

Final remarks

Workshop: Analysis of Longitudinal Data

12th Nov 2024

Jaroslav Harezlak

Armando Teixeira-Pinto



Scientific questions

- Horses (science) vs. carts (statistics)
- Interpretation of the results (software output)
- Clinical trials vs. observational data vs. "data" from the web

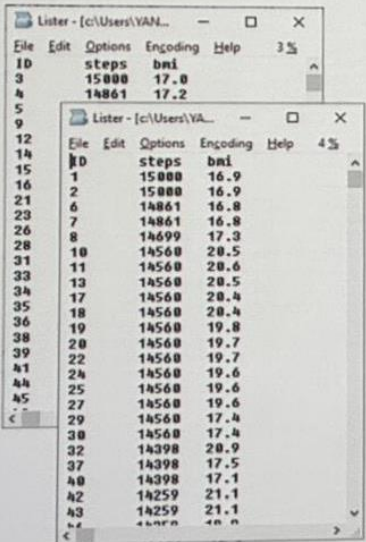
Statistical analysis

- Data exploration: plots, summaries, EDA
- More complex is not necessarily better, for example, if you have pre-post design or if just have a few clusters (two or three), you might consider fixed effects only
- ML/AI vs. statistical methods
- More data is not necessarily better (noise, heterogeneity, etc.)

Fun example

<https://www.biorxiv.org/content/10.1101/2020.07.30.228916v1.full.pdf>

Two files(F/M):
Steps, BMI

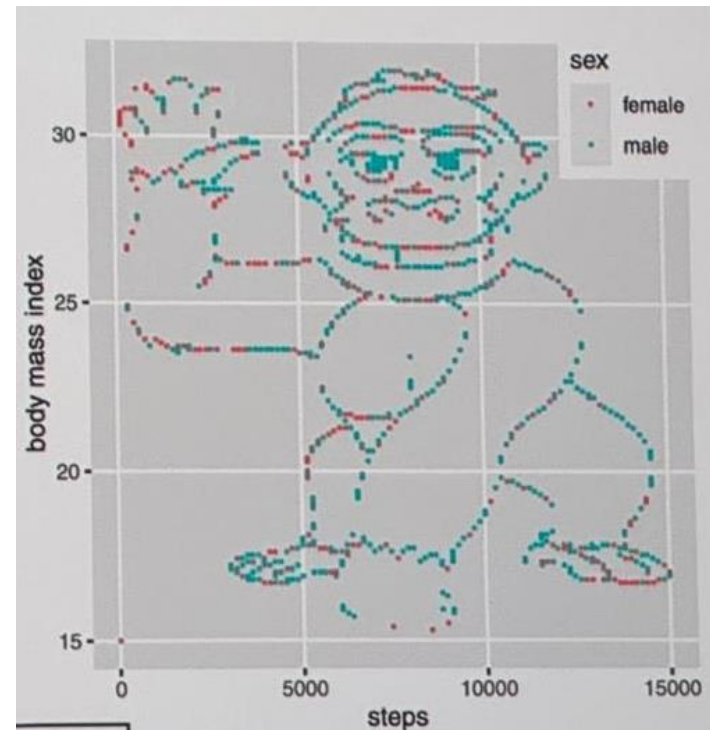
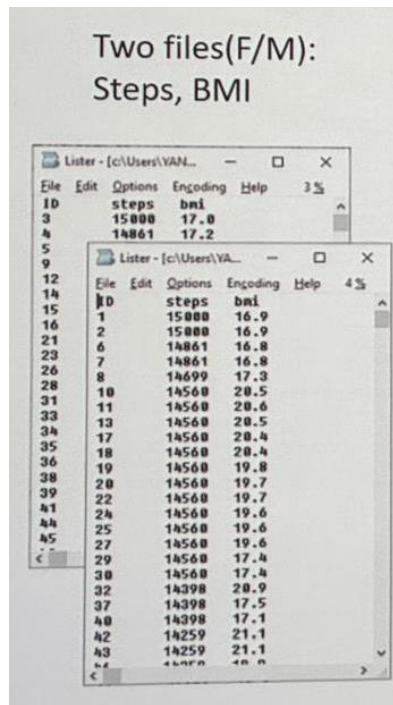


The image shows two overlapping Notepad++ windows. The top window, titled 'Listner - [c:\Users\YAN...', displays a table with two columns: 'steps' and 'bmi'. The bottom window, also titled 'Listner - [c:\Users\YAN...', displays a larger table with three columns: 'ID', 'steps', and 'bmi'. Both windows have a menu bar with 'File', 'Edit', 'Options', 'Encoding', and 'Help'. The top window's status bar shows '3%' and the bottom window's status bar shows '4%'.

| ID | steps | bmi |
|----|-------|-------|
| 10 | 15000 | 17.0 |
| 3 | 15000 | 17.0 |
| 4 | 14861 | 17.2 |
| 5 | | |
| 9 | | |
| 12 | | |
| 14 | | |
| 15 | 1 | 15000 |
| 16 | 2 | 15000 |
| 21 | 6 | 14861 |
| 23 | 7 | 14861 |
| 26 | 8 | 14699 |
| 28 | 10 | 14560 |
| 31 | 11 | 14560 |
| 33 | 13 | 14560 |
| 34 | 17 | 14560 |
| 35 | 18 | 14560 |
| 36 | 19 | 14560 |
| 38 | 20 | 14560 |
| 39 | 22 | 14560 |
| 41 | 24 | 14560 |
| 44 | 25 | 14560 |
| 45 | 27 | 14560 |
| | 29 | 14560 |
| | 30 | 14560 |
| | 32 | 14398 |
| | 37 | 14398 |
| | 40 | 14398 |
| | 42 | 14259 |
| | 43 | 14259 |
| | 44 | 14259 |

Fun example

<https://www.biorxiv.org/content/10.1101/2020.07.30.228916v1.full.pdf>



Design and data collection

- Confounding, bias
- Data harmonization
- Data missingness
- Causal inference?

Topics not covered

- **Diagnostics** Very similar to GLM
- **Other distributions for random effects** Normal distribution works well
- **GEE for MAR** Not commonly used
- **GLS and ANOVA for repeated measures** Classical methods
- **Joint survival/longitudinal models** Advanced methods
- **Computational methods** Advanced methods