

HW 12 – Math 58B

your name here

not due ever

Assignment Summary (Goals)

- Multiple Regression
 - choosing variables
 - confidence intervals for mean response
 - prediction intervals for individual response
 - R^2 and adjusted R^2

Note: this HW isn't ever due, so the solutions are already posted. Also, from Lab 12 and the class notes, there is a lot of R code. There is virtually no scaffolding of R code in this assignment as a way for you to test yourself with the R code. But feel free to go look at the solutions if you get stuck!

Q2. RailTrails Consider the RailTrail data available in the mosaic package.

```
# try this to see the variables: ?RailTrail
data("RailTrail")
```

- Build a linear model with `cloudcover` used to predict the `volume` of riders. For the process: make a scatter plot (include a regression line), run the linear model (find the coefficients), check to see if `cloudcover` is significant, and calculate the adjusted R^2 .
- Are the technical conditions met for the model?
- Find a model which predicts `volume` of ridership using some set of the explanatory variables which seem best to you. Use coefficient p-values as well as adjusted R^2 to make your determination.
- Using model from e., predict the `volume` of ridership for each of the 90 days (i.e., \hat{Y}). Create a plot of residuals vs predictions (add a horizontal line at 0). Create another plot of predictions (\hat{Y}) vs observed value (Y) (add a line with intercept 0 and slope 1: `geom_abline(intercept = 0, slope = 1)`)
- Did you notice that the residual plot seems to have increasing variability? let's try a log transformation on the response variable. In order to compare it to the observed value, we'll need to transform it back (using `exp()`). Create the linear model using `log(volume)` as the response variable and re-make the residual and prediction plots.
- Not on exam 2, might be considered to enhance the project** Which model has a higher adjusted R^2 , the model with `volume` as the response variable or the model with `log(volume)` as the response variable?
- Not on exam 2, might be considered to enhance the project** Find the CI and PI for the `log(volume)` of riders using the model with `avgtemp` and `cloudcover`. Unfortunately, the intervals need to be generated separately.

```
praise()
```

```
## [1] "You are kickass!"
```