

# Final Study Guide

*Final - Tuesday March 19th 2pm - 3:30pm Weniger 287*

*2019-03-06*

## **Tuesday March 19th 2pm - 3:30pm Weniger 287**

### **The final is closed book.**

I will provide statistical tables if you need them (so you should know how to use them).

You should bring a calculator (although arithmetic errors are generally forgiven).

**All material on the midterm study guide** is also examinable.

### **You should be able to:**

- State the assumptions required for making inferences in regression.
- Rank the assumptions in rough order of importance.
- Describe the consequences of violating a particular assumption.
- Describe/sketch residual plots that should be examined to diagnose problems with the regression assumptions.
- Sketch a residual plot that illustrates a violation of a particular assumption (non-linearity, non-constant variance or non-normality).
- Given a residual plot, describe evidence you see for violations of the regression assumptions.
- Suggest a remedy for a particular violation.
- Describe three ways a point may be considered “unusual”.
- Name three case influence statistics and describe conceptually how they measure “unusualness”.
- Identify from a scatterplot (in the simple linear regression case) if a point is likely to be high leverage, influential and/or an outlier.
- Describe a limitation of case influence statistics.
- Describe what is meant by multicollinearity.
- Describe how multicollinearity might be detected.
- Discuss the consequences of multicollinearity.
- Describe the assumption that generalized least squares is designed to relax.
- Derive the generalized least squares estimates (for known  $\Sigma$ ).
- Give an example of data where using weighted least squares is desirable.
- Conduct a lack-of-fit test.
- Interpret the result of a lack-of-fit test.
- Describe the goal of robust regression techniques.
- Describe why we might transform the response and/or the explanatory variables.

- Choose a transform of the response based on a Box-Cox plot.
- Interpret a parameter estimate based on a regression with a log transformed response.
- State the additional assumption required to make inferences about medians in a regression using a log transformed response.
- Give a reason why variable selection might be recommended.
- Give a reason why variable selection might be avoided.
- Describe the process of model selection by forward selection/backward elimination.
- Give an advantage and a disadvantage of stepwise methods.
- Perform one step of forward selection/backward elimination.
- Name and describe four model selection criteria.
- Discuss the similarities and differences between model selection criteria.
- Discuss why it is dangerous to use the same data to fit a predictive model and to evaluate the model's predictive ability.
- Describe two regularized regression methods.
- Describe why might we prefer biased estimates (or predictions).
- Describe/sketch an example of a predictive model that would have low/high variance and low/high bias.
- Discuss the difference in goals between explanation and prediction.
- Describe the difference between linear and logistic regression models.
- Describe the difference between linear and non-linear regression models.