

Lesson 18 - In Class Activity

Due End of Class - 20 Pts

In groups of 2-3 complete the following and turn in one PDF/Word document generated from RMarkdown.

Read the prompted for Four-Year Graduation Rates in Relation to Tuition and Type of School on Pgs 380-381 of the text.

Note that the data can be accessed from:

```
Graduation <- read.csv("https://raw.githubusercontent.com/nick3703/MA376/master/Graduation.csv")
Graduation <- Graduation %>% mutate(Type = as.factor(Type))
```

The research question is: Are the four-year graduation rates of four-year universities in California related to the annual tuition and/or whether the school is a private or public university, and is there an interaction between these two?

Answer the following questions:

1. Identify the observational and experimental units
2. Determine if this was an experiment or an observational study. Why does this question matter for our analysis?
3. Did the study use random sampling, random assignment, both, or neither? How is this related to the previous question?
4. Identify the response variable. Is it quantitative or categorical? Develop a visualization of the spread of the response variable.
5. SKIP
6. Identify the sources of explained variation and 2-3 sources of unexplained variation.
7. Find the overall mean, standard deviation, and SSTotal. Use these values to write out the fitted model for a **Single means model**. Include the SE of the residuals, σ^2 .

```
single.mean <- lm(GradRate~1,data=Graduation)
summary(single.mean)
```

```
##
## Call:
## lm(formula = GradRate ~ 1, data = Graduation)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -36.98  -21.73   -0.98   20.52   49.02
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  41.980      3.449   12.17  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.39 on 49 degrees of freedom
```

Mean is $\beta_0 = 41.98$, and $\sigma = 24.39$, for the model:

$$y_i = \beta_0 + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma)$$

8. Create a scatterplot comparing four year graduation rate and tuition. Describe the relationship you observe. Would you consider a linear model reasonable?
9. Write out, using proper notation, the linear regression model that would predict the four-year graduation rate from tuition. Using R, fit the model and give the fitted model as well. Give R^2 and the SE of the residuals from this model as well.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$i = \text{Subject}$$

$$\epsilon_i \sim N(0, \sigma)$$

10. In the context of this study what is the interpretation of the slope and the intercept.
11. Does this model explain more variation (statistically) than the single means model (null model)? How are you deciding?
12. Skip
13. Now, develop an appropriate model to compare School Type with tuition. What type of model will you use? Using proper notation write out the model. Using R, fit the model and record the fitted model as well.

```
our.lm<-lm(GradRate ~ Tuition+Type,data=Graduation)
summary(our.lm)
```

```
##
## Call:
## lm(formula = GradRate ~ Tuition + Type, data = Graduation)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -39.566 -14.548   0.829  11.312  38.383
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.699e-02  1.116e+01   0.006   0.9952
## Tuition     1.315e-03  2.874e-04   4.576 3.47e-05 ***
## TypePublic  1.856e+01  9.908e+00   1.873  0.0673 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.24 on 47 degrees of freedom
## Multiple R-squared:  0.4032, Adjusted R-squared:  0.3778
## F-statistic: 15.87 on 2 and 47 DF,  p-value: 5.403e-06
```

14. Do you prefer this model over the single means? Why or why not?
15. Write out and fit a statistical model using both School Type as well as Tuition to explain graduation rates. Record the fitted model, the R^2 , and the SE of the residuals.
16. Explain what an interaction would mean in context of this study
17. SKIP
18. Carry out a two-variable analysis including the interaction of tuition and type of school. Record the prediction equation, R^2 , and residual SE.
19. Write two equations, one to predict the four-year graduation rates from tuition for private schools, and one to predict the four-year graduation rates from tuition for public schools. How do the slopes and intercepts differ?
20. State the null/alternative hypothesis to test whether the observed interaction of type of school and tuition is statistically significant.
21. Do you believe there is convincing evidence of an interaction? How are you deciding? Ensure you check the validity conditions.
22. Assuming the validity conditions are true, predict the graduation rate for a private California university with an annual tuition of \$50,000
23. Two intervals, each at 95% confidence level, were created using the results from a regression analysis for type of school = private, and tuition = \$10,016: (1.6, 33.3) and (-21.5, 56.4). Which of these is a confidence interval and which of these is a prediction interval? How are you deciding?
24. Summarize your conclusions from the model as they relate to the original statistical question.