AE 13: Conditions for Multiple Linear Regression

Rail Trail

Driver: _____, Reporter: _____, Gopher: _____

> **!** Important
>
> - Open RStudio and create a subfolder in your AE folder called "AE-13".
> - Go to the Canvas and locate your `AE-13` assignment to get started.
> - Upload the `ae-13.qmd` and `rail_trail.csv` files into the folder you just created. The `.qmd` and PDF responses are due in Canvas. You can check the due date on the Canvas assignment.

Packages + data

```
library(tidyverse)
library(ggformula)
library(broom)
library(knitr)
library(mosaic)
library(mosaicData)

rail_trail <- read_csv("rail_trail.csv")
```

The data for this AE is based on data from the Pioneer Valley Planning Commission (PVPC) and is included in the **mosaicData** package. The PVPC collected data for ninety days from April 5, 2005 to November 15, 2005. Data collectors set up a laser sensor, with breaks in the laser beam recording when a rail-trail user passed the data collection station. More information can be found here.

Analysis goal

The goals of this activity are to:

- Determine whether the conditions for inference are satisfied in this multi-predictor setting.

Exercise 0

Fit a linear model to predict `volume` from ALL of the other predictors. The resulting model is called the "full model". Hint: if you use the formula `volume ~ .` in your `lm` command, it will automatically include all predictors. Once you have fit your model, use `tidy` to print it out. Have the reporter for you group write down the model on the white board. Does the model include any interaction terms?

Exercise 1

Augment the model you created above using the `augment` function. Generate a scatter plot of the residuals vs. the fitted values for this model.

Exercise 2

Make two plots:

1. Residuals vs. `precip`.
2. Residuals vs. `day_type`.

Exercise 3

Based on the three plots you've made and the four on the screen, do you think the linearity condition is satisfied?

Exerise 4

We check the constant variance assumption in the same way we do with SLR. Does the constant variance condition seems to be satisfied?

Exercise 5

Generate a histogram of the residuals. If you have time, also generate a QQ-plot of the residuals. Do you believe the normality condition is satisfied?

Exercise 6

How do you think would you go about checking the independence condition?

To submit the AE

> **!** Important
>
> - Render the document to produce the PDF with all of your work from today's class.
> - Upload your QMD and PDF files to the Canvas assignment.