

Stat 414 – Day 8 Random Effects and Model Diagnostics

Last Time: Estimating Random Effects

- Random intercepts model $Y_{ij} = \beta_0 + u_j + \epsilon_{ij}$
 - Intercept is the expected (population mean) response for average subject
 - $\hat{\beta}_0 = \bar{y}$ if balanced or avg of \bar{y}_j
 - Fixed effect: $\hat{u}_j = \bar{y}_j$ (but then use pooled standard error)
 - Random effect: $\hat{u}_j = w_j \bar{y}_j + (1 - w_j) \bar{y}$ (“shrinkage estimator”) where $w_j = \tau^2 / (\tau^2 + \sigma^2/n_j)$ (aka the “reliability” of group j)

So then what is the standard error of \hat{u}_j ? Can we find a confidence interval around this estimator? Do we want to?

Fixed effect:

Confidence interval for Jones:

model1\$fitted.values[bball\$Player == "Jones"][1]	= .37957 – 0.17730 = 0.2023
new.dat <- data.frame(player="Jones") predict(model1, newdata = new.dat, interval = 'confidence')	.2023 ± t*(37)(.0448/sqrt(11)) = (.1749, .2296)

Confidence interval for Jones’ “effect”:

<pre> Coefficients: (Intercept) 0.379571 0.007523 50.453 < 2e-16 *** playerf1 -0.084571 0.016722 -5.058 1.18e-05 *** playerf2 -0.177298 0.013351 -13.280 1.19e-15 *** playerf3 -0.056237 0.016722 -3.363 0.0018 ** playerf4 0.170429 0.016722 10.192 2.72e-12 *** playerf5 -0.022753 0.013351 -1.704 0.0967 . </pre>	$-0.177298 \pm 2.026(.01335) = (-.204, -.150)$
--	--

Random effect:

<pre> Random effects: Groups Name Variance Std.Dev. Player (Intercept) 0.019648 0.14017 Residual 0.002008 0.04481 Number of obs: 43, groups: Player, 6 Fixed effects: Estimate Std. Error t value (Intercept) 0.37885 0.05772 6.564 </pre>	<pre> ranef(model2) \$Player (Intercept) Anderson -0.08244295 Jones -0.17494910 Mitchell -0.05458408 Rodriguez 0.16828693 Smith -0.02182608 Suarez 0.16551528 </pre>
--	---

Confidence interval for Jones’ “effect”:

$$\text{Comparative standard error for } u_j = \sqrt{\frac{1}{\frac{1}{\tau^2} + \frac{1}{n_j} + \frac{1}{\sigma^2}}}$$

<pre> ranef(model2, condVar= TRUE, drop=TRUE) \$Player Anderson Jones -0.08244295 -0.17494910 attr("postVar") [1] 0.0003290190 0.0001808414 </pre>	$-0.174949 \pm 2.026(.01345) = (-.2022, -.1477)$
---	--

Partly, you are more interested in the distribution of effects, rather than these individual players. We have estimated that this distribution is approximately normal with mean 0 and variance 0.01965.

(a) What is the expected average for a player in the 84th percentile?

So would we even look at these random effects estimates?

(b) What assumptions have we made about these “level 2 effects”?

We will add checking these residuals as part of our model diagnostics (Section 10.6).

- Check standardized level 2 residuals for normality though doesn't always guarantee real effects follow a normal distribution. Also check for unusual observations.
- Plot against the Level 2 units and other Level 2 variables (e.g., nonlinearity)
- Plot squared residuals against Level 2 variables to check for heteroscedasticity

(c) Try `> plot(model2)` and `> plot(ranef(model2))`

Example 2: (Example 4.1 in text): We want to predict language test scores (langPOST) in Grade 8 students (~ age 11) in elementary schools in the Netherlands based on their (verbal) IQ.

(a) Identify the fixed and random effects in this context. Identify level 1 and level 2.

(b) Use R to create the null model. How many students and how many schools are in the dataset? Does this model appear to be valid?

(c) Using the null model, what do you predict for the language score of a randomly selected student?

(d) What standard error would you put around your estimate? Is this the same as the standard deviation of all the language scores in the sample? Why or why not?

(e) Is it reasonable to pick out the schools with the largest positive residuals and conclude they are doing something better than the other schools?

(f) Now we want to include pupil (verbal) IQ as a predictor of language test performance. Is this a Level 1 or Level 2 predictor?

(g) Write out an appropriate statistical “random intercepts” model. How many parameters?

(h) Include the IQ variable in the model and provide an interpretation of the slope and intercept from the parameter estimates. What would a graph of this model look like? Does this model appear to be valid?

(i) Is the effect of IQ statistically significant?

(j) What is a “typical deviation” in these regression lines from the overall regression line? Are these distances negligible? What is a “low line”? What is a “high line”?

(k) What is a “typical deviation” of a student from his/her school regression line? More or less than between the lines?

(l) What is the “residual” intraclass correlation coefficient?

(l) How do the residual variance and the random intercept variance compare to the empty model? How do the likelihood and AIC/BIC values compare? Telling you?

(m) Now that we have “controlled for IQ,” can we pick out the schools with the most positive random effects and declare them superior?

Model for hurricane data (HW 3) effect coding for hurricane gender:

$$Y_{ij} = \beta_0 + \beta_1 \times \text{name.gender}_{ij} + u_j + \epsilon_{ij}$$

- Intercept: Expected rating for “average hurricane” for average subject
- $2\beta_1$ = average rating for females – average rating for males, adjusted for subject
- With lmer, confint gives you confidence intervals for the fixed and random components (with lme, use intervals()), results will vary slightly)

Approach 1: Paired analysis on the means (diff = avg.male – avg.female)

```

      Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.2257     0.0497  4.541 0.00000773 ***
    
```

We are 95% confident that the mean rating is 0.128 to 0.323 higher for male names compared to female names.

Approach 2: Mixed model with hurricane-name-gender as fixed effect (1 = female, 0 = male)

Fixed effects:

```

      Estimate Std. Error t value
(Intercept)  4.38555     0.04472  98.073
HurrGend     -0.22572     0.04255  -5.305
HurrGend     -0.3091269 -0.1423182
    
```

Approach 3: Ignore subject/treat all observations as independent

```

Coefficients:
      Value Std.Error t-value p-value
(Intercept) 4.385549 0.03351924 130.83675 0
HurrGend    -0.225723 0.04740337 -4.76174 0
    
```

```

> confint(gls(Score ~ HurrGend, data = hurr_long))
      2.5 % 97.5 %
(Intercept) 4.3198526 4.4512456
HurrGend    -0.3186314 -0.1328136
    
```

All these approaches are assuming the hurricane-name-gender effect is the same across the subjects, but that there is subject-to-subject variation in ratings.

Also see new review problem.