Multilevel Linear Regression Models

Brady T. West

Review: The European Social Survey (ESS) Data

- **Data** collected in face-to-face interviews from national sample of 1,703 adults in Belgium.

Review: The European Social Survey (ESS) Data

- **Data** collected in face-to-face interviews from national sample of 1,703 adults in Belgium.
- **Variables:** respondent ID, interviewer ID, 22 variables measuring attitudes and opinions of respondents on various topics … **interested in interviewer effects on data!**

Review: The European Social Survey (ESS) Data

- **Data** collected in face-to-face interviews from national sample of 1,703 adults in Belgium.
- **Variables:** respondent ID, interviewer ID, 22 variables measuring attitudes and opinions of respondents on various topics … **interested in interviewer effects on data!**
- Have final respondent **weights** (based on complex sample design), along with interviewer-specific response rates (percentage scale).

Revisiting Random Effects

Random effects: random variables that allow coefficients in regression model to randomly vary depending on randomly selected cluster *(e.g., neighborhood)* or subject *(e.g., in a longitudinal study)*

Revisiting Random Effects

Random effects: random variables that allow coefficients in regression model to randomly vary depending on randomly selected cluster *(e.g., neighborhood)* or subject *(e.g., in a longitudinal study)*

Multilevel Models also known as:

Random coefficient models

Varying coefficient models

Subject-specific models

Hierarchical linear models

Mixed-effects models

Example Model Specification

Model for a **continuous dependent variable *Y***,
measured on **person *i*** within **cluster *j***

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + u_{1j} x_{1ij} + e_{ij}$$

Fixed effects

Random effects

Error

Example Model Specification, cont'd

- **Fixed effects:** *regression coefficients* or *regression parameters* ~ *Unknown constants* defining relationships between predictors and dependent variables that we wish to estimate.
- **Random effects:** *random variables*; need to define distributions!

Example Model Specification, cont'd

- **Fixed effects:** *regression coefficients* or *regression parameters* ~ *Unknown constants* defining relationships between predictors and dependent variables that we wish to estimate.
- **Random effects:** *random variables*; need to define distributions!

Recall: Multilevel model because have **explicit interest** in estimating variance of random cluster effects!

Example Model Specification, cont'd

Common distributions for random effects and random error terms ~ Normal with mean 0 and specified variances and covariances

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + u_{0j} + u_{1j} x_{1ij} + e_{ij}$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix} \equiv D \right)$$

$$e_{ij} \sim N(0, \sigma^2)$$

Variance-covariance Matrix of Random Effects (*D*)

Errors, independent of random effects

Multilevel Specification

Alternative way of specifying model

Level 1: $$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + e_{ij}$$

Random coefficients (**not parameters!**)

Level 2:
$$\beta_{0j} = \beta_0 + u_{0j}$$
$$\beta_{1j} = \beta_1 + u_{1j}$$

When combined, we have the same model!

Why the Multilevel Specification?

Level 1:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + e_{ij}$$

Level 2:

$$\beta_{0j} = \beta_0 + u_{0j}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

- Specification clearly defines role of covariates measured at higher levels in multilevel models
- View each **Level 2** equation for random coefficient as intercept-only regression model *(where DV is a random coefficient)*!
- Explain variance in random effects by adding fixed effects of Level-2 covariates to models!

Why the Multilevel Specification?

$$\beta_{0j} = \beta_0 + u_{0j}$$

- Fit model, compute estimated variance of random intercepts:

$$\hat{\sigma}_0^2 = 2$$

- Include fixed effect of subject gender in model
  (assume a longitudinal study): $$\beta_{0j} = \beta_0 + \beta_2 MALE_j + u_{0j}$$

- Now, $\hat{\sigma}_0^2 = 1 \rightarrow$ explained 50% of variance in intercepts with fixed effect of gender!

Estimating the Model Parameters

Computational technique

MLE = maximum likelihood estimation

Idea: What values of model parameters that would make observed data ***most likely***?

Estimating the Model Parameters

Computational technique

MLE = maximum likelihood estimation

Idea: What values of model parameters that would make observed data ***most likely***?

Use software like Python to compute MLEs of fixed effects and variance components, in addition to standard errors

Testing the Model Parameters

Compute confidence intervals or test hypotheses for model parameters

Test null hypotheses (e.g., fixed effect is zero, or variance component is zero – random effects don't vary!), can use **likelihood ratio testing**

Idea: Does probability *(likelihood)* of observed data change substantially if we remove a given parameter (or parameters) from model?

Testing the Model Parameters

Compute confidence intervals or test hypotheses for model parameters

Test null hypotheses (e.g., fixed effect is zero, or variance component is zero – random effects don't vary!), can use **likelihood ratio testing**

Idea: Does probability *(likelihood)* of observed data change substantially if we remove a given parameter (or parameters) from model?

Reading this week: provides specific details on how to perform these types of tests for parameters in multilevel models!

ESS Example

- **Interviewers** in ESS = random selections from a larger pool of interviewers that might have been hired.
- Relationship of **trust in police** (TRSTPLC) with person's **attitude** about whether people generally try to help others (PPLHLP).
- **Observations clustered by interviewer** ~ random effects can account for this.
- **Fit multilevel model** to see if interviewers are having an effect on intercept and/or slope in our model!

An Example: Interpretation

MLE of fixed effect of TRSTPLC is positive (0.14) and significant ($p < 0.01$)
$\rightarrow$ those with higher levels of trust in police tend to have higher levels of faith in people helping others

An Example: Interpretation

MLE of fixed effect of TRSTPLC is positive (0.14) and significant ($p < 0.01$)
→ those with higher levels of trust in police tend to have higher levels of faith in people helping others

MLE of intercept (3.89) is also significant ($p < 0.01$)
→ mean on help scale (0 to 10) for those with zero trust in police

An Example: Interpretation

Estimated **variance** of random **intercepts** = 0.696
Estimated **variance** of random **slopes** = 0.012
Both significant based on likelihood ratio tests!

An Example: Interpretation

Estimated **variance** of random **intercepts** = 0.696
Estimated **variance** of random **slopes** = 0.012
Both significant based on likelihood ratio tests!

Interviewers are varying significantly around overall fixed effects; they have unique intercepts and unique slopes!

Model Diagnostics

Examine whether our **assumptions** about distributions of random effects and random errors were **reasonable**!

Does the **model** seem to **fit well**?

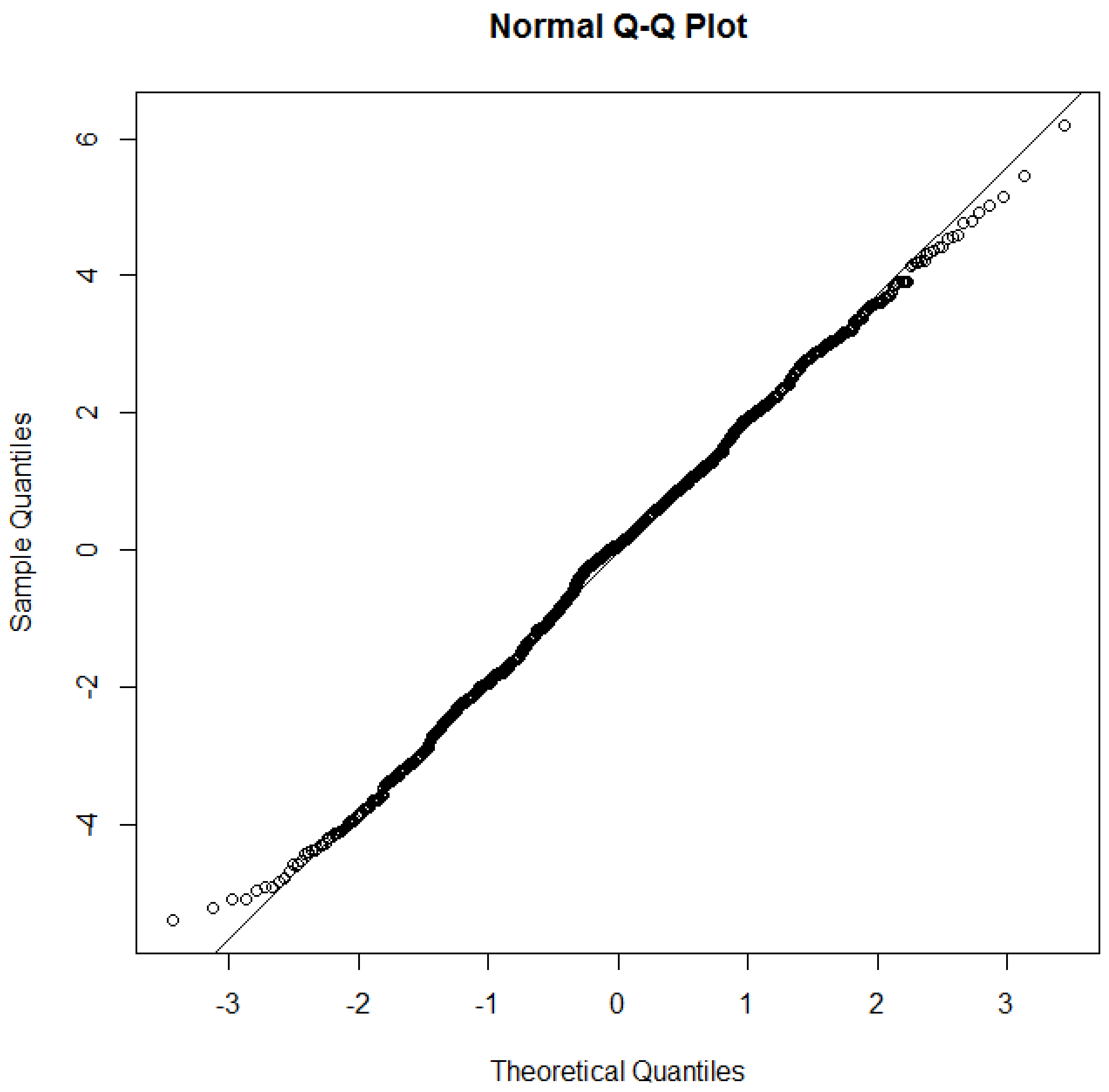
Model Diagnostics

Examine whether our **assumptions** about distributions of random effects and random errors were **reasonable**!

Does the **model** seem to **fit well**?

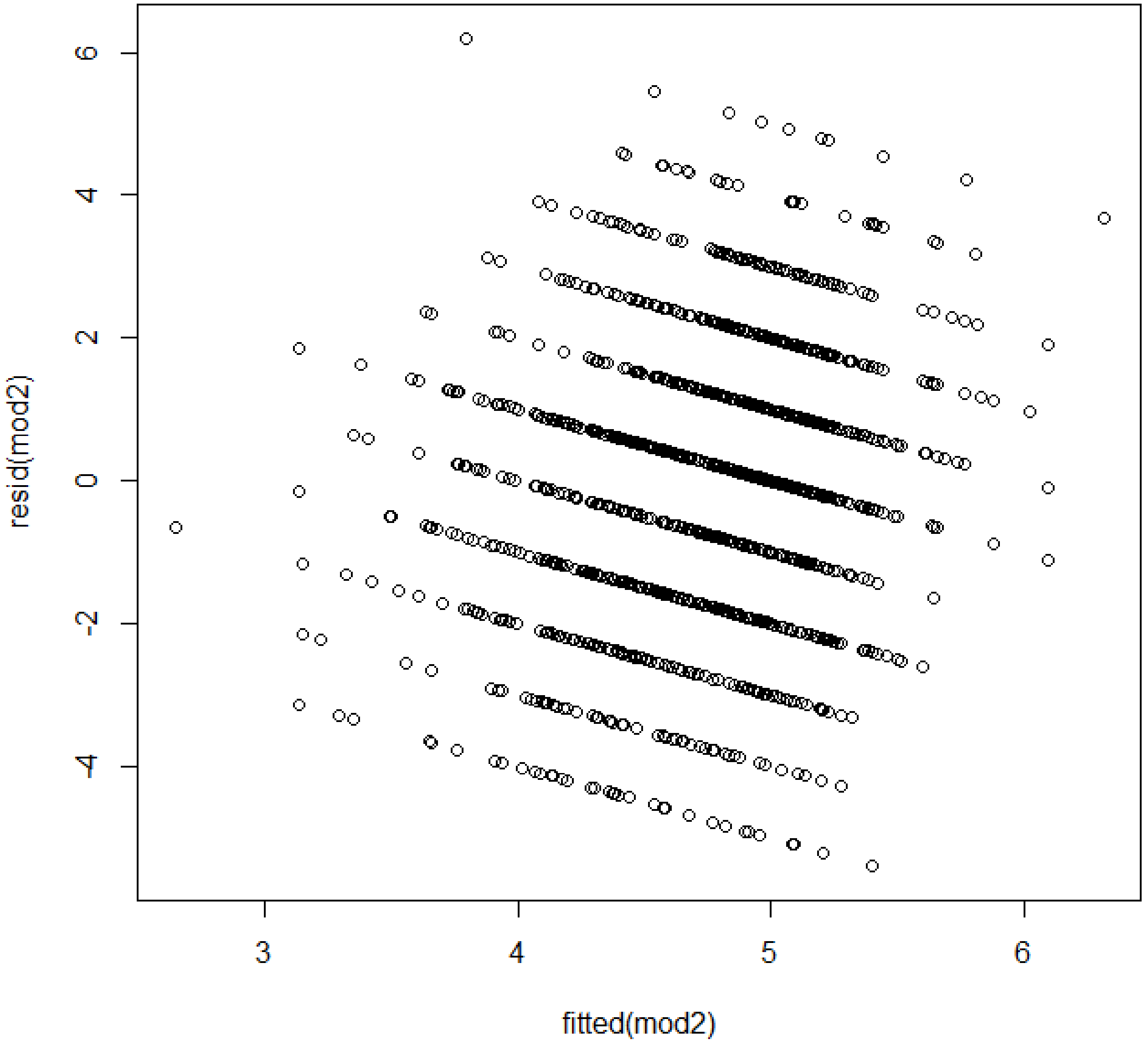
1. Look at **distribution of residuals** *(just like in linear regression!)*
2. Look at distributions of **predicted** values of random interviewer effects, or EBLUPs; are there outliers?

Residual Diagnostics: Normality



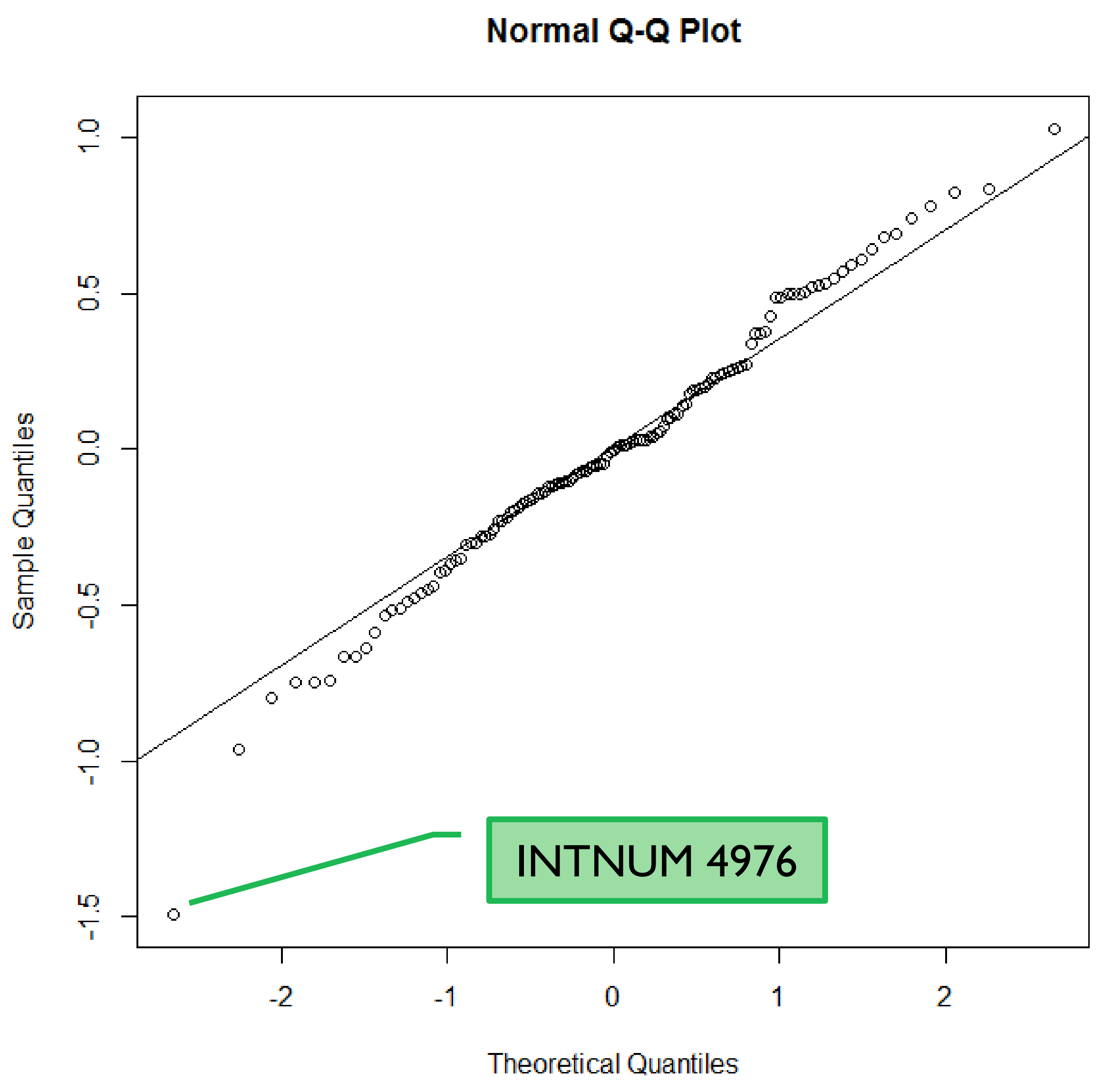
QQ plot suggests residuals are normally distributed + no outliers!

Residual Diagnostics: Constant Var.



Scatterplot of residuals against fitted values suggests **no concerns with constant variance of errors**

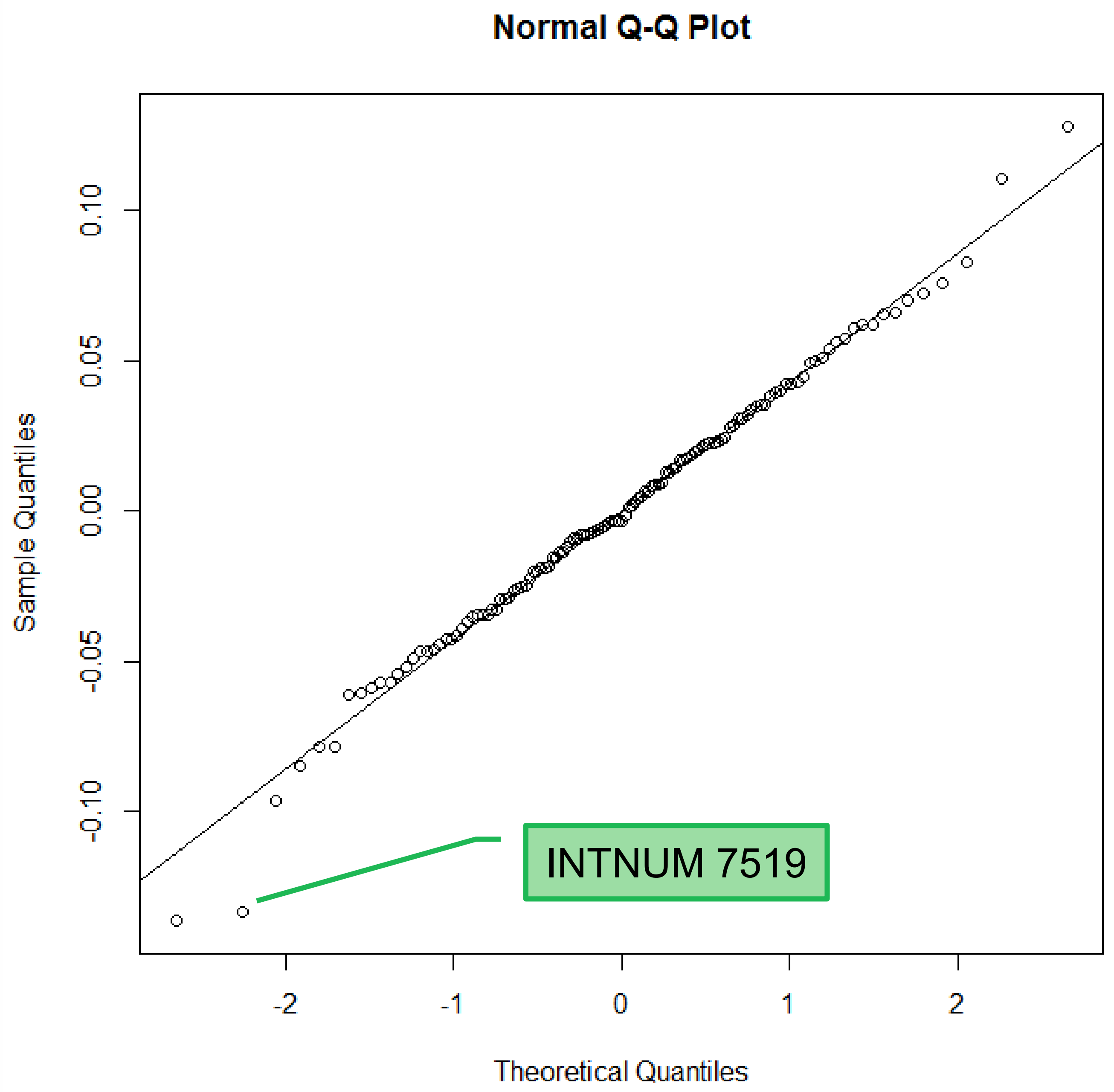
EBLUPs for Random Intercepts



QQ plot suggests **random effects on intercept normally distributed**

One outlier = Interviewer #4976

EBLUPs for Random Slopes

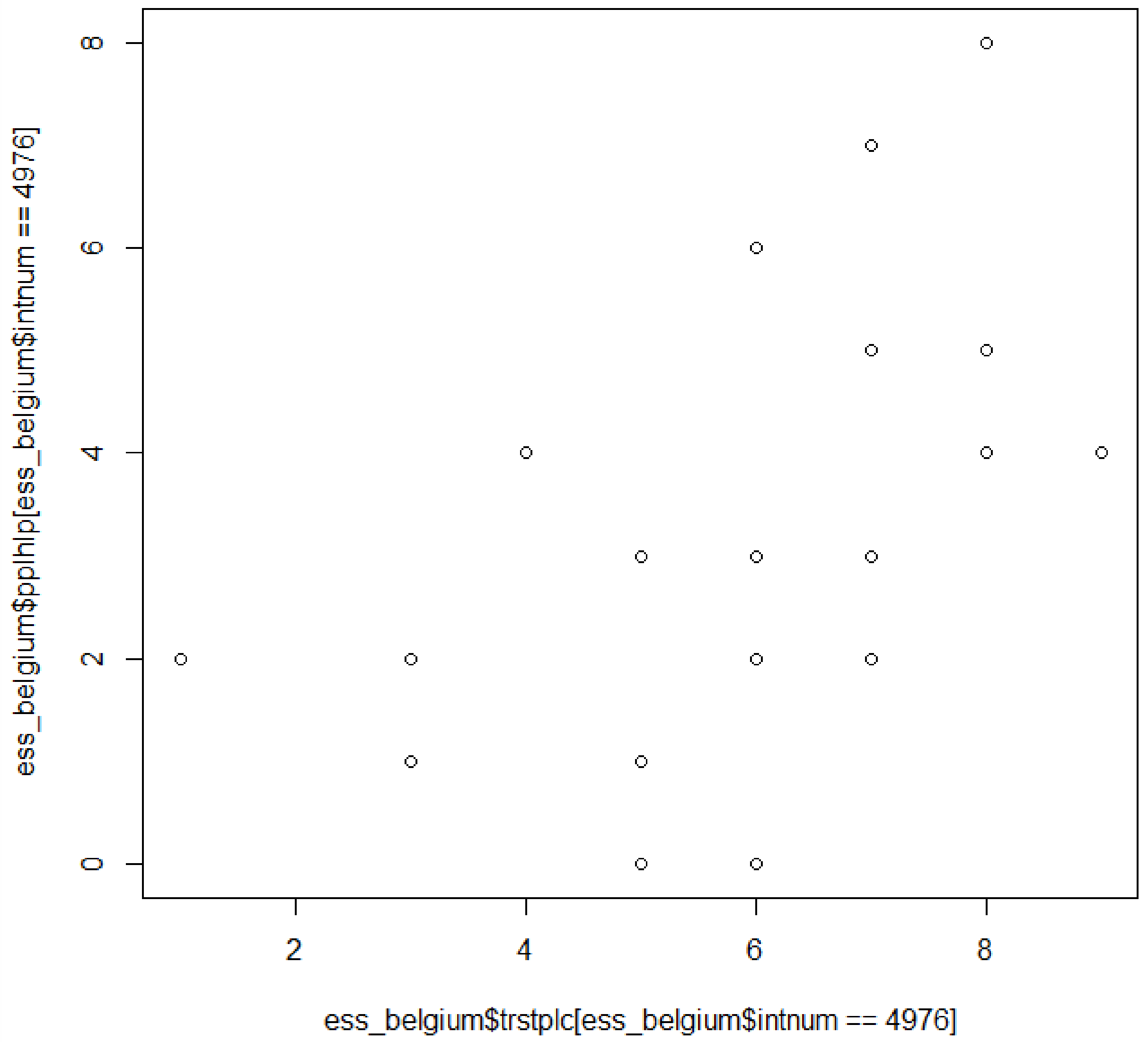


QQ plot suggests
random effects on slope are normally distributed

One outlier = Interviewer #7519

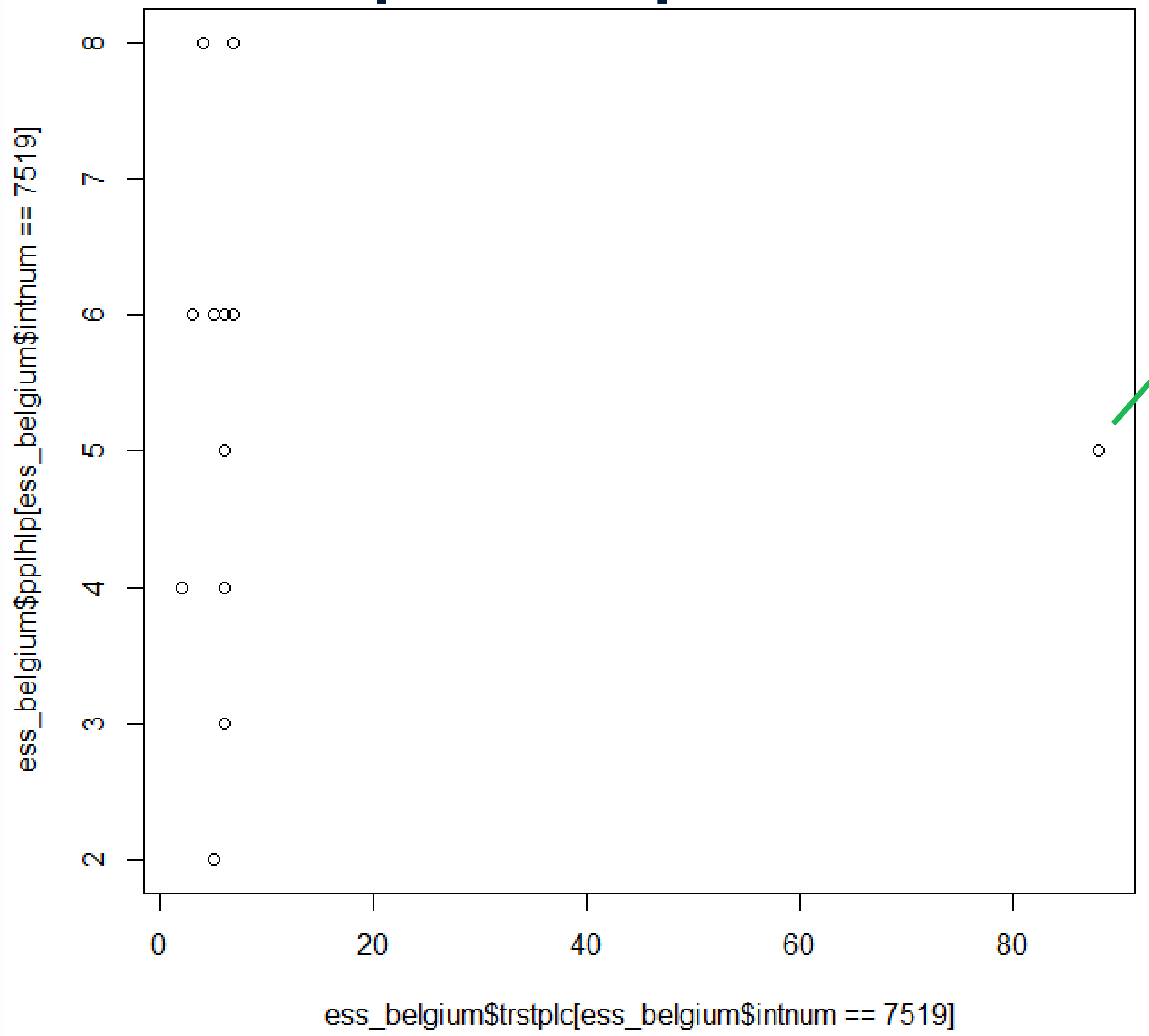
Look at the Data for the Outliers

Interviewer 4976: many responses < 4 for helpfulness!



Look at the Data for the Outliers

Interviewer 7519: unique slope caused by missing data!



!!!!
88 = missing data
…need to rerun!

Conclusions from Example

- ESS interviewers producing unique intercepts and unique slopes
- Variance not necessarily good: adds uncertainty to estimates of parameters! Should re-evaluate variance after removing outliers.
- If each interviewer working random subsample of full sample, should produce similar intercepts and slopes, assuming common model holds

Conclusions from Example

- ESS interviewers producing unique intercepts and unique slopes
- Variance not necessarily good: adds uncertainty to estimates of parameters! Should re-evaluate variance after removing outliers.
- If each interviewer working random subsample of full sample, should produce similar intercepts and slopes, assuming common model holds

Next step: add interviewer-level covariates to level-2 equations for random intercepts and slopes to see if explains this variance … Hypothesize **interviewer attitudes** explain some of variance!

What's Next?

- **What if dependent variable is binary?** → multilevel **logistic** regression models for binary variables in clustered data sets
- **Revisit** logistic regression model for smoking (NHANES)

Deep dive reading on multilevel linear regression models:
West, Welch, and Galecki (2014), *Linear Mixed Models*