Syllabus for MTH 214/ MTHH 501, Applied Statistics with Computers

Books: We use the book *Exploring Statistics: a modern introduction to data analysis and inference* Kitchens.

This is a *projected* course syllabus. Most topics will take approximately 1 week or 2 class meetings. Exceptions will likely be topics 8, 9 and 11. Topic 12 will be covered if time allows. Notice there are two scheduled exams and a comprehensive Final exam.

Topic 1: Collecting and Understanding Data.

- Lecture Material Topics to include discussion on data, collection of data, being skeptical of data collection, biased data sets.
- **Computer skills and assignment** Become familiar with the R software: storing data, applying a function.
- Media watch Bring in an article from a newspaper or website where data is collected and summarized. Critically discuss the collection in the language developed in class.

Topic 2: Univariate Data.

- Lecture Material Cover the basics of observational data: types of data, stem-and-leaf plots, measures of center and measures of spread.
- **Computer skills and assignment** Learn graphical means to view univariate data: histograms, frequency polygons, pie charts and boxplots. Discuss basic vocabulary to describe a distribution.
- Media watch Bring in an example from a newspaper or a website where a graphical presentation of univariate data is made. Comment on the strengths and weaknesses of the graphic.

Topic 3: Bivariate Data

- Lecture Material Cover contingency tables, correlation (Pearson and Spearman) and least squares regression.
- **Computer skills and assignment** Table manipulation, side-by-side boxplots, segmented bar graphs, scatterplots, fitted-line plots, brushing, subsetting data, outliers.
- Media watch Bring in examples of bivariate data. This can be graphically presented or in tabular format. Comment on the relationship implied by the data or the description thereof.

Topic 4: Probability and Distributions

- Lecture Material Cover basic concepts: definitions, intuitive examples, independence, conditional probability, Bayes theorem, the binomial and normal distributions.
- **Computer skills and assignment** Simulate binomial data, normally distributed data and other types of continuous data: exponential, uniform etc. Learn the normal plot (qqplot of data against normal quantiles) for comparison of data with normally distributed data. Z-scores.
- Media watch Find an article where an inference is made based on a data set. Describe the population and the sample and comment on the inference made.

Topic 5: Sampling Distributions

- Lecture Material Introduction to the sampling distribution of a statistic. In particular, the sampling distribution of the average or proportion as appropriate. Discussion of the Central limit theorem.
- **Computer skills and assignment** Learn to simulate data to investigate the sampling distributions mentioned in lecture.
- Media watch Find an example in the media where the distribution of a random variable is discussed.
- Test 1 Review and Exam.

Topic 6: Describing Distributions

- Lecture Material Work through a few case studies using a combination of graphical analysis and appeal to normally distributed random variables. Discuss appropriate measures of central tendency for the data set. Describe a framework for exploratory data analysis with a mathematical perspective.
- **Computer skills and assignment** Practice reading normal plots of real data, characterize data sets as symmetric or skewed, long of short tailed as appropriate, investigate biased versus unbiased statistics, compare variances of unbiased statistics (mean and median).
- Media watch Find an article that presents data and a summary. Comment on the steps made in the analysis of the data.

Topic 7: Confidence Interval Estimation

- **Lecture Material** Estimation of μ , the population mean and an introduction to the z-interval, the *t*-interval. Confidence interval for the median.
- **Computer skills and assignment** Use of R to find the confidence intervals. Simulation of polling to understand the confidence interval.
- Media watch Find an example of a poll. Identify the statistic, the margin of error and describe what is meant by confidence in this number.

Topic 8: Hypothesis Testing

- Lecture Material Introduction to hypothesis testing and application to tests for proportion, mean and median.
- Computer skills and assignment Use of R to find the *p*-value for a statistic.

Media watch Find a journal article where hypothesis testing or *p*-values are used. Comment.

Topic 9: Inferences about Differences

Lecture Material two-sample test of proportion, two sample T-test, Pooled T-test, Rank-sum test.

Computer skills and assignment R implementation of the tests, applications to real world data sets.

Media watch Find an illustration in the media of a two-sample test. Comment.

Topic 10: Chi-Square tests

- Lecture Material Analysis of categorical data: the χ -square goodness of fit and test for independence.
- **Computer skills and assignment** Use R to perform Chi-Square tests, use computer simulation to understand the test.

Media watch Find an example of categorical bivariate data, and compute an appropriate test.

<u>Test 2</u> Review and Exam. Topic 11: Regression Analysis

- Lecture Material The linear regression model, inference about the coefficients, estimation and prediction.
- **Computer skills and assignment** Linear regression in R, residual plots, predicting values, multiple linear regression and quadratic regression models.

Media watch Find an example of curve fitting to real-world data.

Topic 12: Analysis of Variance

Lecture Material Comparison of several means, non-parametric Kruskal-Wallis test.

Computer skills and assignment R implementation of these ideas.

Media watch Find an example of ANOVA in the media or in a journal article. Discuss.