Lab 11 - Math 58B: SLR inference

due Tuesday April 21, 2026

your name here

```
library(tidyverse)
library(broom)
library(praise)
```

Lab Goals

The goal of the lab is to put into practice the tools we've been using to analyze linear models. In particular, we will:

- run a linear model
- remember R^2
- perform inference on the linear model
- check linear model technical conditions using residual plots
- create both prediction and mean intervals for a given "X" value

Getting started

For formatting the the linear model output, use the **broom** package. Also, **tidyverse** will continue to be used for `ggplot()` and data wrangling.

For using data wrangling verbs, you might see: https://r4ds.had.co.nz/transform.html (remembering the verbs `filter()`, `arrange()`, `select()`, `mutate()`, `summarize()`, and `group_by()`.)

For `ggplot()`, look for inspiration on this cheat sheet:

https://raw.githubusercontent.com/rstudio/cheatsheets/main/data-visualization.pdf

Load packages & data

The data is on homes from Ames, IA from the **openintro** package.

> Data set contains information from the Ames Assessor's Office used in computing assessed values for individual residential properties sold in Ames, IA from 2006 to 2010.

```
library(tidyverse)
library(openintro)
data(ames)
```

To Turn In

Q1. Look at the data

```
?ames
```

What are the observational units? What are the variables? Are the variables numeric or categorical? How many levels does each categorical variable have? (Just report two or three things that you see when you look at the information about the data.)

Q2. Plots to predict `price`

Include at least three different plots whose aim is to identify a linear relationship for predicting the home `price`. Which of the three (or more) variables do you think is the best predictor of `price`? Explain.

Q3. Residual plots

Using the same three variables as above, create three different residual plots. For each residual plot, describe any violations of the technical conditions that you see. Your code should be something like this:

```
lm(responsevariable ~ explanatoryvariable, data = yourdataname) |>
     augment() |>  # in the broom package
     ggplot(aes(x = .fitted, y = .resid)) +
     geom_point() +
     geom_hline(yintercept = 0)
```

Q4. Transform the data.

Using transformations like the natural log (`log()`) or the square (`^2`) or the square root (`sqrt()`), transform either the explanatory or response (or both!) variables in order to create residual plots that do not show violations of the technical conditions.

Hint: you probably won't get perfect residual plots, but they should be better than Q2!

Q5. Find a model

Using the variables from Q4 that led to residual plots with the best technical conditions, find the estimated linear model (least squares regression line) and write it down. For example, you might have:

$$\widehat{\sqrt{price}} = 376.9 + 65.69 \cdot Fireplaces$$

Hint: Your R code might look like this:

```
lm(responsevariable ~ explanatoryvariable, data = yourdataname) |>
     tidy()   # in the broom package
```

Q6. Inference on the slope

Run a full hypothesis test to assess the discernibility of the slope from the model in Q5. Include the hypotheses, test statistic, p-value, and conclusion. Remember that your conclusion will be in the transformed space.

Q7. Interval estimates

Using the model in Q5, create two types of intervals: a confidence interval for the mean response and a prediction interval for the individual responses.

After you create the intervals, back transform the response variable and confirm that the intervals make sense by looking at the plot in Q2. Explain.

Your R code will look something like:

```
lm(sqrt(price) ~ Fireplaces, data = ames) |>
  augment(newdata=data.frame(Fireplaces = 3), interval = "confidence")

lm(sqrt(price) ~ Fireplaces, data = ames) |>
  augment(newdata=data.frame(Fireplaces = 3), interval = "prediction")
```

Q8. 2nd explanatory variable

Using the same model from Q5, add a second categorical variable to the linear model plot using color (for both the points and the line). (We will create the full linear model next week, for now just make the plot with the two explanatory variables.) Answer the following question: are the different colored lines close to parallel?

```
praise()
```

```
[1] "You are fine!"
```