

STA 4210 Final Exam Topic List

What is the criterion for fitting a line to data pairs?

How do we form a confidence interval for the actual slope in the model or test a hypothesis about it?

How do we form a confidence interval for the average Y at some particular X-value? How do we form a prediction interval for a new Y at some particular X-value? What is the difference between these two objectives? Why is there more variability in the prediction problem?

How do we detect and how do we remedy the following assumption violations: 1 Nonlinearity of the regression function, 2 Nonconstancy of the error variance, 3 Outliers, and 4 Omission of Important Predictors.

What is the interpretation of the Total Sum of Squares, the Error Sum of Squares, the Regression Sum of Squares and the Coefficient of Determination?

What does the Bonferroni Inequality say and how do we construct simultaneous confidence intervals based on this inequality? How is the Working-Hotelling confidence band constructed? When would it be preferred to the Bonferroni method? How do both of these methods apply to prediction intervals?

How do we express the multiple linear regression model in matrix form?

How are polynomial models created? What are the effects of interaction terms and the purpose of higher order models?

How is the overall ANOVA test conducted and what is it testing?

How are inferences conducted for a) the slope parameters, b) the mean response at a particular set of predictor variables and c) a predicted response at a particular set of predictor variables?

What happens to the SSR and the coefficient of determination when we include more variables into the linear model? What is the purpose of the Adjusted R-square?

What are Extra Sums of Squares? What is their purpose and how are tests performed on these Extra SS?

What is a partial R-square and how is it interpreted?

What is multicollinearity? Which inferences are affected by its presence and which are not? What are the four ways to "fix" multicollinearity?

How should indicator predictor variables be used? What do they do structurally and how do we create separate categories when there are three or more categories for a single response?

How does the All Subsets regression work as a method to search for the best model? What are the criteria that we can use to determine the best subset of variables? What is an appropriate search strategy for a good model?

How do we identify points that are extreme in their Y value or in their X values? What do these quantities each measure? What are DFFITS, DFBETAS and Cook's D each measuring and what are they designed to identify?

What does the variance inflation factor measure and what is it designed to identify?

How does weighted least squares work and how do we use it to adjust inferences when changes in variability are present? How does Ridge Regression work and what is its purpose? How do we use weighted least squares to diminish the affects of influential data points?

What ingredients occur within a time series? What problems occur when we use our usual least squares regression to model a time series? How do we model and how do we detect the dependence usually seen in time series? How do we fit the autocorrelation model? How do we make forecasts based on this model? How do we include seasonal effects into the autocorrelation model?

Be able to read R output and construct appropriate models, estimators, confidence intervals, tests, and confidence and prediction intervals at a particular X set based on the output?

Know the appropriate steps to determine the best model?

Know the potential problems in model building, what diagnostics to apply in order to identify whether they are present and what remedial measures are available to fix them.

Be able to interpret appropriate graphs and computer output.