Week 1 Reading Guide

## Chapter 3: Linear Regression

1. **What does it mean when we say** and are *unbiased* estimates of and ?
2. **When is the standard error of the slope (**) the smallest? Why does this make sense?
3. **What is the residual standard error (RSE) an estimate of?**
4. **Would you always want to choose a model with the lowest RSE? Why or why not?**
5. **Is there a threshold for a “good”** ? Meaning, any $R^2 below this threshold would suggest there is no linear relationship between the variables?

Suppose you fit four different simple linear regressions modeling the relationship between a penguin’s body weight and (1) its flipper length, (2) its bill depth, (3) its bill length, and (4) its species. Then you fit a multiple linear regression where all four variables were included in the same model.

1. **Will the coefficients for each variable in the multiple linear regression be the same as the coefficient from their respective simple linear regression? If not, in what cases will they be the most different?**
2. **What is p-value for newspaper (in Table 3.4) conditional on?**
3. **When the sample size () is large, what F-statistics would result in you rejecting in an “overall F-test”**[[1]](#footnote-20)
4. **Why should we use an overall F-test instead of individual t-tests to decide if any of the variables in a multiple linear regression have a relationship with the response?**
5. **What does it mean when we say forward selection is a “greedy” approach to variable selection?**
6. **What model fit statistics could we use in a multiple linear regression?**
7. **Will residual standard error always decrease when a new variable is added to the model?**
8. **How is the slope () interpreted in a simple linear regression with a binary categorical variable?**
9. True or False: Changing the baseline group changes all of the estimated / predicted responses (means).
10. True or False: Changing the baseline group changes how the coefficients are interpreted.
11. True or False: Changing the baseline group changes the p-values for each coefficient.
12. **Why should not “significant” main effects be kept in an interaction model?**
13. **In a simple linear regression, when there is a non-linear relationship between the variables how would you expect the residual versus fitted plot to look?** Feel free to sketch a picture!
14. **What happens to the standard errors when the errors in a regression are correlated?**
15. **Suppose you plot the residuals versus fitted values for a regression you fit and you see evidence of a funnel pattern. What should you do?**
16. **Should you always remove “outliers” from your model?**
17. **In order for an “outlier” to heavily influence the slope of a regression line, where does it need to be located?** Feel free to sketch a picture!
18. **What happens to the standard errors when collinearity is present in a multiple regression?**
19. **When collinearity is present in a multiple regression, what type of error are we more likely to make when conducting hypothesis tests?**
20. **How would you remedy a regression that is found to have multicollinearity?**

1. [↑](#footnote-ref-20)