THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

CHE384, From Data to Decisions: Measurement, Uncertainty, Analysis, and Modeling

# Review Questions by Lecture (42-73)

Chris A. Mack
*Adjunct Associate Professor*

http://www.lithoguru.com/scientist/statistics/



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 42: Multiple Regression

- Define multiple regression, and how it applies to both bivariate and multivariate data
- What mathematical techniques are used for OLS multiple regression?
- Explain *interaction* and what it means for multiple regression
- Explain *multicollinearity* and what it means for multiple regression
- Provide three unique purposes for a model



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 43: Comparing Models

- Why can't $R^2$ be used to compare models with different number of parameters?
- Explain the adjusted $R^2$ and how it is used
- What is an "information criterion" and how is it used?
- The use of which information criterion results in the most parsimonious model?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 45: Best Subset Regression

- Understand the use of the partial F-test for subset models
- How is Mallow's $C_p$ used to evaluate subset models?
- How is the likelihood ratio used to evaluate subset models?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 47: Multicollinearity

- What is multicollinearity?
- What happens to OLS if two predictor variables have perfect correlation?
- What is the opposite of perfect correlation between predictor variables?
- What is a correlation matrix and how is it used?
- What happens to models and predictions when multicollinearity exists?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 48: Standardized Variables

- What is variable standardization, and why is it used?
- In general, will standardization help with problems of multicollinearity?

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 50: Detecting Multicollinearity

- What are the advantages and disadvantages of using the correlation matrix for detecting multicollinearity?
- How do the Variance Inflation Factors address the disadvantage of the correlation matrix?
- How do we use eigenvalues and the condition number to detect multicollinearity?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 51: Addressing Multicollinearity

- Name as many ways as you can for dealing with multicollinearity
- When can and can't you solve mulitcollinearity by collecting new data?
- When is it wise to drop a correlated predictor variable?
- Understand the basics of ridge regression



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 53: Principal Component Analysis

- Explain how constraints between regressor variables lead to multicollinearity
- Why is PCA sometimes described as a rotation of the parameter space?
- What does a small eigenvalue tell you about that principal component?
- How can PCA be used to improve regression?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 55: Robust Estimation

- Explain robustness and the breakdown point
- What are some common robust location estimators?
- What are some common robust scale estimators?
- What is the main disadvantage of using robust estimators?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 56: Robust Regression

- What is the breakdown point for OLS?
- Explain the basic operation of M-estimators for linear regression
- What are some of the difficulties and complications for robust regression?
- How do you choose among the robust regression options?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 58: Generalized Linear Modeling

- What are the three requirements of a generalized linear model?
- What are some common distribution/link function pairs?
- What is the distribution/link function pair for logistic regression?
- Name three examples of where you might want to use a logistic regression

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 59: Other Regression Topics

- What is an indicator variable and how is it used in regression?
- If I have a categorical variable with four levels, how many indicator variables are needed to represent it?
- What are the problems with non-linear regression?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 62: Building Models

- What are the three automated model search approaches?
- Explain omitted variable bias
- How can model building result in uncounted degrees of freedom?
- What are the three model validation approaches?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 64: Introduction to Design of Experiments

- Name the three types of inputs to a process
- Define design of experiments
- What are the three uses of experimental design?
- For regression to a straight line, what is the most efficient design?
- For regression to a straight line, what design produces uniform leverage?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 65: Regression Design

- Can you name all six principles of designing for regression?
- What is optimal design?
- What is the difference between repeats and replicates?
- What is randomization used for?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 67: Blocking in Experimental Design

- What is blocking?
- Why use blocking rather than randomization?
- What is covariate analysis and when is it used?
- Explain the randomized complete block
- Explain the Latin Square design



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 68: Factorial Design of Experiments

- What is full factorial design?
- What kind of design is often used for screening variables?
- What is the Hierarchy Principle?
- What is fractional factorial design and what are its advantages and disadvantages?
- Explain aliasing in fractional factorial design

THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 71: Response Surface Modeling

- When is response surface modeling preferred over two-level factorial design?
- What is wrong with varying our factors one-at-a-time?
- Describe some of the RSM designs we have discussed
- Why are center points often repeated in RSM?
- What are the four properties we look for in RSM?



THE UNIVERSITY OF TEXAS AT AUSTIN — WHAT STARTS HERE CHANGES THE WORLD

## Lecture 72: Final Thoughts on Design of Experiments

- Explain sequential DOE
- What is the simplex design and when is it used?
- Describe the Taguchi methods
- What are some of the cautions about using DOE and RSM?