Department of Mathematical Sciences
Appalachian State University

December 6, 2007

STT2810 Course Objectives

Part I: Exploring and Understanding Data

Chapter 1: Stats Starts Here

Think first. Know where you're headed and why. It will save you a lot of work.

Show is about the mechanics of calculating statistics and graphical displays, which are important (but are not the most important part of Statistics).

Tell what you've learned. You must explain your results so that someone else can understand your conclusions.

Chapter 2: Data

- Be able to identify the *Who*, *What*, *When*, *Where*, *Why*, and *How* of data, or recognize when some of this information has not been provided.
- Be able to identify the cases and variables in any data set.
- Be able to classify a variable as categorical or quantitative depending on its use.
- For any quantitative variable be able to identify the units in which the variable has been measured (or note that they have not been provided).

- Be able to describe a variable in terms of its *Who*, *What*, *When*, *Where*, *Why*, and *How* (and be prepared to remark when that information is not provided).

Chapter 3: Displaying and Describing Categorical Data

- Be able to recognize when a variable is categorical and choose an appropriate display for it.
- Understand how to examine the association between categorical variables by comparing conditional and marginal percentages.

- Be able to summarize the distribution of a categorical variable with a frequency table.
- Be able to display the distribution of a categorical variable with a bar chart or pie chart.

- Know how to make and examine a contingency table.
- Know how to make and examine displays of the conditional distributions of one variable for two or more groups.

Tell

- Be able to describe the distribution of a categorical variable in terms of its possible values and relative frequencies.
- Know how to describe any anomalies or extraordinary features revealed by the display of a variable.
- Be able to describe and discuss patterns found in a contingency table and associated displays of conditional distributions.

Chapter 4: Displaying and Summarizing Quantitative Data

Think

- Be able to identify an appropriate display for any quantitative variable.
- Be able to guess the shape of the distribution of a variable by knowing something about the data.
- Be able to select a suitable measure of center and a suitable measure of spread for a variable based on information about its distribution.
- Know what the median is: it is the middle value of a distribution or data set.
- Know the basic properties of the mean: The mean is the point at which the distribution or data set balances.
- Know that the standard deviation summarizes how spread out all the data are around the mean.
- Understand that the median and IQR resist the effects of outliers, while the mean and standard deviation do not.
- Understand that in a skewed distribution, the mean is pulled in the direction of the skewness (toward the longer tail) relative to the median.

Show

- Know how to display the distribution of a quantitative variable with an appropriate graph.
- Know how to compute the mean and median of a set of data.
- Know how to compute the standard deviation and IQR of a set of data.

Tell

- Be able to describe the distribution of a quantitative variable in terms of its shape, center, and spread.
- Be able to describe any anomalies or extraordinary features revealed by the display of a variable.
- Know how to describe summary measures in a sentence. In particular, know that the common measures of center and spread have the same units as the variable that they summarize, and should be described in those units.

- Be able to describe the distribution of a quantitative variable with a description of the shape of the distribution, a numerical measure of center, and a numerical measure of spread. Be sure to note any unusual features, such as outliers, too.

Chapter 5: Understanding and Comparing Distributions

Think

- Be able to select a suitable display for comparing groups. Understand that histograms show distributions well, but are difficult to use when comparing more than two or three groups. Boxplots are more effective for comparing several groups, in part because they show much less information about the distribution of each group.
- Understand that how you group data can affect what kinds of patterns and relationships you are likely to see. Know how to select groupings to show the information that is important for your analysis.
- Be aware of the effects of skewness and outliers on measures of center and spread. Know how to select appropriate measures for comparing groups based on their displayed distributions.
- Understand that outliers can emerge at different groupings of data, and that whatever their source, they deserve special attention.

Show

- Know how to make side-by-side histograms on comparable scales to compare the distributions of two groups.
- Know how to make side-by-side boxplots to compare the distributions of two or more groups.
- Know how to describe differences among groups in terms of patterns and changes in their center, spread, shape, and unusual values.

Tell

- Know how to compare the distributions of two or more groups by comparing their shapes, centers, and spreads. Be prepared to explain your choice of measures of center and spread for comparing the groups.
- Be able to describe trends and patterns in the centers and spreads of groups especially if there is a natural order to the groups, such as a time order.
- Be prepared to discuss patterns in a time plot in terms of both the general trend of the data and the changes in how spread out the pattern is. Be able to use a smooth trace as a general guide to the long-term patterns.
- Be able to describe the distribution of a quantitative variable in terms of its shape, center, and spread.
- Be able to describe any anomalies or extraordinary features revealed by the display of a variable.
- Know how to compare the distributions of two or more groups by comparing their shapes, centers, and spreads.
- Know how to describe patterns over time shown in a timeplot.

- Be able to discuss any outliers in the data, noting how they deviate from the overall pattern of the data.

Chapter 6: The Standard Deviation as a Ruler and a Normal Model

Think

- Understand how adding (subtracting) a constant or multiplying (dividing) by a constant changes the center and/or spread of a variable.
- Recognize when standardization can be used to compare values.
- Understand that standardizing uses the standard deviation as a ruler.
- Recognize when a Normal model is appropriate.

Show

- Know how to calculate the z-score of an observation.
- Know how to compare values of two different variables using their z-scores.
- Be able to use Normal models and the 68-95-99.7 Rule to estimate the percentage of observations falling within 1, 2, or 3 standard deviations of the mean.
- Know how to find the percentage of observations falling below any value in a Normal model using a Normal table or appropriate technology.
- Know how to check whether a variable satisfies the **Nearly Normal Condition** by making a Normal probability plot or a histogram.

Tell

- Know what z-scores mean.
- Be able to explain how extraordinary a standardized value may be by using a Normal model.

Part II: Exploring Relationships Between Variables

Chapter 7: Scatterplots, Association, and Correlation

Think

- Know that a scatterplot can help identify possible relationships between two quantitative variables.
- Know how to identify the roles of the variables and to place the response variable on the y-axis and the explanatory variable on the x-axis.
- Know the conditions for correlation and how to check them.
- Know that correlations are between -1 and + 1, and that each extreme indicates a perfect linear association.
- Understand how the magnitude of the correlation reflects the strength of a linear association as viewed in a scatterplot.
- Know that the correlation has no units.

- Know that the correlation coefficient is not changed by changing the center or scale of either variable.
- Understand that causation cannot be demonstrated by a scatterplot or correlation.

Show
- Know how to make a scatterplot by hand (for a small set of data) or with technology.
- Know how to compute the correlation of two variables.
- Know how to read a correlation table produced by a statistics program.

Tell
- Be able to describe the direction, form, and strength of a scatterplot.
- Be prepared to identify and describe points that deviate from the overall pattern.
- Be able to use correlation as part of the description of a scatterplot.
- Be alert to misinterpretations of correlation.
- Understand that finding a correlation between two variables does not indicate a causal relationship between them. Beware the dangers of suggesting causal relationships when describing correlations.

Chapter 8: Linear Regression

Think
- Be able to identify response (y) and explanatory (x) variables in context.
- Understand how a linear equation summarizes the relationship between two variables.
- Recognize when a regression should be used to summarize a linear relationship between two quantitative variables.
- Be able to judge whether the slope of a regression makes sense.
- Know how to examine your data for violations of the **Straight Enough Condition** that would make it inappropriate to compute a regression.
- Understand that the least squares slope is easily affected by extreme values.
- Know that residuals are the differences between the data values and the corresponding values predicted by the line and that the least squares criterion finds the line that minimizes the sum of the squared residuals.

Show
- Know how to find a regression equation from the summary statistics for each variable and the correlation between the variables.
- Know how to find a regression equation using your statistics software and how to find the slope and intercept values in the regression output table.
- Know how to use regression to predict a value of y for a given x.
- Know how to compute the residual for each data value and how to display them.

Tell
- Be able to write a sentence explaining what a linear equation says about the relationship between y and x, basing it on the fact that the slope is given in y-units per x-unit.

- Understand how the correlation coefficient and the regression slope are related. Know how R^2 describes how much of the variation in y is accounted for by its linear relationship with x.
- Be able to describe a prediction made from a regression equation, relating the predicted value to the specified x-value.

Part III: Gathering Data

Chapter 11: Understanding Randomness

Think
- Be able to recognize random outcomes in a real-world situation.
- Be able to recognize when a simulation might usefully model random behavior in the real world.

Show
- Know how to perform a simulation either by generating random numbers on a computer or calculator, or by using some other source of random values such as dice, a spinner or a table of random numbers.

Tell
- Be able to describe a simulation so that others could repeat it.
- Be able to discuss the results of a simulation study and draw conclusions about the question being investigated.

Chapter 12: Sample Surveys

Think
- Know the basic concepts and terminology of sampling (see the preceding list).
- Recognize population parameters in descriptions of populations and samples.
- Understand the value of randomization as a defense against bias.
- Understand the value of sampling to estimate population parameters from statistics calculated on representative samples drawn from the population.
- Understand that the size of the sample (not the fraction of the population) determines the precision of estimates.

Show
- Know how to draw a simple random sample from a master list of a population, using a computer or a table of random numbers.

Tell
- Know what to report about a sample as part of your account of a statistical analysis.
- Report possible sources of bias in sampling methods. Recognize voluntary response and nonresponse as sources of bias in a sample survey.

Part IV: Randomness and Probability

Chapter 14: From Randomness to Probability

Think

- Understand that random phenomena are unpredictable in the short term but show run regularity.
- Be able to recognize random outcomes in a real-world situation.
- Know that the relative frequency of an event of a random phenomenon settles down a value called the (empirical) probability. Know that this is guaranteed for independent events by the Law of Large Numbers.
- Know the basic definitions and rules of probability.
- Recognize when events are disjoint and when events are independent. Understand the difference, and that disjoint events cannot be independent.

Show

- Be able to use the facts about probability to determine whether an assignment of probabilities is legitimate. Each probability must be a number between 0 and 1 and the sum of the probabilities assigned to all possible outcomes must be 1.
- Know how and when to apply the Addition Rule. Know that events must be disjoint for the Addition Rule to apply.
- Know how and when to apply the Multiplication Rule. Know that events must be independent for the Multiplication Rule to apply. Be able to use the Multiplication Rule to find probabilities for combinations of independent events.
- Know how to use the Complement Rule to make calculating probabilities simpler. Recognize that probabilities of "at least ..." are likely to be simplified in this way.

Tell

- Be able to use statements about probability in describing a random phenomenon. You will need this skill soon for making statements about statistical inference.
- Know and be able to use correctly the terms *sample space*, *disjoint events*, and *independent events*.

Part V: From the Data at Hand to the World at Large

Chapter 18: Sampling Distribution Models

Think

- Understand that the variability of a statistic (as measured by the standard deviation of its sampling distribution) depends on the size of the sample. Statistics based on larger samples are less variable.
- Understand that the Central Limit Theorem gives the sampling distribution model of the mean for sufficiently large samples regardless of the underlying population.

Show

- Be able to demonstrate a sampling distribution by simulation.
- Be able to use a sampling distribution model to make simple statements about the distribution of a proportion or mean under repeated sampling.

Tell

- Be able to interpret a sampling distribution model as describing the values taken by a statistic in all possible realizations of a sample or randomized experiment under the same conditions.

Chapter 19: Confidence Intervals for Proportions

Think

- Understand confidence intervals as a balance between the precision and the certain of a statement about a model parameter.
- Understand that the margin of error of a confidence interval for a proportion changes with the sample size and the level of confidence.
- Know how to examine your data for violations of conditions that would make inference about a population proportion unwise or invalid.

Show

- Be able to construct a one-proportion z-interval.

Tell

- Be able to interpret a one-proportion z-interval in a simple sentence or two. Write such an interpretation so that it does not state or suggest that the parameter of interest is itself random, but rather that the bounds of the confidence interval are the random quantities about which we state our degree of confidence.

Chapter 20: Testing Hypotheses about Proportions

Think

- Be able to state the null and alternative hypotheses for a one-proportion z-test.
- Know the conditions that must be true for a one-proportion z-test to be appropriate and how to examine your data for violations of those conditions.
- Be able to identify and use the alternative hypothesis when testing hypotheses. Understand how to choose between a one-sided and two-sided alternative hypothesis and be able to explain your choice.

Show

- Be able to perform a one-proportion z-test.

Tell

- Be able to write a sentence interpreting the results of a one-proportion z-test.
- Know how to interpret the meaning of a P-value in nontechnical language, making clear that the probability claim is made about computed values under the assumption that the null model is true and not about the population parameter of interest.

Part VI: Learning about the World

Chapter 23: Inferences about Means

- Know the assumptions required for t-tests and t-based confidence intervals.
- Know how to examine your data for violations of conditions that would make inference about the population mean unwise or invalid.
- Understand that a confidence interval and a hypothesis test are essentially equivalent. You can do a two-tailed hypothesis test at level of significance α with a $1 - \alpha$ confidence interval, or a one-tailed test with a 1 - 2α confidence interval.

- Be able to compute and interpret a t-test for the population mean using a statistics package or working from summary statistics for a sample.
- Be able to compute and interpret a f-based confidence interval for the population mean using a statistics package or working from summary statistics for a sample.

- Be able to explain the meaning of a confidence interval for a population mean. Make clear that the randomness associated with the confidence level is a statement about the interval bounds and not about the population parameter value.
- Understand that a 95% confidence interval does not trap 95% of the sample values.
- Be able to interpret the result of a test of a hypothesis about a population mean.
- Know that we do not "accept" a null hypothesis if we cannot reject it. We say that we fail to reject it. Understand that the P-value of a test does not give the probability that the null hypothesis is correct.