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 (1-41)

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 1: The Knowledge Hierarchy

- Define "statistics" and "data analysis"
- What are some of the problems affecting data in the real world?
- What are some of the problems affecting data analysis in the real world?
- What are the four levels of the knowledge hierarchy?
- What are the five steps in the decision making process?

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

Lecture 2: Data and Measurement

- What are all the components of data?
- Explain the measurement model
- Explain the difference between precision and accuracy
- Explain the difference between repeatability and reproducibility
- Can you describe a measurement model for a measurement that we haven't discussed?

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

Lecture 3: Data Example

- What are the six parts of the concept "data"?
- Which two parts can sometimes (but not always) be taken for granted?
- How often do you see data presented where the remaining four parts are carefully documented?

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

Lecture 4: Process Modeling

- What are the three tasks in process modeling?
- Explain the relationship between model building and regression
- What are the two major outputs of regression?
- Define "linear regression"

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

Lecture 5: Regression Review, part 1

- What is a residual?
- What are the desired properties of a "best fit" parameter estimate?
- Explain MLE, maximum likelihood estimation.
- What are the assumptions of ordinary least squares (OLS) estimation?

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

Lecture 6: Regression Review, part 2

- Why is the matrix formulation of OLS commonly used?
- What is the hat matrix? Why is it important?
- Be able to use the results of this lecture to calculate standard errors and confidence intervals for model parameters and predicted model values

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

Lecture 8: Regression Review, part 2

- Name the six assumptions in OLS regression
- Define the mean square error (MSE) of a parameter estimate
- Describe the Anscombe graphs and what they teach us about regression

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

Lecture 10: What is the Distribution of the Residuals?

- Does the central limit theorem guarantee that error distributions are Normal?
- What are two common ways an actual distribution departs from Normality?
- What are the problems with using a histogram to assess distributions?

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

Lecture 11: Q-Q and Normal Probability Plots

- What are the steps for generating a normal probability plot?
- How does one interpret the shape of the normal probability plot?
- How can we test for normality using a normal probability plot?
- What other distributions can we test against a given data set?

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

Lecture 13: Testing for Skewness

- How is skewness defined?
- For positive skewness, what is the shape of the pdf?
- Be able to test a sample data set for skewness. What test statistic is used? What is its sampling distribution?

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

Lecture 14: Testing for Kurtosis

- How is kurtosis defined?
- For positive excess Kurtosis, what is the shape of the pdf?
- Be able to test a sample data set for excess kurtosis. What test statistic is used? What is its sampling distribution?

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

Lecture 16: Shapiro-Wilk Test for Normality

- Why do people like the Shapiro-Wilk test for normality?
- Be able to run the Shapiro-Wilk test in R
- What are the difficulties in interpreting the results of any hypothesis test for normality?

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

Lecture 17: Testing for Outliers, part 1

- What is an outlier?
- What is the difference between an outlier and a spurious data point?
- How does the Box and Whisker plot identify outliers?
- Be able to perform the Dixon Q test. What can go wrong with this test?

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

Lecture 18: Testing for Outliers, part 2

- What is a studentized outlier?
- What is the maximum possible value for a studentized outlier?
- Be able to perform the Grubbs' test, using the tables provided.

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

Lecture 19: Some Final Thoughts on Outliers

- Why should one focus on identifying the cause of an outlier?
- Name the four options for what to do with an outlier that can't be ignored
- What is an important alternative to outlier testing and rejection?
- Describe the recommended testing sequence for outliers

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

Lecture 21: Leverage in Regression

- Define leverage
- What is the role of the hat matrix in determining leverage?
- What is the difference between internally and externally studentized residuals?
- How should residuals be studentized for statistical testing?
- Know how to create and interpret the Williams Graph

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

Lecture 22: Influence in Regression

- Define influence
- Name several metrics of influence
- Explain what is meant by a "fragile" regression
- How does a measure of influence affect the way we approach outliers?

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

Lecture 24: Heteroscedasticity: When Variance Varies

- Define homoscedasticity and heteroscedasticity
- What are the consequences of heteroscedasticity to your regression?
- What are some of the causes of heteroscedasticity?
- What are the advantages of either the Bartlett test or the Brown-Forsythe test?

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

Lecture 26: Correcting for Heteroscedasticity

- Under what circumstances is it appropriate to take remedial action against heteroscedasticity?
- What are the two main approaches to addressing heteroscedasticity?
- When should you choose weighted regression vs. data transformation?

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

Lecture 28: Weighted Linear Regression

- Why would we ever want to do weighted regression?
- What is a weighted mean?
- How do the weights relate to the variance of each y value?
- How do we estimate weights?
- How does weighted regression affect our analysis of the residuals?

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

Lecture 30: Total Regression, part 1

- When do I have to worry about error in the x-variable?
- What is total regression (also called errors-in-variables regression)?
- Explain the effective variance approximation
- How does x uncertainty affects our OLS slope estimate for a straight-line model?
- When does error in the x-variable result in heteroscedasticity?

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

Lecture 31: Total Regression, part 2

- What is orthogonal regression and when is it useful?
- What is geometric mean regression and when is it useful?
- What do the method of moments and the asymptotic behavior of OLS (as $n \to \infty$) teach about the effect of x uncertainty on slope?

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

Lecture 32: Total Regression, part 3

- What is Deming regression and when is it useful?
- How does Deming regression relate to OLS, geometric mean regression, and orthogonal regression?
- How can we use replicates to estimate measurement uncertainty?
- When is it improper to use Deming regression when we now the uncertainty in x?

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

Lecture 34: The Wrong Model

- How can we use graphing to detect equation error?
- What is required to perform a chi-square test of the residuals to find equation error?
- What is the reduced chi-square and what is its expected value?
- What are the assumptions inherent in the chi-square test?

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

Lecture 35: The Wrong Model, Part 2

- What is the coefficient of determination and how is it calculated?
- What is the overall regression F-test and how is it used?
- What are the assumptions inherent in the F-test?
- What is model validation?
- Define model scope

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

Lecture 37: Independence of Residuals

- What can cause correlated (non-independent) residuals?
- Be able to generate and interpret a lag plot if the data sequence (order) is known
- Why is randomization of data order important in experimental design?
- What is a runs test and how is it performed?

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

Lecture 39: Autocorrelation in Time Series

- What is an AR(1) model?
- What is the Durbin-Watson test and what can it tell you?
- How can we estimate the correlation parameter ρ of an AR(1) model?
- How can we correct for autocorrelation when doing a regression?