Math 216 Final Exam Study Guide

The first midterm will cover Chapters 1 through 7 in your textbook, as well as additional data visualization techniques discussed in class. You are encouraged to bring a calculator (scientific or graphing) to the test, but you will not be allowed to use a laptop during the test. You may bring two 3"x5" index cards with notes (front and back) for use during the test.

You should be able to do each of the tasks listed below and understand the concepts associated with each task. See the end of this document for a list of suggested practice problems from the textbook.

Section 1.2 – Data Basics

- Determine whether a given variable is numerical or categorical.
- Determine whether a given numerical variable is discrete or continuous.
- Determine from a scatterplot if two variables are positively associated, negatively associated, or not associated.

Section 1.3 – Numerical Data

- Calculate the mean and standard deviation of a small data set.
- Calculate the first quartile, median, and third quartile of a data set.
- Identify outliers in a data set using the interquartile range method.
- Determine from a histogram if a data set is left skewed, right skewed, or symmetric.
- Determine from a histogram if the mean of a data set is likely less than, equal to, or greater than the median.
- Compare multiple data sets using histograms.
- Match a data set's histogram with its boxplot.
- Classify a summary statistic (mean, median, etc.) as robust or not.
- Compare data sets using summary statistics (mean, standard deviation, etc.).

Section 1.4 – Categorical Data

- Interpret a bar, segmented bar, or mosaic plot.
- Discuss why pie charts are inferior. (Just kidding.)
- Compare multiple data sets using boxplots.

Section 1.5 – Data Collection Principles

- Identify more or less effective sampling techniques.
- Discuss potential problems with particular sampling techniques.
- Identify potential lurking variables in observational studies.

Section 1.7 - Experiments

- Discuss how a given experiment does or does not fulfill the principles of controlling, randomization, replication, and blocking.

Section 2.1 – Defining Probability

- Identify common misconceptions about probability (base rate fallacy, representativeness heuristic, conjunction fallacy) in examples of probability estimations.
- Discuss ways in which an example could be thought of as following the empirical or subjective notions of probability.
- Use a Venn diagram, along with the General Addition and Complement Rules, to calculate probabilities.
- Translate a narrative example of probabilities to event/set notation, and vice versa.
- Determine if two events described as unions, intersections, or complements of other events are equal.
- Shade a Venn diagram to represent a particular event described as unions, intersections, or complements of other events.
- Distinguish between the common notion of independence (that one event does not cause the other) and the mathematical notion.
- Calculate probabilities involving independent events using the Product Rule.

Section 2.2 – Continuous Distributions

- Use a simple (continuous or discrete) probability density function to calculate probabilities.

Section 2.3 – Conditional Probability

- Calculate conditional probabilities given sufficient information about an example.
- Calculate probabilities using conditional probabilities and/or the General Multiplication Rule.
- Use conditional probabilities to determine if two events are independent.
- Model a probabilities problem using a tree diagram and use that tree diagram to calculate probabilities. (Optionally, use Bayes' Theorem to solve these kinds of problems.)

Section 2.4 – Sampling from a Small Population

- Calculate probabilities involving sampling with or without replacement from small populations.
- To discuss the relationship among independence, sampling without replacement, and sample size.

Section 2.5 – Random Variables

- Determine the probability distribution for a given random variable given sufficient information on that random variable.
- Find the expected value, variance, and standard deviation for a random variable given its probability distribution.

- Find the expected value, variance, and standard deviation for a linear combination of random variables given the expected value and variance of the component random variables.

Section 3.1 – Normal Distributions

- To analyze the graph of a normal distribution, particularly how means, standard deviations, and z-scores are represented
- To compute probabilities for a random variable with a standard or non-standard normal distribution using a normal probability table

Section 3.2 – Probability Plots

- To construct and interpret a standard normal probability plot for a small data set
- To interpret a standard normal probability plot for a large data set

Section 3.4 – Binomial Distributions

- To determine whether a given experiment is a binomial experiment
- To model a given scenario using an appropriate binomial random variable
- To compute probabilities using binomial probability distributions
- To compute the expected value, variance, and standard deviation of a binomial random variable

Section 3.5.2 – Poisson Distributions

- To model a given scenario using an appropriate Poisson random variable
- To compute probabilities using Poisson probability distributions
- To compute the expected value, variance, and standard deviation of a Poisson random variable

Section 4.1 & 4.4 – The Central Limit Theorem

- To discuss the conditions under which the Central Limit Theorem applies
- To determine the approximate distribution of a sample mean (including its mean and standard deviation) given the mean and standard deviation of the underlying distribution, assuming that the Central Limit Theorem applies
- To compute probabilities for a sample mean given the mean and standard deviation of the underlying distribution, assuming the Central Limit Theorem applies and using a normal probability table

Section 4.2 – Confidence Intervals

- To describe (accurately!) the meaning of an X% confidence interval
- To discuss conditions under which an X% confidence interval can be constructed from a sample mean
- To construct the X% confidence interval for the mean of a population given a sample mean and either the sample or population standard deviation

- To determine the sample size necessary to construct a X% confidence interval of a specified width for the mean of a population given the population standard deviation
- To discuss the factors that determine the width of a confidence interval (e.g. confidence level, sample size, and population or sample standard deviation)

Section 4.3 – Hypothesis Tests

- To determine appropriate null and alternative hypotheses for a given hypothesis test
- To describe the Type I and II errors associated with a given hypothesis test and to evaluate the relative seriousness of these errors
- In the context of a one-sided or two-sided hypothesis test for a population mean when the population standard deviation is known or the sample size is large enough to permit use of the sample standard deviation…
    - To determine whether to reject or accept the null hypothesis at a specified significance level by using an appropriate confidence interval
    - To compute the p-value for a specified hypothesis test and set of data
    - To determine whether to reject or accept the null hypothesis at a specified significance level using the p-value for a set of data
    - To discuss the relationships among the population or sample standard deviation, the sample size, the p-value for a set of data, and the difference between the actual and assumed population means.

Section 5.1 – Paired Data

- To determine if two sets of observations can be considered paired data
- To construct the X% confidence interval for the difference in population means for two populations based on paired samples from those populations
- To conduct a one-side or two-sided hypothesis test for the difference in population means for two populations based on paired samples from those populations

Section 5.2 – Difference of Two Means

- To construct the X% confidence interval for the difference in population means for two populations based on unpaired samples from those populations
- To conduct a one-side or two-sided hypothesis test for the difference in population means for two populations based on unpaired samples from those populations
- To justify the standard error for the difference of two means using results from section 2.5 on linear combinations of random variables

Section 5.3 – Single Population Proportion

- To construct the X% confidence interval for a population proportion given a sample proportion
- To determine the margin of error of a poll given its confidence level and sample size

- To determine the sample size needed for a poll with a given confidence level in order to have a given margin of error
- To conduct a one-sided or two-sided hypothesis test for a population proportion based on a sample from that population

Section 5.4 – Difference of Two Proportions

- To construct the X% confidence interval for the difference in population proportions given samples from those populations
- To conduct a one-side or two-sided hypothesis test for the difference in population proportions given samples from those populations (including under the special condition that H_0: $p_1 = p_2$)

Section 6.1 – Small Sample Inference for the Mean

- To discuss the assumptions that must be satisfied in order to use a *t*-distribution in a confidence interval or hypothesis test
- To assess the normality of a set of data through probability plots, histograms, and rules of thumb.
- To construct the X% confidence interval for a population mean given a small sample from that population (assuming the population is normally distributed)
- To conduct a one-side or two-sided hypothesis test for the population mean given a small sample from that population (assuming the population is normally distributed)

Section 6.2 – Small Sample Inference for the Difference of Two Means

- To construct the X% confidence interval for the difference in population means for two populations based on small samples from those populations (assuming those populations are normally distributed)
- To conduct a one-side or two-sided hypothesis test for the difference in population means for two populations based on small samples from those populations (assuming those populations are normally distributed)

Section 7.1 – Residuals and Correlation

- To discuss possible sources of random error in a linear model in the context of particular variables, including sampling variability
- To match a scatterplot with a possible linear model with the associated residual plot
- To describe the nature and strength of the correlation between two variables in a data set given the correlation coefficient for that data set

Section 7.2 – Least Squares Regression

- To apply the least squares criterion to sample data to select lines of best fit
- To discuss conditions under which least squares regression is appropriate
- To identify the least squares line for a set of data from summary statistics

- To describe the nature of the correlation between two variables in a data set given the line of best fit to that data
- To interpret the meaning of the slope of a least squares line in the context of particular variables
- To use a least squares line to make predictions and discuss conditions under which those predictions are less valid (e.g. extrapolation)
- To interpret R^2 for a set of data in the context of particular variables

Section 7.4 – Inference for Linear Regression

- To construct a X% confidence interval for the slope of a line of best fit to a set of data given a point estimate of the slope and the standard error for that estimate
- To conduct a one-side or two-sided hypothesis test for the slope of a line of best fit to a set of data given a point estimate of the slope and the standard error for that estimate

Data Visualization

- Interpret bubble charts, heatmaps, and treemaps.
- Evaluate the quality of an infographic, particularly its use of spatial relationships and color to convey meaning and its clarity on the sources of its data.

End-of-Chapter Exercises

- Section 1.2 – 1.3, 1.5, 1.9
- Section 1.3 – 1.13, 1.15, 1.19, 1.23, 1.25, 1.27, 1.29, 1.31
- Section 1.4 – 1.35
- Section 1.5 – 1.37
- Section 1.6 – 1.39, 1.41, 1.45
- Section 1.7 – 1.49
- Section 2.1 – 2.1, 2.3, 2.5, 2.7, 2.9
- Section 2.2 – 2.15
- Section 2.3 – 2.19, 2.21, 2.23, 2.27, 2.29
- Section 2.4 – 2.31, 2.33
- Section 2.5 – 2.39, 2.41, 2.45, 2.47
- Section 3.1 – 3.1, 3.3, 3.5, 3.7, 3.9, 3.11, 3.13
- Section 3.2 – 3.19
- Section 3.4 – 3.27, 3.29, 3.31, 3.33
- Section 3.5 – 3.43, 3.45
- Section 4.1 – 4.3, 4.5, 4.7
- Section 4.2 – 4.11, 4.13, 4.15, 4.17, 4.19, 4.21
- Section 4.3 – 4.23, 4.27, 4.29, 4.31, 4.35
- Section 4.4 – 4.37

- Section 5.1 – 5.1
- Section 5.2 – 5.3, 5.5, 5.7
- Section 5.3 – 5.9, 5.13, 5.15, 5.17, 5.21, 5.29
- Section 5.4 – 5.31, 5.33, 5.37
- Section 6.1 – 6.1, 6.5, 6.9, 6.11
- Section 6.2 – 6.13, 6.17, 6.19, 6.21
- Section 7.1 – 7.1, 7.3, 7.9, 7.15
- Section 7.2 – 7.19, 7.21, 7.25
- Section 7.4 – 7.31, 7.33, 7.35, 7.37(d)