Introduction to Data Science

Analysis of LOCUS Scores

Thomas Maierhofer

National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
University of California, Los Angeles

November 7th, 2017

Table of Contents

1. Introduction
2. Statistical Background
   - Handling Missing Observations
   - Mixed Model Representation of Rasch Model
3. Model for LOCUS Results
   - Proposed Model Formula
   - Preliminary Results

Introduction

Background

- 1555 students took pilot course Introduction to Data Science (IDS)
- Student performance was measured before and after completing IDS, using Levels of Conceptual Understanding in Statistics (LOCUS), see `https://locus.statisticseducation.org/`
- Research Question: Which parameters influence student performance?

Difficulties

- Some students have missing pretest and/or posttest scores
- The two LOCUS forms A (pretest) and B (posttest) were not administered to all students as designated

$\Rightarrow$ Original statistical analysis did not take into account missing values

Table of Contents

1. Introduction
2. Statistical Background
   - Handling Missing Observations
   - Mixed Model Representation of Rasch Model
3. Model for LOCUS Results
   - Proposed Model Formula
   - Preliminary Results

Handling Missing Observations

Original Analysis

$y_j^{post} = \beta_0 + \beta_1 y_j^{pre} + \dots$

- can not handle missing pretest or posttest score

New Analysis

$y_j = \beta_0 + \beta_1 \, TestTime_j + \dots$

- missing values do not matter
- two observations for students with pretest (TestTime $= 0$) and posttest (TestTime $= 1$) score, one observation for students with pretest or posttest score, no observations for students without any scores

for student/observation j.

Interpretation of Additional Parameters

Original Analysis

$y_j^{post} = \beta_0 + \beta_1 y_j^{pre} + \beta_2 PLE_j + \dots$

- β_0: estimate for reference group
- β_2: effect of PLE on posttest adjusted for pretest

New Analysis

$y_j = \beta_0 + \beta_1 TestTime_j + \beta_2 PLE_j + \beta_3 \{PLE * TestTime\}_j + \dots$

- β_1: improvement from pretest ($\text{TestTime} = 0$) to posttest ($\text{TestTime} = 1$) in reference group
- β_2: effect of Primary Language English (PLE) on pretest
- β_3: effect of PLE on improvement from pretest to posttest

for student/observation j.

Table of Contents

1. Introduction
2. Statistical Background
   - Handling Missing Observations
   - Mixed Model Representation of Rasch Model
3. Model for LOCUS Results
   - Proposed Model Formula
   - Preliminary Results

Mixed Model Representation of Rasch Model

Rasch Model (Rasch, 1993)

$$\mathbb{P}(y_{ij} = 1) = \frac{1}{1 + \exp\big(-(b_j - \delta_i)\big)}$$

$$\Leftrightarrow \text{logit}\big(P(y_{ij} = 1)\big) = b_j - \delta_i$$

with

- b_j: the ability of student j
- δ_i: the difficulty of question i

Equivalent Mixed Model (Kamata, 1998, 2001)

$$\text{logit}\big(\mathbb{P}(y_{ij} = 1)\big) = b_j - \delta_i$$

with

- b_j: a random intercept for student j
- δ_i: a fixed effect for question i

Table of Contents

1. Introduction
2. Statistical Background
   - Handling Missing Observations
   - Mixed Model Representation of Rasch Model
3. Model for LOCUS Results
   - Proposed Model Formula
   - Preliminary Results

Proposed Model Formula

$$\begin{aligned}\text{logit}\big(\mathbb{P}(y_{ijk} = 1 | x_{ijk})\big) &= \beta_0 + \delta_i + b_j + d_k + h_k \mathit{TestTime} + \\ &\quad \beta_1 \mathit{TestTime} + \beta_2 \mathit{PLE} + \beta_3 \{\mathit{PLE} * \mathit{TestTime}\} + .\end{aligned}$$

with

- $-\delta_i$: fixed effect for question i ($=$ question difficulty)
- b_j: random intercept for student j ($=$ student ability)
- d_k: random intercept for teacher k on pretest
- h_k: random slope for TestTime per teacher k (teacher effect on improvement from pretest to posttest)
- β_1: improvement in reference group
- β_2: effect of PLE (pretest)
- β_3: effect of PLE on improvement

for question i, student j, and teacher k.

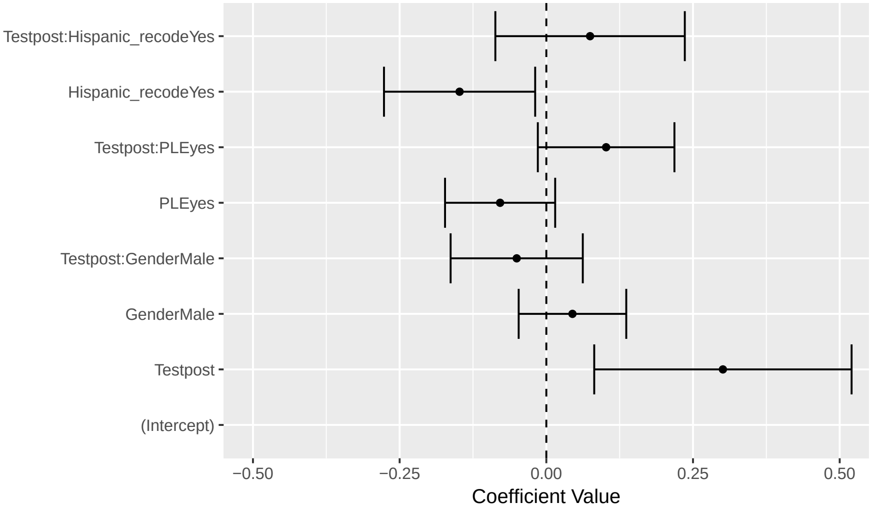
Table of Contents

1 Introduction

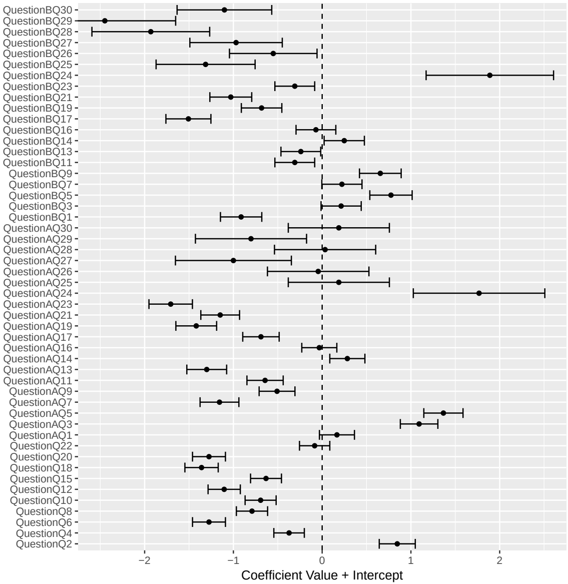
2 Statistical Background
- Handling Missing Observations
- Mixed Model Representation of Rasch Model

3 Model for LOCUS Results
- Proposed Model Formula
- Preliminary Results

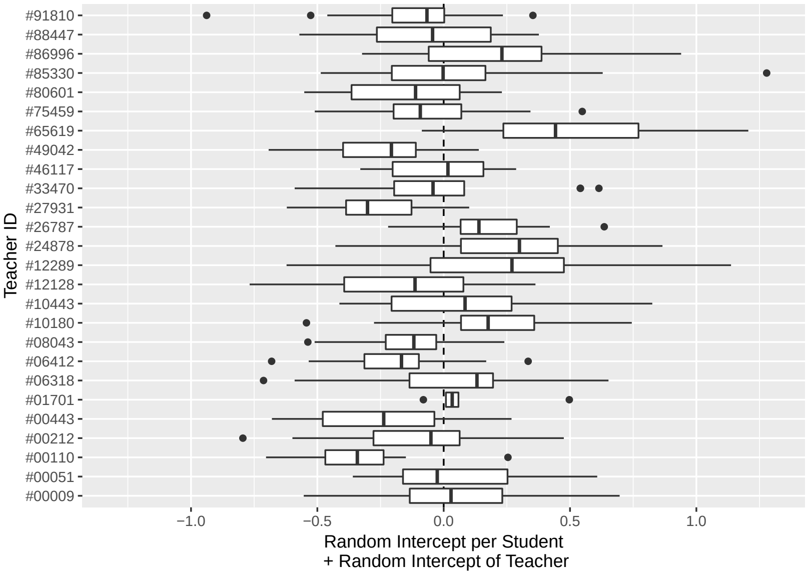
Coefficient Plot: Fixed Effects



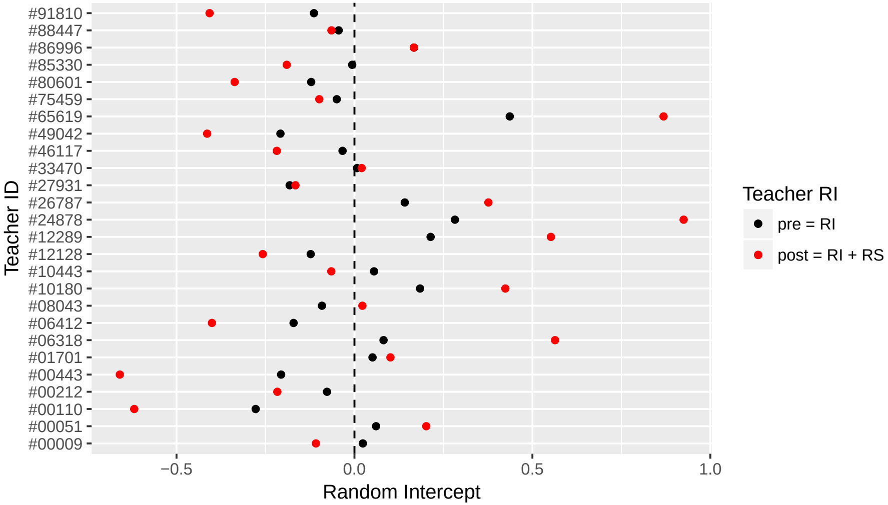
Coefficient Plot: Question Difficulty



Coefficient Plot: Student Ability



Coefficient Plot: Teacher Effects



Literature

Kamata, A. (1998). One-parameter hierarchical generalized linear logistic model: An application of HGLM to IRT.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement 38*(1), 79–93.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests.* ERIC.

Additional Material

R Code for Mixed Model

```
mgcv::gamm(Score ~ Question +
             Test * (Gender + Hispanic + PLE),
           random = list(Teacher_ID = ~1 + Test,
                         LAUSD_ID = ~1),
           family = "binomial",
           data = student_data_long)
```

Original Analysis

First Step: Rasch Model

$$\text{logit}\big(P(y_{ijk} = 1)\big) = b_{jk} - \delta_i$$

with

- b_{jk}: the ability of student j for teacher k
- δ_i: the difficulty of question i

Second Step: Mixed Model

$$b_{jk} = \beta_0 + h_k + \beta_1 y_{jk}^{pre} + \beta_2 PLE + \dots$$

with

- b_{jk}: estimated student ability from Rasch model
- h_k: random intercept for teacher k
- β_1 effect of pretest score on student ability
- β_2 effect of PLE on student ability