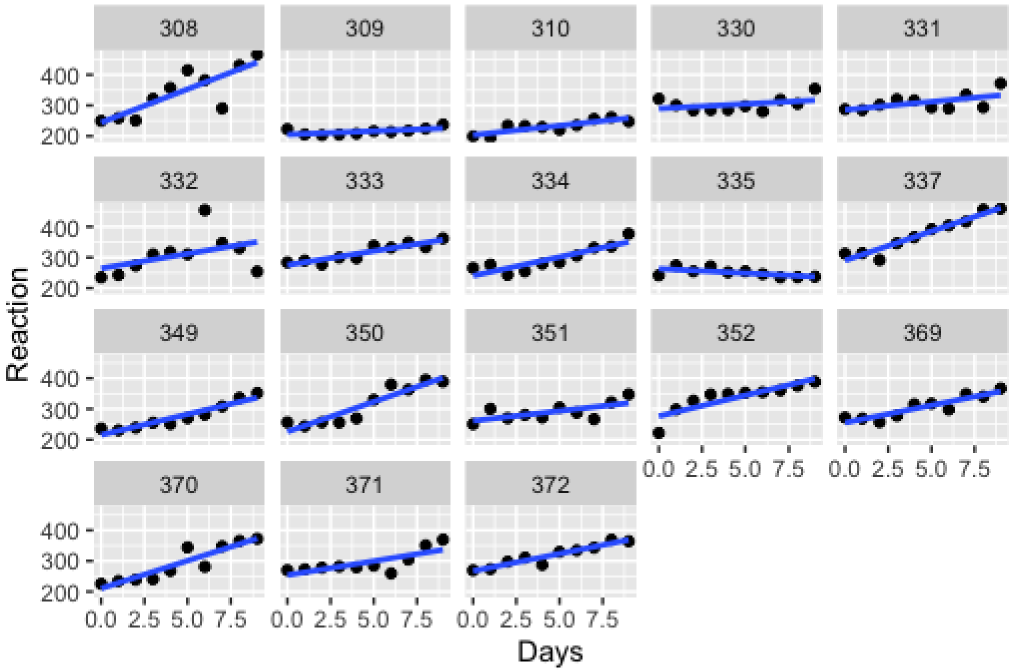
Data

- Linguistic study case
  - The data contains 84 observations on the voice pitch (or frequency) from 6 subjects (3 females and 3 males) under 7 scenarios with 2 attitudes (informal or polite).
- Sleep deprivation study case
  - Average reaction time per day for subjects restricted to 3 hours of sleep per night over 10 days for 180 observations.
- Autism study case
  - 155 children between the ages of 2 and 13 who were diagnosed with either autism spectrum disorder or non-spectrum developmental delays at age 2.
    - `childid`: Child ID
    - `sicdegp`: Sequenced Inventory of Communication Development group (an assessment of expressive language development) with levels low, mod, and high
    - `age2`: Age (in years) centred around age 2 (age at diagnosis)
    - `vsae`: Vineland Socialization Age Equivalent
    - `gender`: Child's gender, female and male
    - `race`: Child's race, white and non-white
    - `bestest2`: Diagnosis at age 2, autism and pdd

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.

Adam Loy, Heike Hofmann (2014). HLMdiag: A Suite of Diagnostics for Hierarchical Linear Models in R. Journal of Statistical Software, 56(5), 1-28. URL http://www.jstatsoft.org/v56/i05/.

Examlple: 🌑 Sleep deprivation study case



Exploratory Data Analysis(EDA):

- Random intercepts and random slopes for different subjects

☞ Using `lmer` function from `lme4` package to fit the linear mixed model

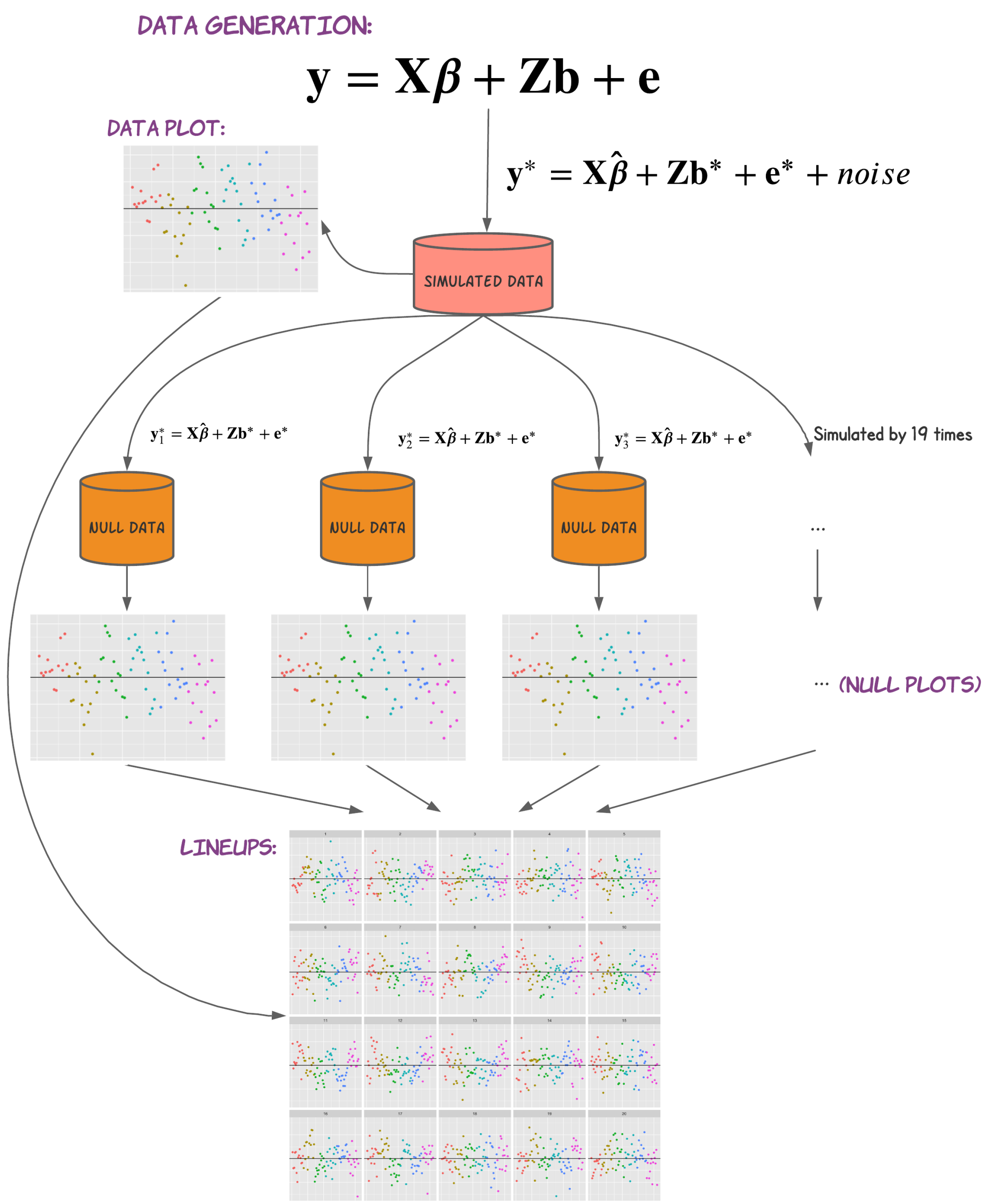
Tukey, J. W. (1977), Exploratory Data Analysis, Boston, MA: Addison-Wesley.

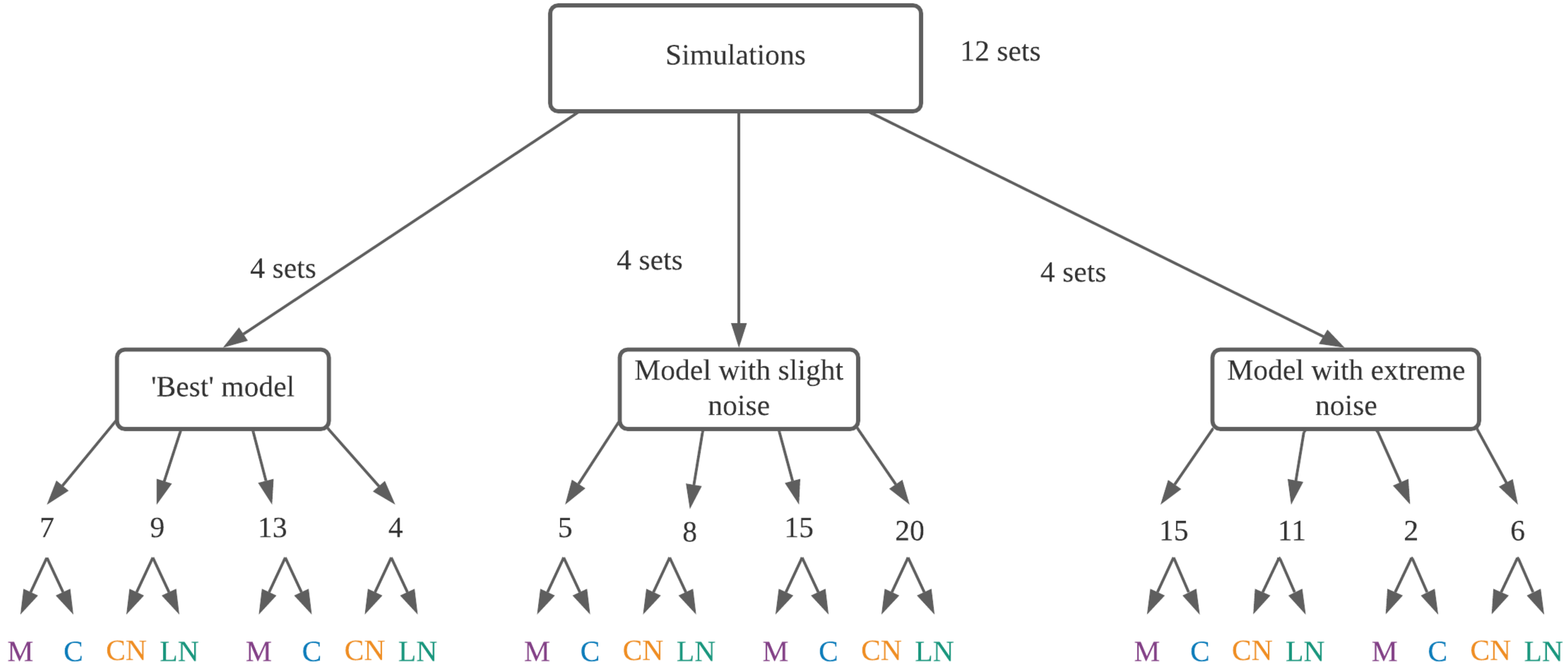
Methodology

Generate 3 versions for each data set (🔔 replicate each version for 4 times) for 4 plots (Residual plot for marginal residuals $\hat{\boldsymbol{\xi}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ and conditional residuals $\hat{\boldsymbol{e}} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{Z}\hat{\boldsymbol{b}}$, QQ plot for conditional residuals and least confounded residual $\mathbf{c}_k^\top \hat{\mathbf{e}}$):

- Simulated from "best" model
  $\mathbf{y}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\mathbf{b}^* + \mathbf{e}^*$
- Added some noise to "best" model
  $\mathbf{y}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\mathbf{b}^* + \mathbf{e}^* + \text{noise}$
  - Linguistic study case: with error term that follows a student t distribution
  - Sleep deprivation study case: added the within unit correlation in the error term by value 25, that is for first 10th observations' error terms are correlated.
  - Autism study case: with error term that follows a student t distribution
- Added extreme noise to "best" model
  $\mathbf{y}^* = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\mathbf{b}^* + \mathbf{e}^* + \text{extreme noise}$
  - Linguistic study case: randomly added 20% of average value of each categories with gender and attitude (F&Polite, F&Informal, M&Polite and M&Informal) to only one response variable of each categories.
  - Sleep deprivation study case: randomly added 20% of mean value to 2 response values of each subject.
  - Autism study case: randomly added 20% of the total median values to 20% of observations

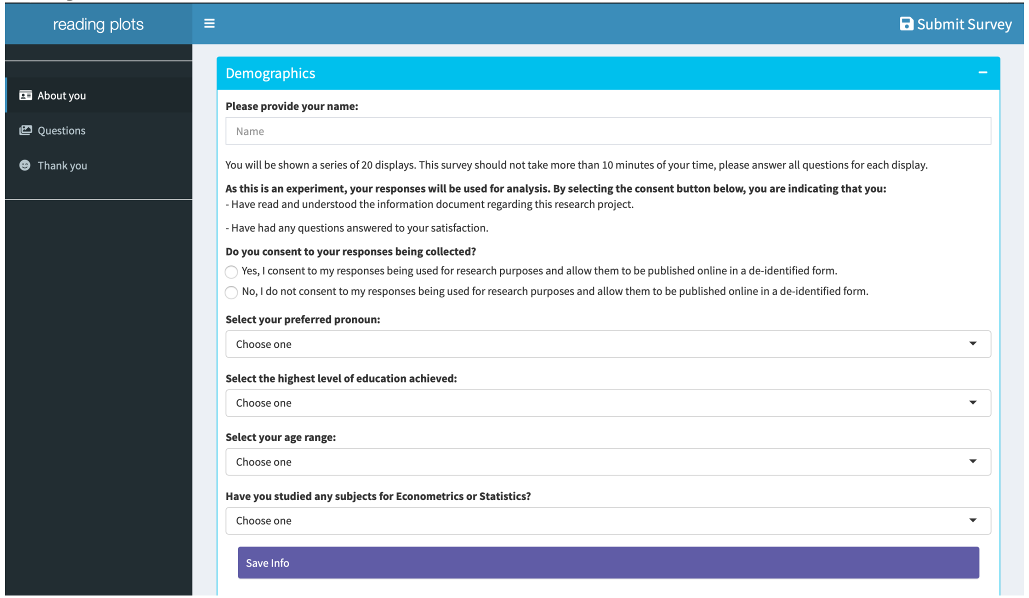
where $\mathbf{b}^* \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{G}})$ and $\mathbf{e}^* \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{R}})$


DATA GENERATION:
$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \mathbf{e}$
$\mathbf{y}^* = \mathbf{X}\hat{\beta} + \mathbf{Z}\mathbf{b}^* + \mathbf{e}^* + noise$
DATA PLOT:
SIMULATED DATA
$\mathbf{y}_1^* = \mathbf{X}\hat{\beta} + \mathbf{Z}\mathbf{b}^* + \mathbf{e}^*$
$\mathbf{y}_2^* = \mathbf{X}\hat{\beta} + \mathbf{Z}\mathbf{b}^* + \mathbf{e}^*$
$\mathbf{y}_3^* = \mathbf{X}\hat{\beta} + \mathbf{Z}\mathbf{b}^* + \mathbf{e}^*$
Simulated by 19 times
NULL DATA
NULL DATA
NULL DATA
...
... (NULL PLOTS)
LINEUPS:

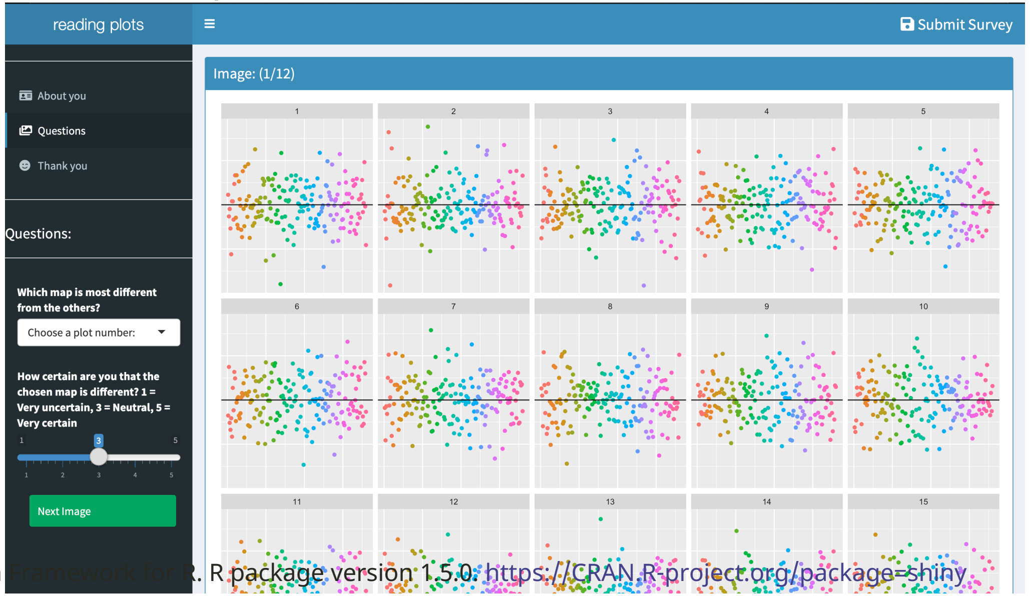

Simulations
12 sets
4 sets
4 sets
4 sets
'Best' model
Model with slight noise
Model with extreme noise
7
9
13
4
5
8
15
20
15
11
2
6
M C CN LN M C CN LN
M C CN LN M C CN LN
M C CN LN M C CN LN

Survey 📸

We are doing a survey on the shiny app by asking observers' name, gender, age, education level and if they've learnt econometrics or statistics before.

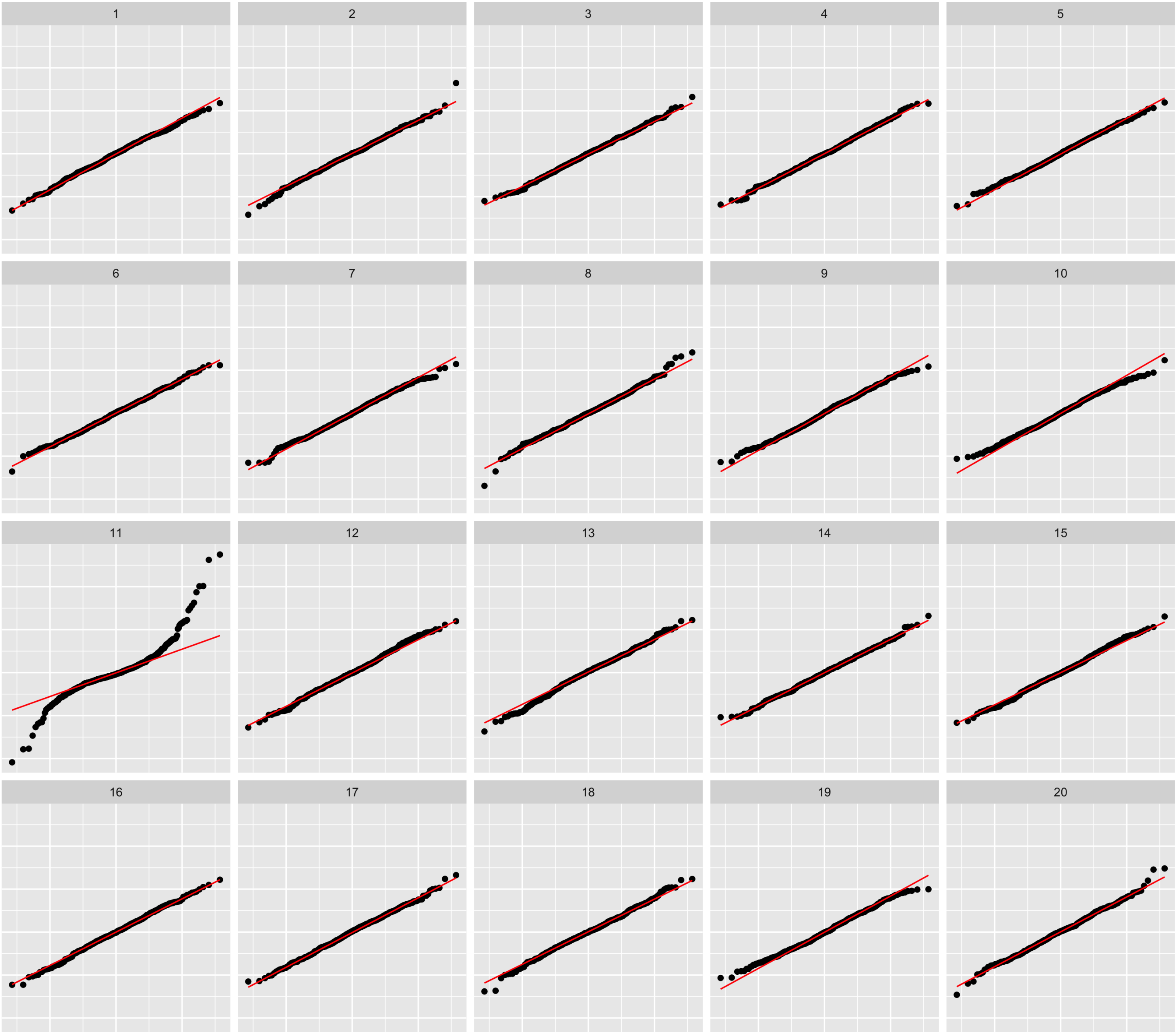


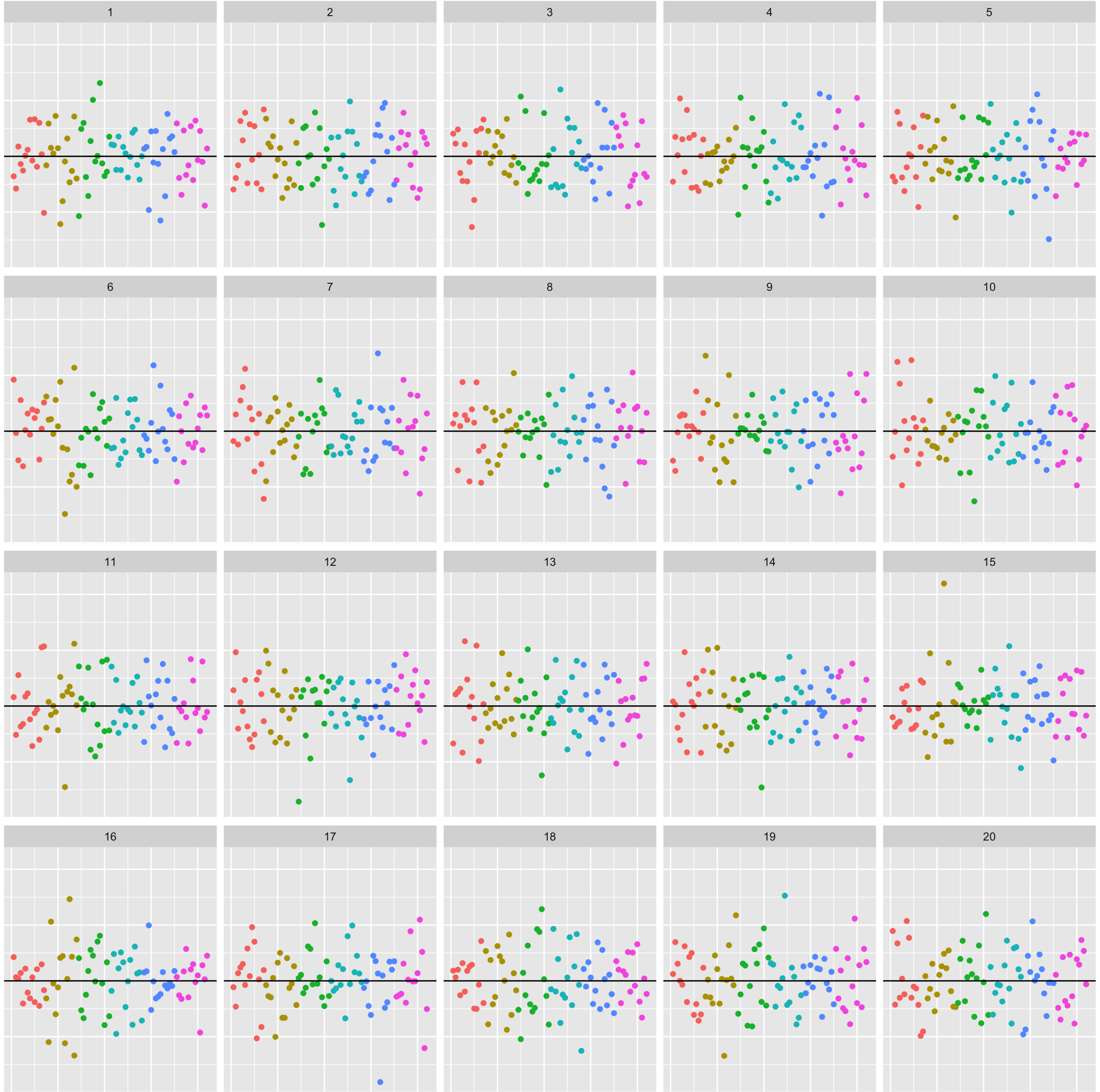
By using the Latin Squared Design, we randomly give each observer 12 different lineups including all 3 data sets with 4 different plot types. The observers are asked the questions "Which map is most different from the others?" and "How certain are you that the chosen map is different?".



Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie and Jonathan McPherson (2020). shiny: Web Application Framework for R. R package version 1.5.0. https://CRAN.R-project.org/package=shiny

Felipe de Mendiburu (2020). agricolae: Statistical Procedures for Agricultural Research. R package version 1.3-3. https://CRAN.R-project.org/package=agricolae


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


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

Results

For K independet observers, let X be the number of observers picking the test statistic from the lineup. Under the null hypothesis $X \sim \mathrm{Binom}_{K,1/m}$.

The $p - value$ of a lineup of size m evaluated by K observers is given as

$$
\begin{aligned}
P(X \geq x) &= 1 - \mathrm{Binom}_{K,1/m}(x-1) \\
&= \sum_{i=x}^{K} \binom{K}{i} \left(\frac{1}{m}\right)^{i} \left(\frac{m-1}{m}\right)^{K-1}
\end{aligned}
$$

where x is the number of observers selecting the actual data plot.

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

Results

Overview of all lineup evaluations

lineups	version1	version2	version3
Marginal residual plot	0/24	1/12	1/15
Condtional residual plot	0/16	1/16	10/22 ***
QQ plot for conditional residual	0/19	0/18	12/16 ***
QQ plot for least confounded residual	1/17	1/13	2/21 .

Note:

Signif. codes: $0 \leq *** \leq 0.001 \leq ** \leq 0.01 \leq * \leq 0.05 \leq . \leq 0.1 \leq'' \leq 1$`

Based on Version 3,

lineups	Autism	Linguistic	Sleep
Condtional residual plot	7/7 ***	3/6 **	0/9
QQ plot for conditional residual	8/10 ***	0/2	4/5 ***

Limitations

1. Small sample
2. Technical problem
3. Although we remove the information about the plots such as axis labels, titles, legends in the lineup to avoid subjective bias. Observers may lose the information about the data.

Conclusion

Instead of simply rejecting the null hypothesis, the use of lineups also provides us the reason that we are rejecting the null hypothesis. However, this process is relative new and relies on the simulation approach, the design of map created, and observers.

Thank you

Slides created via the R package **xaringan**.