Cumulative Link Models

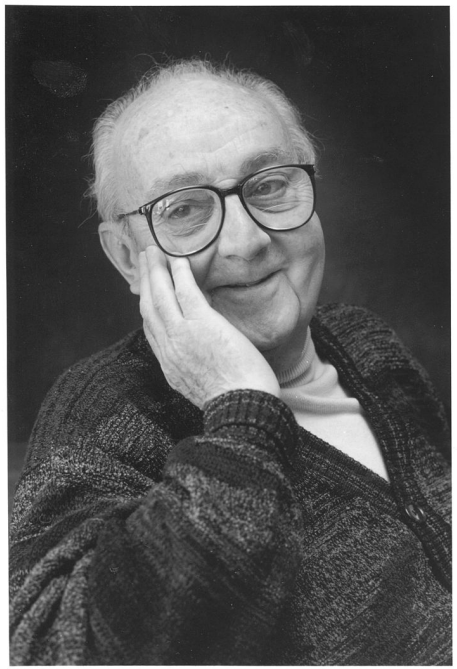
Josh Lospinoso, PhD

josh@lospi.net

October 17, 2015

Overview

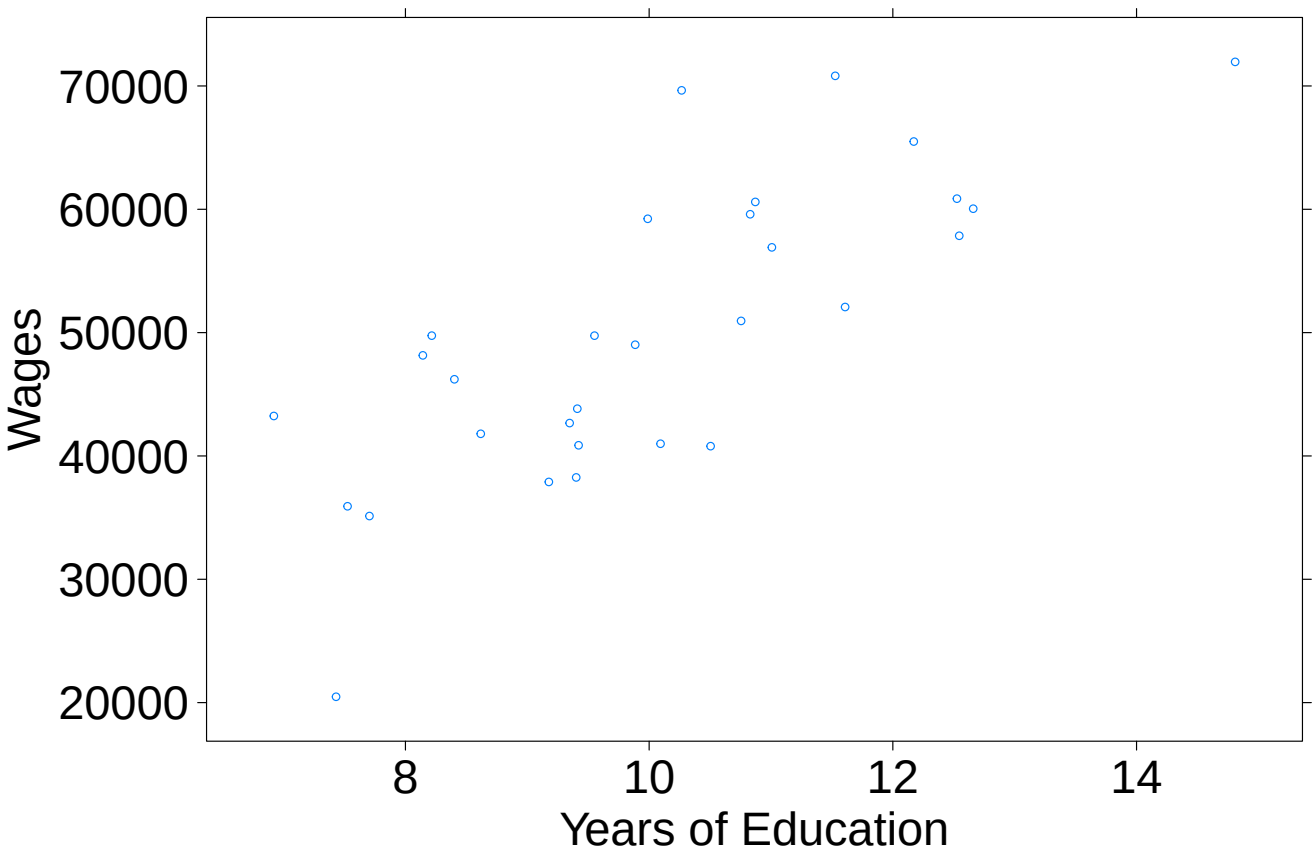
Let's start with some philosophy (on statistical modeling...)



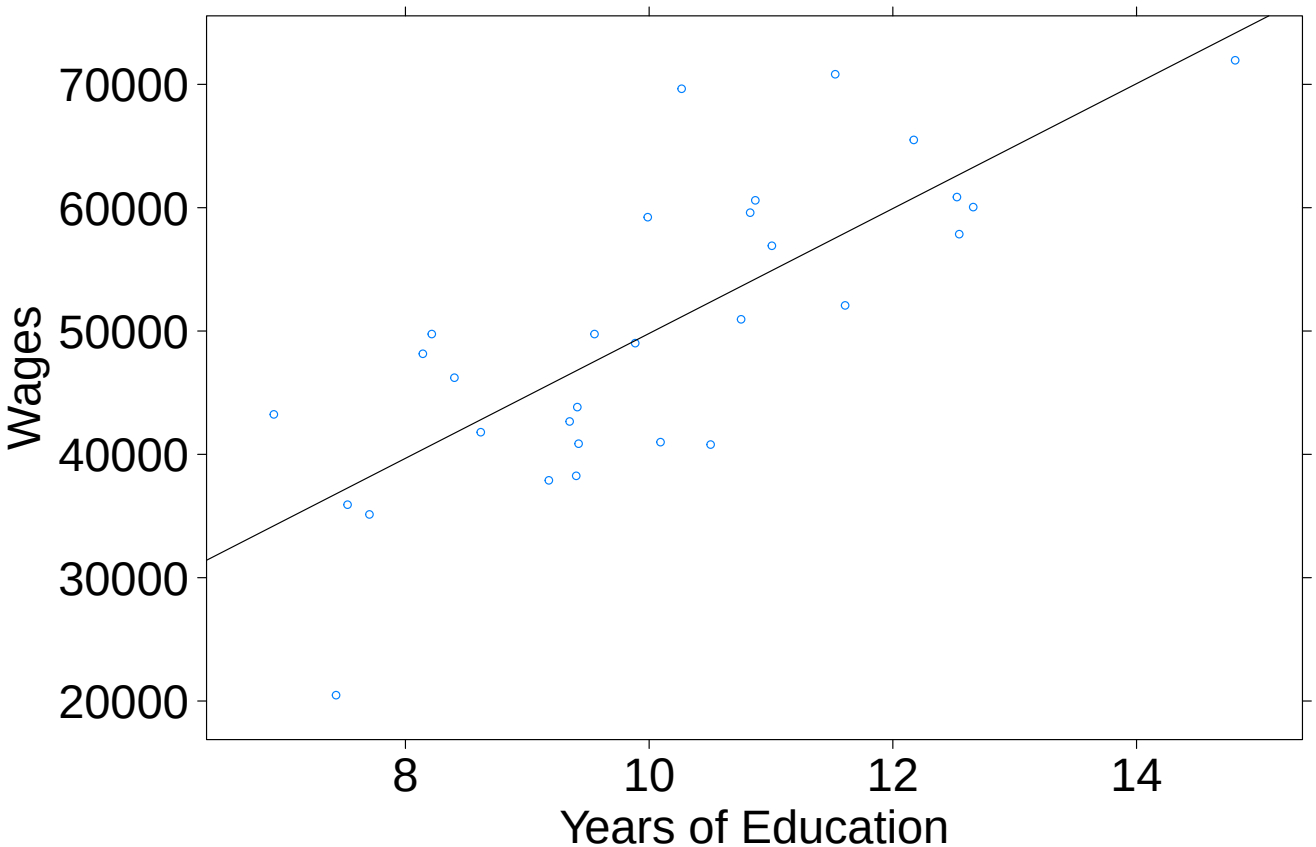
"All models are wrong but some are useful."

George E.P. Box (1978)

Linear Regression



Linear Regression



Slope $\approx$ 5000 dollars per year (and, importantly, we can get p-values...)

Such models are certainly not how reality works.

But they can *illuminate* some phenomena we care about.

Researchers in many disciplines have been taking this approach for decades. It works (and, importantly, scholarly journals tend to recognize this kind of analysis!)

Parametric and Non-parametric Statistics

Consider these hypotheses (taken from Stuart et. al. 1999):

1. that a normal distribution has a specified mean and variance
2. it has a given mean but unspecified variance
3. a distribution is of normal form with both mean and variance unspecified
4. that two unspecified continuous distributions are identical

1/2 are *parametric* hypotheses, 3/4 are *non-parametric*.

Continuous Data

- Speed
- Weight
- Distance
- Time

Continuous data *not countable*.

Discrete Data

- Die roll
- Number of patients
- MCAT score

Discrete data is *countable*.

Categorical Data

- Gender
- Race
- Hair color

Categorical data is about *group membership*.

Ordinal Categorical Data

- Stages of illness (e.g. stage 3 cancer)
- Questionnaires (e.g. "In general, how is your health?")
- Modified Fitzpatrick scale (i.e. complexion)

Categorical data is about *ordered group membership*.

Ordinal Categorical Data

In this talk, we focus on a kind of data that is common when studying pathologies.

Ordinal, **categorical** data exists on an **arbitrary numerical scale**. The exact numerical quantity **has no significance beyond its ability to establish ranking**.

I hope to convince you that this kind of data needs **special treatment**! (Linear models and ANOVAs simply won't do.)

Cumulative Link Model

CLMs are well suited to ordinal, categorical data. They:

1. provide a flexible regression framework to answer nuanced questions,
2. allow you to wring every ounce of power out of your data, and
3. can protect you from some kinds of inferential errors.

If you have this kind of data, CLMs are your go-to tool!

A Running Example

Volume 32, Issue 1 (March 2015)

< Previous Next >

Current Issue
Available Issues

Journal Information

Print ISSN: 0748-8068
Frequency: Quarterly

◂ Previous Article Volume 32, Issue 1 (March 2015) Next Article ▸

Add to Favorites Share Article Export Citations Track Citations (RSS | Email) Permissions

Full-text PDF

Article Citation:
Robert H. Sorensen, Matthew P. Connor, Joshua A. Lospinoso, Danielle J. Lospinoso, and Jose E. Barrera (*2015*) Combination Laser Resurfacing With Facial Plastic Surgery Is Superior to Lasers Alone. American Journal of Cosmetic Surgery: March 2015, Vol. 32, No. 1, pp. 12-17.

ORIGINAL ARTICLES

Combination Laser Resurfacing With Facial Plastic Surgery Is Superior to Lasers Alone

Robert H. Sorensen, MD, Captain, USAF, MC; Matthew P. Connor, MD, Captain, USAF, MC; Joshua A. Lospinoso, PhD, Captain, USA; Danielle J. Lospinoso, DO, Captain, USA, MC; Jose E. Barrera, MD, FACS, Lieutenant Colonel, USAF

From the Department of General Medical Education (Dr Sorensen), Department of Otolaryngology (Dr Connor, Dr Barrera), and Department of Dermatology (Dr D. J. Lospinoso), San Antonio Military Medical Center, Fort Sam Houston, Tex, and Portia Statistical Consulting LLC, San Antonio, Tex (Dr J. A. Lospinoso).

The authors listed have not published or submitted any related articles from this same study. All authors listed contributed such that they meet sufficient criteria to be listed as an author. None of the authors have financial or personal relationships that could inappropriately influence (or bias) the authors' decisions, work, or manuscript.

Corresponding author: Robert H. Sorensen, 9554 Dunstable, San Antonio, TX 78239 (e-mail: roberthsorensen@gmail.com).

Hypothesis: When treating rhytidosis, combining facial plastic surgery with CO_2 fractional laser treatment is superior to laser treatment alone.

Two pathologists evaluate Modified Fitzpatrick and Glogau before and after treatment

For each observation in Sorensen (2015), we have:

- Glogau after procedure(s)
- Modified Fitzpatrick after procedure(s)
- Laser?
- Pathologist ID
- Patient ID
- Estimated Age
- Modified Fitzpatrick before procedure(s)
- Glogau before procedure(s)

Textual Presentation

There is some non-zero probability that a pathologist rates a patient each of the possible Glogau/Modified Fitzpatrick levels.

Each explanatory term (e.g. laser, estimated age, and Glogau before procedures) can have a profound effect on these probabilities.

We want to infer about these effects!

Technical Presentation

Ordinal response variables Y_i falls in category j with probability π_{ij}.

Cumulative probabilities are therefore:

$$\gamma_{ij} = P(Y_i \leq j) = \pi_{i1} + ... + \pi ij \tag{1}$$

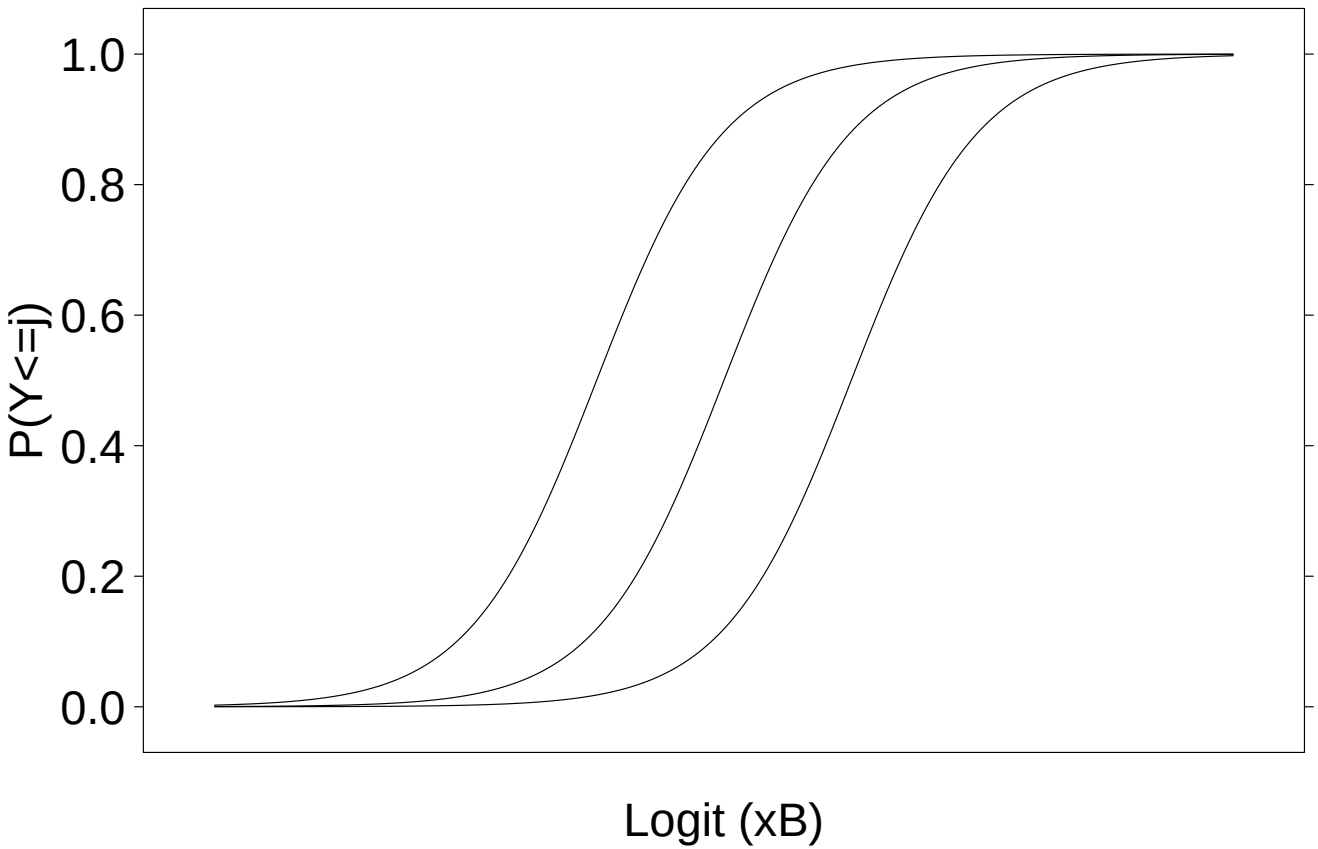
The CLM is a regression of *cumulative logits*

$$\text{logit}(\gamma_{ij}) = \theta_j + x_i^T \beta \tag{2}$$

where

- x are the explanatory terms
- β is the *effect* parameter
- θ_j is a constant for category j

Graphical Presentation



What Could Go Wrong?

Consider a fictitious dataset generated by a CLM:

- 100 observations
- Response Y on a scale from 1 to 5
- Two explanatory terms: X_1 and X_2
- Y depends on X_2, but not on X_1

Using this setup, we generate hundreds of thousands of sample datasets, e.g.:

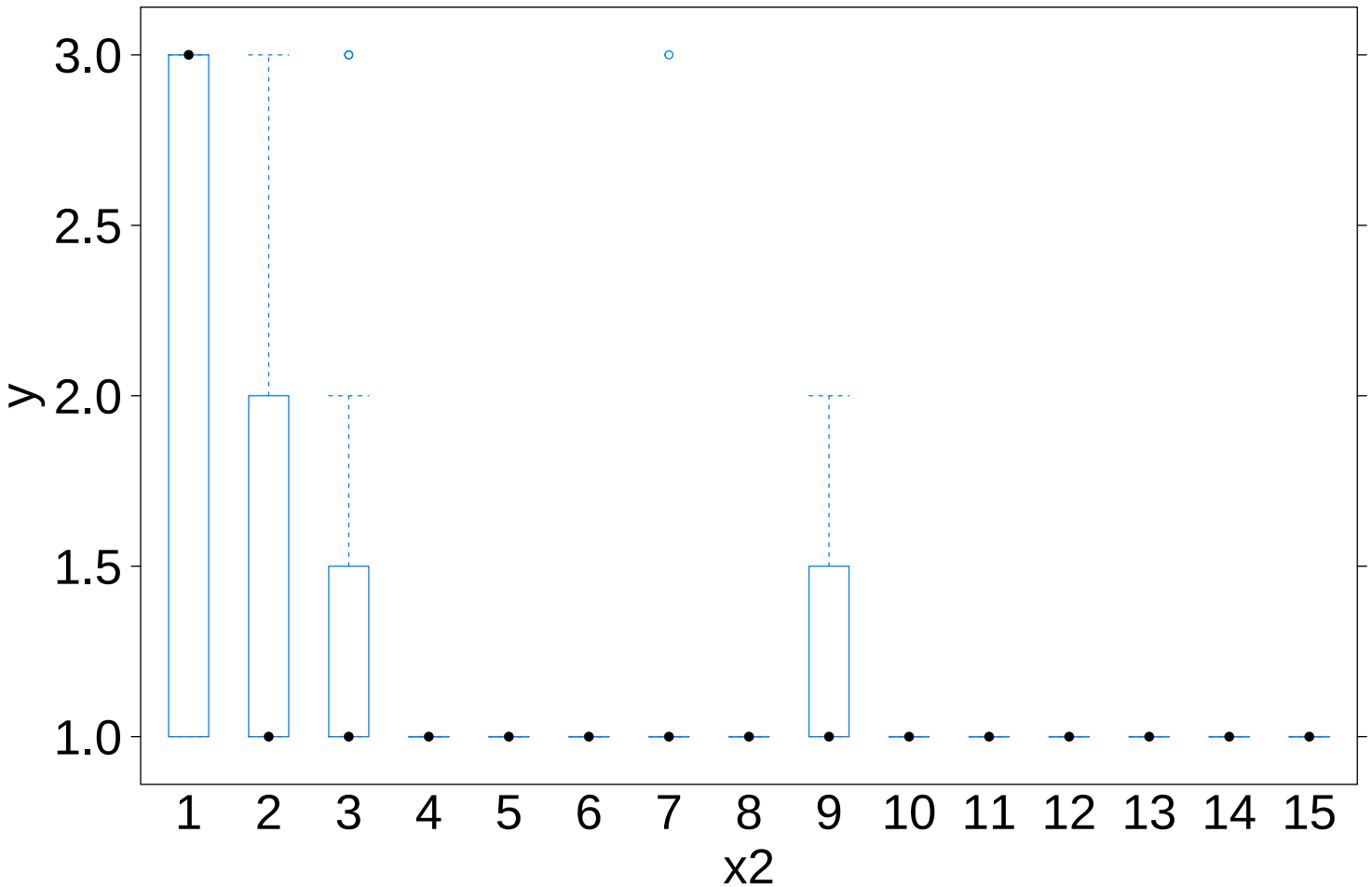


Figure : Sample dataset generated by CLM.

We can estimate the parameters easily using the R package `ordinal`:

Estimating with `ordinal`

```
> my.model <- clm(y~x1+x2, data=sample.data)
> confint(my.model)

       2.5 %      97.5 %
x1 -1.316768 -0.05512462
x2 -0.676549  0.42723845
```

The generative model has X_1's parameter at -0.5

One (unfortunately) typical approach to ordinal data is to use a Gaussian linear model, e.g. nested in an ANOVA approach.

- What happens if we use a *linear model* to estimate CLM data?
- How much *power* do we lose?
- Are our *false positives* inflated?

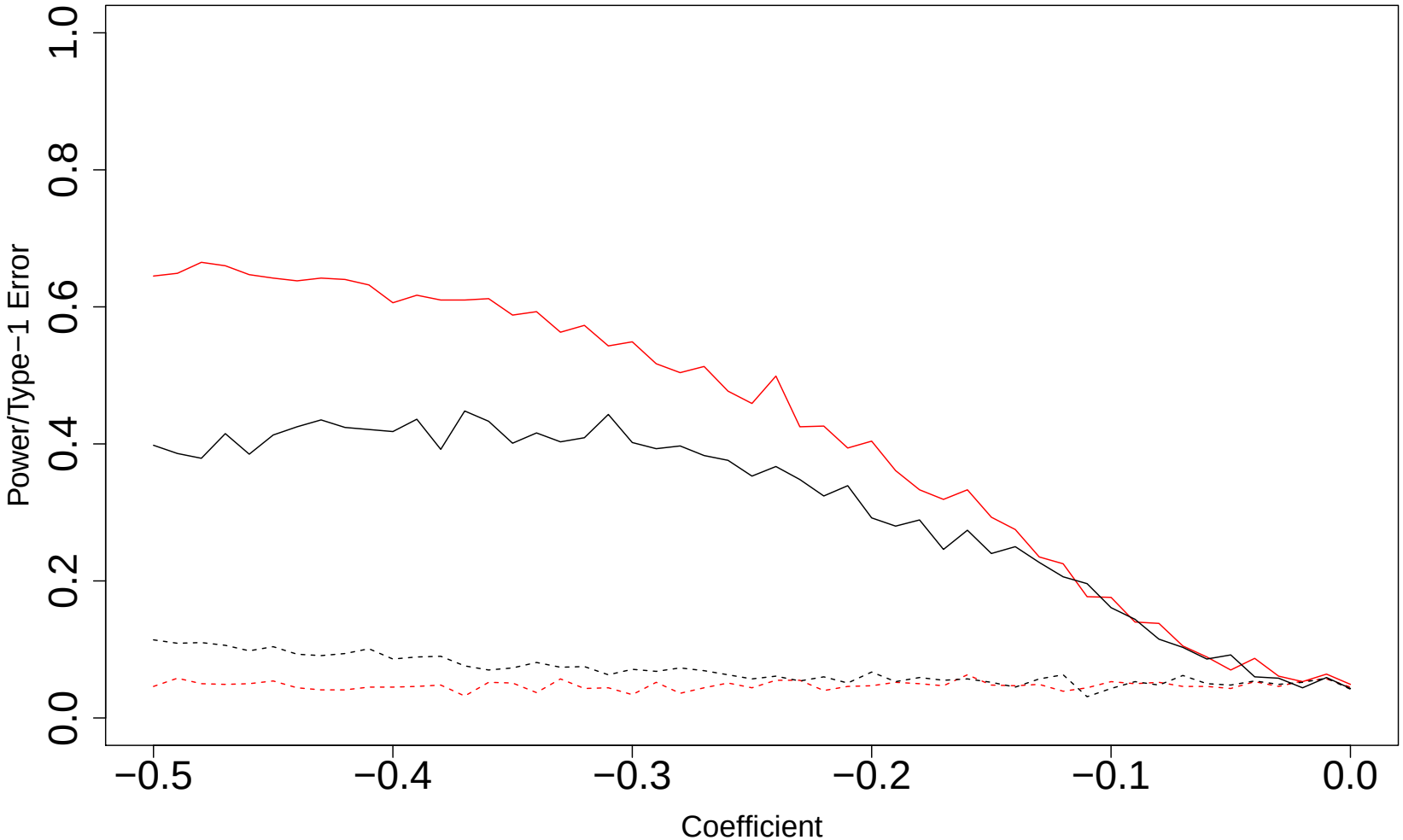


Figure : Red: CLM t-test. Black: LM t-test. Solid: X1. Dotted: X2.

Some other (more technical) reasons not to use a linear model (see Christensen 2015):

- *standard errors are inappropriate*: residuals cannot be normally distributed
- *inflation of information in the data*: linear models assume a scale between categories

Example

Volume 32, Issue 1 (March 2015)

< Previous Next >

Current Issue
Available Issues

Journal Information

Print ISSN: 0748-8068
Frequency: Quarterly

◂ Previous Article Volume 32, Issue 1 (March 2015) Next Article ▸

Add to Favorites Share Article Export Citations Track Citations (RSS | Email) Permissions

Full-text PDF

Article Citation:
Robert H. Sorensen, Matthew P. Connor, Joshua A. Lospinoso, Danielle J. Lospinoso, and Jose E. Barrera (*2015*) Combination Laser Resurfacing With Facial Plastic Surgery Is Superior to Lasers Alone. American Journal of Cosmetic Surgery: March 2015, Vol. 32, No. 1, pp. 12-17.

ORIGINAL ARTICLES

Combination Laser Resurfacing With Facial Plastic Surgery Is Superior to Lasers Alone

Robert H. Sorensen, MD, Captain, USAF, MC; Matthew P. Connor, MD, Captain, USAF, MC; Joshua A. Lospinoso, PhD, Captain, USA; Danielle J. Lospinoso, DO, Captain, USA, MC; Jose E. Barrera, MD, FACS, Lieutenant Colonel, USAF

From the Department of General Medical Education (Dr Sorensen), Department of Otolaryngology (Dr Connor, Dr Barrera), and Department of Dermatology (Dr D. J. Lospinoso), San Antonio Military Medical Center, Fort Sam Houston, Tex, and Portia Statistical Consulting LLC, San Antonio, Tex (Dr J. A. Lospinoso).

The authors listed have not published or submitted any related articles from this same study. All authors listed contributed such that they meet sufficient criteria to be listed as an author. None of the authors have financial or personal relationships that could inappropriately influence (or bias) the authors' decisions, work, or manuscript.

Corresponding author: Robert H. Sorensen, 9554 Dunstable, San Antonio, TX 78239 (e-mail: roberthsorensen@gmail.com).

Hypothesis: When treating rhytidosis, combining facial plastic surgery with CO_2 fractional laser treatment is superior to laser treatment alone.

Two pathologists evaluate Modified Fitzpatrick and Glogau before and after treatment

Table : Golgau via CLM estimation (probit)

	Estimate	p-value
Treatment	-0.78	0.07
Rater	-0.18	0.61
Est. Age	0.10	0.00
Fitz. 345	-1.42	0.00
Glogau 34	1.61	0.02

Table : Modified Fitzpatrick via CLM estimation (probit)

	Estimate	p-value
Treatment	-0.06	0.83
Rater	-1.29	0.00
Est. Age	0.06	0.00
Fitz. 345	-0.54	0.09
Glogau 34	0.15	0.71

References

P. McCullagh and J. Nelder (1989)

Generalized Linear Models, *Chapman and Hall*, Second Ed.

Sorensen, Robert H., et al. (2015)

"Combination Laser Resurfacing With Facial Plastic Surgery Is Superior to Lasers Alone." American Journal of Cosmetic Surgery 32.1 (2015): 12-17.

Christensen, R. H. B. (2015)

ordinal—Regression Models for Ordinal Data. R package version 2015.6-28. http://www.cran.r-project.org/package=ordinal/

Stuart A., Ord J.K, Arnold S. (1999)

Kendall's Advanced Theory of Statistics: Volume 2A—Classical Inference and the Linear Model, sixth edition.

Questions