Utilizing Informative Priors in Group Comparisons with Completely Censored Data

Luwis Diya

23/10/2025

Johnson&Johnson
Innovative Medicine

Introduction

Censored data: Common in the pharmaceutical industry

- e.g. differential expression of cytokines in the discovery space

Approach: Use Tobit or censored regression models

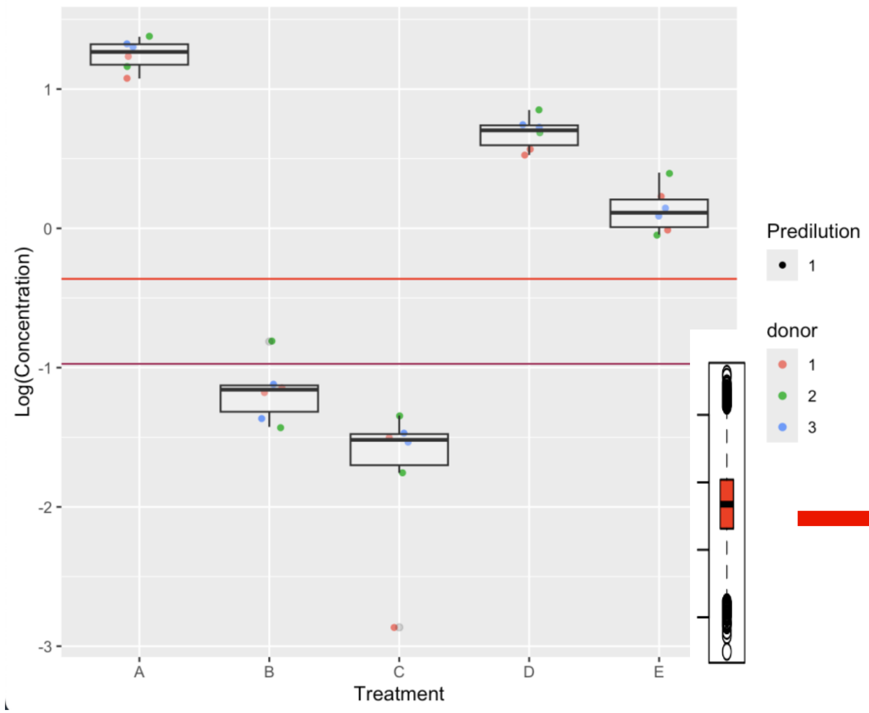
- **Assumption**:
  - The **Wald method** assumes that all group means are estimable.
  - **Problem**: Any complete censored group makes estimates unreliable.

Potential solutions:

- **Profile Likelihood approach**
- **Bayesian Approach** with *informative priors*

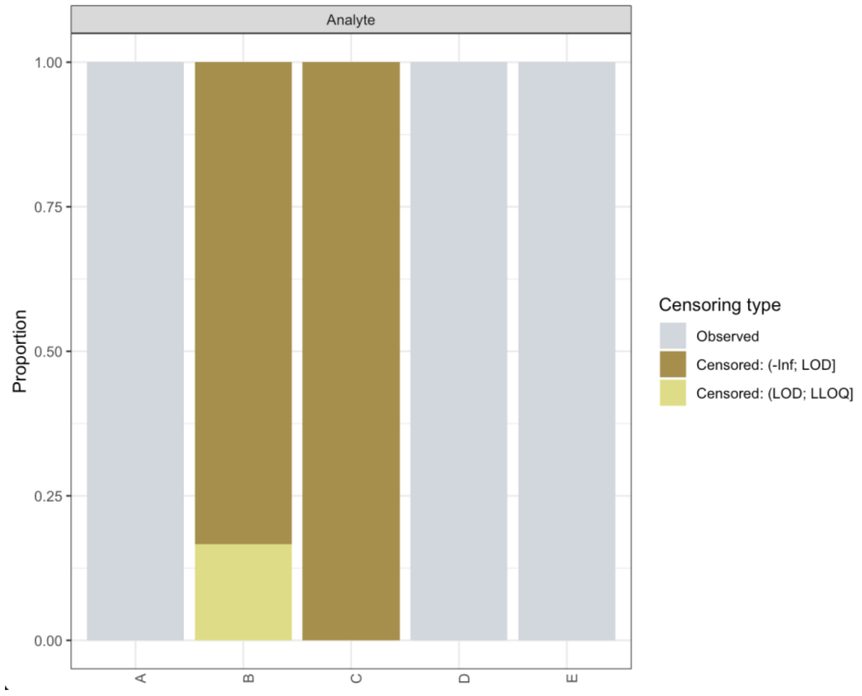
Data

- **Response**: Conc of Cytokine X
- **Treatments**: A, B, C, D and E
- **Detection Limits:**
  - **LOD** (*Limit of Detection*) : 0.378
  - **LLOQ** (*Lower Limit of Quantification*) : 0.6958



Blanks: N(-1.66, 0.27^2)

- Censoring Challenges:
  - **Types**: Left- and interval-censored values
  - **Critical issue**: Entirely censored treatment groups(2 groups)



Bayesian analysis results

Non informative priors:

- An issue with Treatment C!

Treatment	Estimate	L95%CI	U95%CI
A	1.26	0.81	1.81
B	-1.13	-1.62	-0.58
C	-6.32×10^{21}	-1.91×10^{22}	-1.95×10^{20}
D	0.70	0.23	1.26
E	0.14	-0.31	0.70

Bayesian analysis results

Non informative priors:

- An issue with Treatment C!
- But why not Treatment B?
  - Mean=(log(LOD)+log(LLOQ))/2

Treatment	Estimate	L95%CI	U95%CI
A	1.26	0.81	1.81
B	-1.13	-1.62	-0.58
C	-6.32×10^{21}	-1.91×10^{22}	-1.95×10^{20}
D	0.70	0.23	1.26
E	0.14	-0.31	0.70

Bayesian analysis results:

Informative prior (Treatment C)

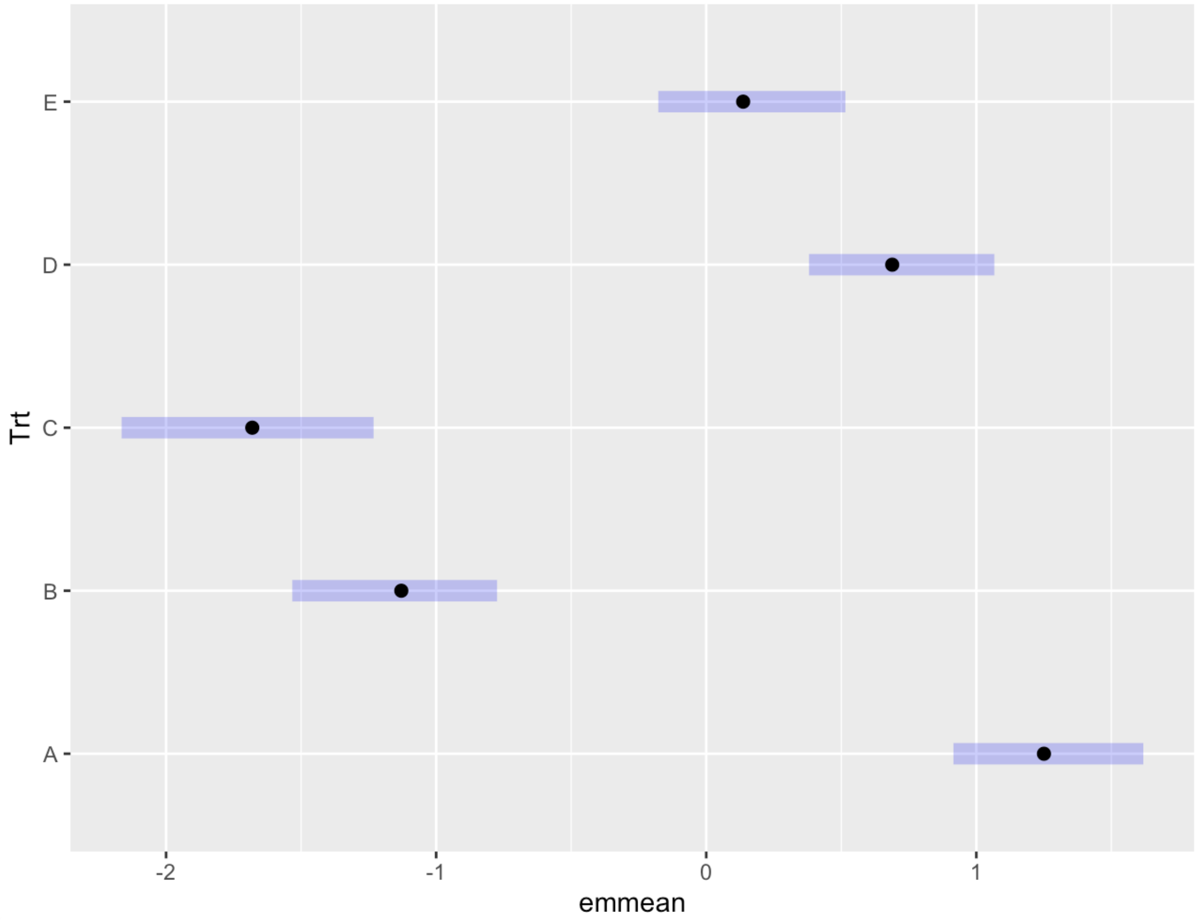
- Based on blank samples/control wells
  - $N(-1.66, 0.27^2)$
- Valid results

Treatment	Estimate	L95%CI	U95%CI
A	1.27	0.95	1.67
B	-1.12	-1.49	-0.71
C	-1.69	-2.18	-1.24
D	0.70	0.40	1.11
E	0.15	-0.16	0.55

Bayesian analysis results

Treatment Group Comparison:

- **Treatment A**: Highest Cytokine X concentration
- **Note**: No difference between **Treatment B** and **Treatment C**



Key Takeaways

What We've Learned

- **Complete censored group**:
  - with only left-censored values lead to unreliable inference
  - but with some interval-censored values lead to valid inference
- **Bayesian approach**:
  - Allows for the incorporation of **informative priors** resulting in valid and robust analyses.

References

1. Tobin, J. (1958). "Estimation of Relationships for Limited Dependent Variables." Econometrica, 26(1), 24-36.
2. Helsel, D. R. (2012). Statistics for censored environmental data using Minitab and R (2nd ed.). John Wiley & Sons.
3. Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. Journal of Statistical Software, 80(1), 1-28. https://doi.org/10.18637/jss.v080.i01

Likelihood of censored data

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \left[\phi\left(\frac{Y_i - x_i^T\beta}{\sigma}\right)^{\delta_{i,0}} \left(1 - \Phi\left(\frac{R_i - x_i^T\beta}{\sigma}\right)\right)^{\delta_{i,R}} \left(\Phi\left(\frac{L_i - x_i^T\beta}{\sigma}\right)\right)^{\delta_{i,L}} \left(\Phi\left(\frac{R_i - x_i^T\beta}{\sigma}\right) - \Phi\left(\frac{L_i - x_i^T\beta}{\sigma}\right)\right)^{\delta_{i,I}} \right]$$

- β is the vector of regression coefficients.
- σ^2 is the variance of the normal distribution.
- x_i is the vector of covariates for observation i.
- $\mu_i = x_i^T\beta$ is the mean of the distribution for observation i.
- $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF of the standard normal distribution, respectively.
- $\delta_{i,0}, \delta_{i,R}, \delta_{i,L}, \delta_{i,I}$ are indicator variables that are 1 if the observation i is exactly observed, right-censored, left-censored, or interval-censored, respectively, and 0 otherwise.