Mixed model assumptions

Last time: mixed model assumptions

$$Price_{ij} = \beta_0 + \beta_1 Satisfaction_{ij} + u_i + \varepsilon_{ij}$$

$u_i \overset{iid}{\sim} N(0, \sigma_u^2) \quad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$

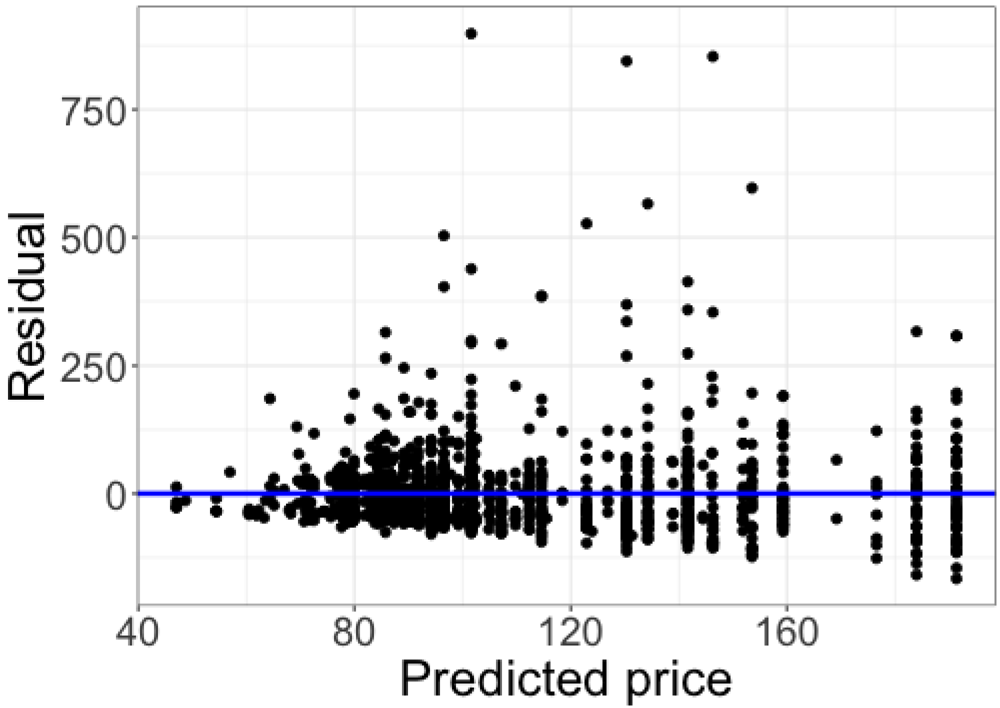
- **Shape assumption:**
  - the overall relationship between satisfaction and price is linear
  - The slope is the *same* for each neighborhood
- **Constant variance assumption:**
  - ε_{ij} has the same variance σ_ε^2 regardless of satisfaction or neighborhood

> How do you think we could check the shape and constant variance assumptions?

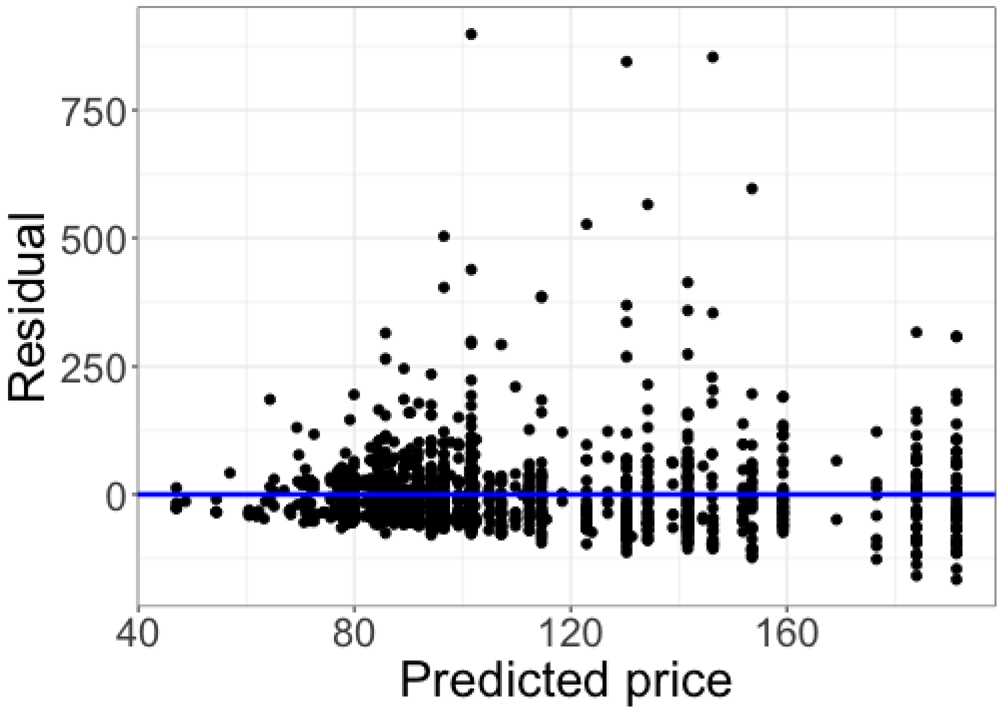
Residual plots

Residuals: $Price_{ij} - \widehat{Price}_{ij}$, where

$$\widehat{Price}_{ij} = \widehat{\beta}_0 + \widehat{\beta}_1 Satisfaction_{ij} + \widehat{u}_i$$



Residual plots



> Do the shape and constant variance assumptions look reasonable?

Checking assumptions

$$Price_{ij} = \beta_0 + \beta_1 Satisfaction_{ij} + u_i + \varepsilon_{ij}$$

$$u_i \overset{iid}{\sim} N(0, \sigma_u^2) \qquad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

- **Normality assumption:** Both $u_i \sim N(0, \sigma_u^2)$ and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

> How do you think we could check the normality assumption?

QQ plots

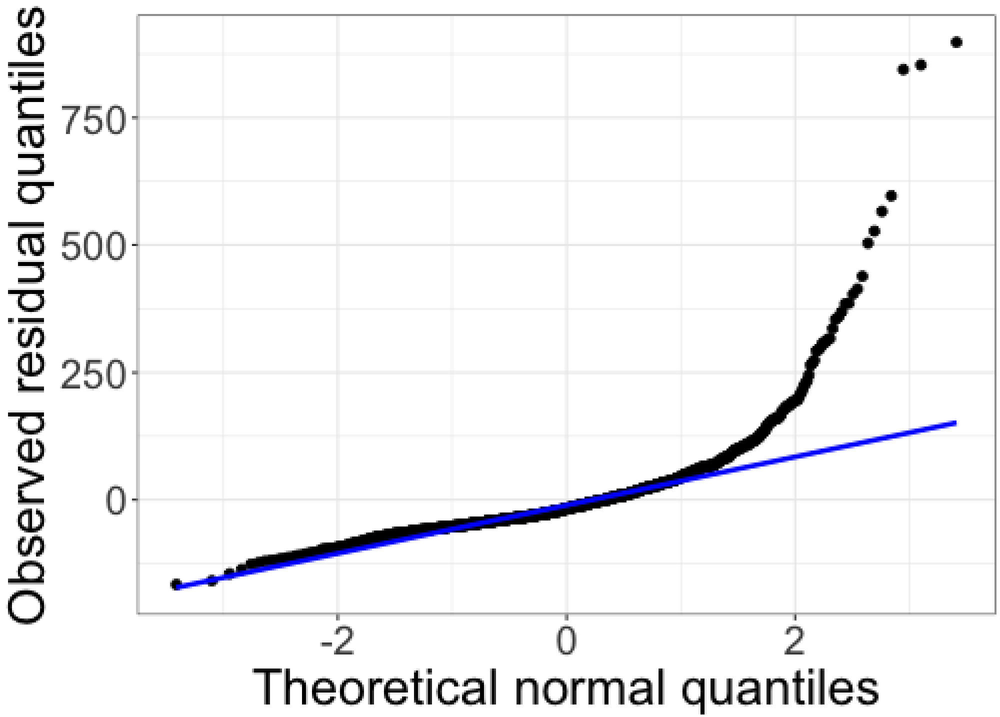
Assumption: $u_i \sim N(0, \sigma_u^2)$

- Check whether the random effect estimates $\widehat{u}_i$ appear normal with a QQ plot

Assumption: $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

- Check whether the residuals appear normal with a QQ plot

QQ plot for the residuals



Do the residuals appear normal?

QQ plots for random effects

To create a QQ plot for the random effects, I need estimates $\widehat{u}_i$ of the random effects for each group.

> How would I calculate $\widehat{u}_i$?

Estimated random effects

R calculates an estimated random effect for each group (i.e., neighborhood):

```
m1 <- lmer(price ~ overall_satisfaction +
             (1 | neighborhood),
           data = bnb)
coef(m1)
```

```
## $neighborhood
##                          (Intercept) overall_satisfaction
## Albany Park                16.367331             14.80912
## Archer Heights              9.863461             14.80912
## Avondale                   40.533851             14.80912
## Beverly                    21.046464             14.80912
## Bridgeport                 13.304517             14.80912
## Brighton Park              28.548742             14.80912
## Burnside                   12.741349             14.80912
## Calumet Heights            11.465091             14.80912
```

Estimated random effects

```
coef(m1)
```

```
...
##                      (Intercept) overall_satisfaction
## Albany Park            16.367331             14.80912
## Archer Heights          9.863461             14.80912
## Avondale               40.533851             14.80912
## Beverly                21.046464             14.80912
## Bridgeport             13.304517             14.80912
## Brighton Park          28.548742             14.80912
...
```

> What is the same for every neighborhood?

Estimated random effects

```
coef(m1)
```

```
...
##                      (Intercept) overall_satisfaction
## Albany Park            16.367331             14.80912
## Archer Heights          9.863461             14.80912
## Avondale               40.533851             14.80912
## Beverly                21.046464             14.80912
## Bridgeport             13.304517             14.80912
## Brighton Park          28.548742             14.80912
...
```

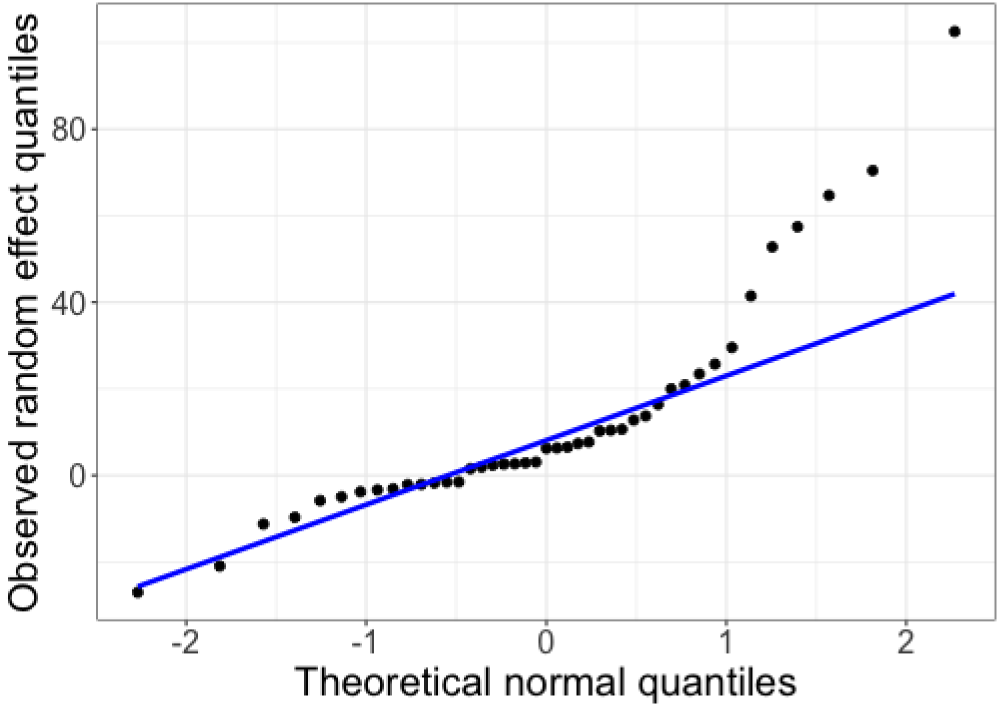
> What is *different* for each neighborhood?

Estimated random effects

```
...
##                        (Intercept) overall_satisfaction
## Albany Park              16.367331             14.80912
## Archer Heights            9.863461             14.80912
## Avondale                 40.533851             14.80912
## Beverly                  21.046464             14.80912
## Bridgeport               13.304517             14.80912
## Brighton Park            28.548742             14.80912
...
```

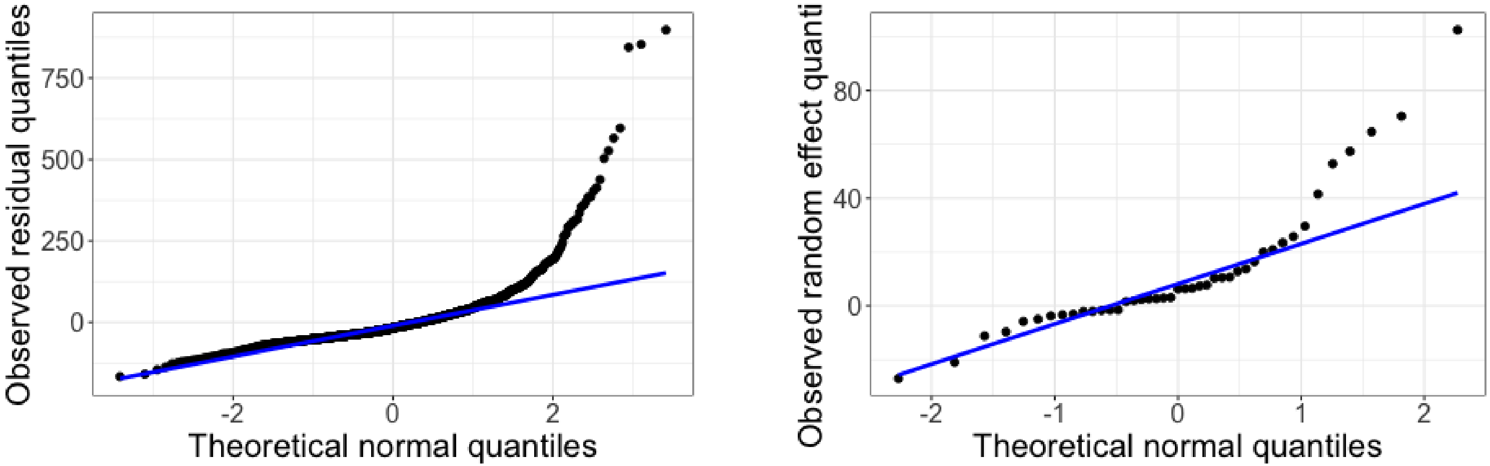
> How do I get the random effect estimates $\widehat{u}_i$?

QQ plot for the random effects



Do the random effects appear normal?

Addressing assumption violations

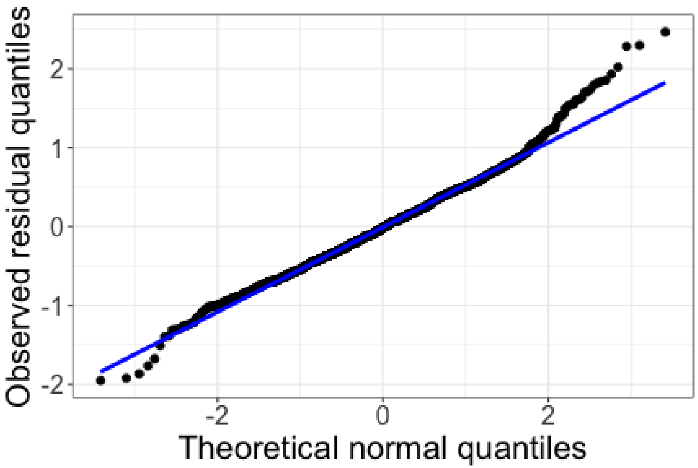


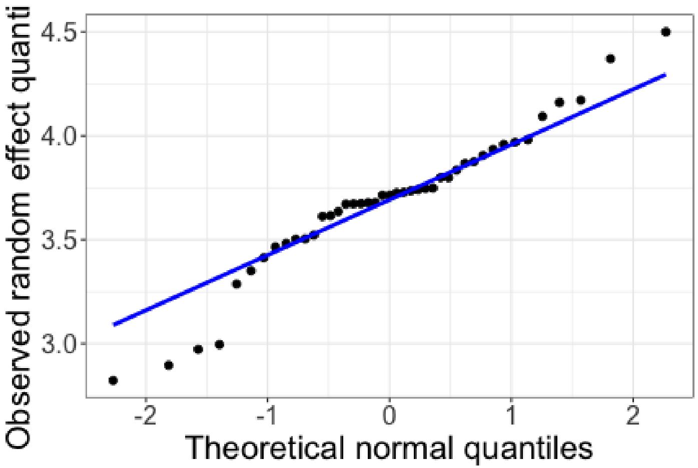
> How could we address violations of the normality assumptions?

Transformations

$$\log(Price_{ij}) = \beta_0 + \beta_1 Satisfaction_{ij} + u_i + \varepsilon_{ij}$$

$$u_i \overset{iid}{\sim} N(0, \sigma_u^2) \qquad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$





Revisiting the shape assumption

$$Price_{ij} = \beta_0 + \beta_1 Satisfaction_{ij} + u_i + \varepsilon_{ij}$$

$$u_i \overset{iid}{\sim} N(0, \sigma_u^2) \qquad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

- This model assumes that the slope is the same for each neighborhood

> How can we change the model to allow the slope to be *different* in different neighborhoods? Discuss with your neighbor for 1-2 minutes, and try to write down what the model would look like. Then we will discuss as a group.

Adding random slopes

$$Price_{ij} = \beta_0 + u_i + (\beta_1 + v_i)Satisfaction_{ij} + \varepsilon_{ij}$$

- β_0 = mean price when satisfaction is 0 (average across neighborhoods)
- $\beta_0 + u_i$ = mean price when satisfaction is 0 in neighborhood i
- β_1 = average change in price for a one-unit increase in satisfaction (average across neighborhoods)
- $\beta_1 + v_i$ = average change in price for a one-unit increase in satisfaction in neighborhood i

Class activity

https://sta214-f22.github.io/class_activities/ca_lecture_29.html

Class activity

Mixed effects models are useful when there are group effects in our data.

> What are the groups in the data, and what are the observations within each group?

Class activity

The researchers hypothesize that anxiety levels depend on the type of performance (large or small ensembles), and that the difference in anxiety levels between large and small ensembles varies from person to person.

> What mixed effects model should the researchers use to investigate their hypothesis?

Class activity

$$Anxiety_{ij} = \beta_0 + u_i + (\beta_1 + v_i)LargeEnsemble_{ij} + \varepsilon_{ij}$$

Interpret the fixed effects and random effects in the model.

Class activity

$$Anxiety_{ij} = \beta_0 + u_i + (\beta_1 + v_i)LargeEnsemble_{ij} + \varepsilon_{ij}$$

> Interpret the fixed effects and random effects in the model.

- β_0 = average performance anxiety before small ensemble and solo performances (average across musicians)
- $\beta_0 + u_i$ = average performance anxiety before small ensemble and solo performances for musician i
- β_1 = average difference in anxiety before large ensemble performances (compared to small/solo performances) (average across musicians)
- $\beta_1 + v_i$ = average difference in anxiety before large ensemble performances for musician i