Statistics 311 Learning Objectives

This document contains a draft of the learning objectives for Introductory Statistics at North Carolina State University

Authored by Roger Woodard, Pam Arroway, Jen Gratton, Steve Stanislav and Jeff Thompson

Data Collection and Surveys:

A1. Given a study, identify population, sample, parameter, and statistic.
A2. Given a study, determine whether it is an observational study or an experiment.
A3. Given a survey sample, determine whether the sample is a simple random sample, a voluntary response sample, a convenience sample, or has other forms of sampling bias.
A4. Given a study, recognize typical forms of biases such as potential undercoverage, nonresponse, question wording, and response bias.
A5. Given a study, describe the potential problem of selecting a sample with severe undercoverage, nonresponse, and response bias.
A6. Given a study, decide if undercoverage, nonresponse, and response bias are likely to cause significant problems in inference.
A7. Given a study, determine whether a SRS or a stratified random sample was selected.
A8. Given a study's objective, explain the advantage of using a stratified random sample over a SRS.
A9. Given a study's objective, decide when to use stratified random sampling or SRS.

Summarizing with Graphics:

B1. Given a set of raw data, identify the individuals and the variables.
B2. Given a variable, determine whether it is categorical or quantitative.
B3. List which graphical methods (pie charts, histograms, etc.) are appropriate for categorical and for quantitative variables.
B4. Given a set of categorical data, create bar chars and pie graphs.
B5. Given a bar chart or pie chart, determine the number of individuals or percentage of individuals for each category.
B6. Given a histogram, determine the number of individuals in a particular range.
B7. Given a set of raw data, create a histogram or stemplot by hand.
B8. Given a histogram or stemplot, describe the distribution's shape (skewed left, skewed right, symmetric, or multimodal), center, and spread.
B9. Given a histogram, stemplot or time plot, identify values that would be considered outliers.
B10. Given a time plot, identify trends and changes in trends.
B11. Given a graphical summary, propose an explanation of the distribution of the data.
B12. Given a description of a variable, predict what shape the histogram of that variable would take.

Summarizing with Numbers:

C1. Given a set of raw data, calculate the principle summary statistics (mean, median, quartiles, inter-quartile range, variance, standard deviation) by hand and using appropriate software.
C2. Explain how the mean and median are related for different shapes of a distribution (skewed left, skewed right or symmetric).
C3. List the following characteristics of the standard deviation
  a. The standard deviation must be greater than or equal to zero.
  b. When standard deviation is equal to zero, there is no spread – every number on the list if the same.
C4. Describe how adding or subtracting the same value to every observation or multiplying or dividing every observation by the same value will change the median, mean and standard deviation.
C5. Given a set of summary statistics (mean, median and standard deviation), find the summary statistics of a data set that would result from a linear transformation of the original data. (A linear transformation means adding or subtracting the same value from each observation and/or multiplying or dividing each observation by the same value).
C6. Given a histogram, be able to determine the approximate location of the median and quartiles.
C7. Match given histograms to given sets of appropriate summary statistics. (For example, mean, median, standard deviation and quartiles).
C8. Explain the impact of outliers on summary statistics such as mean, median and standard deviation.
C9. Given a set of raw data or Five Number Summary, create a boxplot.
C10. Given a boxplot, determine the shape of the distribution (skewed right, skewed left or symmetric).
C11. Given side-by-side boxplots, contrast key features of the groups represented by the boxplots.

Basic Probability:

D1. Explain what the long-run frequency interpretation of probability is.
D2. Given an experiment, describe the sample space.
D3. Given an event, list the outcomes that make up the event.
D4. Given a set of probabilities, determine if they are legitimate. That is, check the following criteria
  a. The probability of any event is a number between zero and one.
  b. If we assign a probability to every possible outcome of a random phenomenon, the sum of these probabilities must be one.
D5. Given a probability distribution, find a specified probability for an event of interest.
D6. Find the probability of an event by summing the probabilities of the individual outcomes that make up the event.
D7. Find the probability that an event does not occur by one minus the probability that the event does occur.

The Normal Distribution:

E1. Given a histogram, explain how it relates to the conceptual density curve for the variable in the corresponding population.
E2. List the key characteristics of the normal distribution.
E3. Given a mean and standard deviation, use the 68-95-99.7 rule to find the percentage of the normal distribution within one, two, or three standard deviations of the mean.
E4. Given a mean, standard deviation, and observed value, calculate the standardized value (z-score).
E5. Given a z-score, use a normal table to find the corresponding probability.
E6. Given a mean and standard deviation, find a specified percentile of the normal distribution.

Sampling Distributions:

F1. Given a study, describe how the law of averages (law or large numbers) applies.
F2. Describe the sampling distribution of a statistic.
F3. Given a study, describe the sampling distribution of x-bar as specifically as possible. This involves stating whether this distribution is at least approximately normal.
F4. Given a population mean(μ), standard deviation(σ), sample size (n) and sample mean, calculate the standardized value (z-score) for a sample mean.

Confidence Intervals:

G1. Given a study, describe what role statistical inference plays in terms of the population and sample.
G2. Given a study, determine whether the study meets the "simple" conditions under which inferences on a population mean may be performed. (For example, requiring a simple random sample).
G3. Given a confidence level C, determine the critical value (z*) from the standard normal table needed to construct the confidence interval.
G4. Explain that confidence intervals are random quantities which vary from sample to sample and that they may miss the true population parameter.
G5. Explain that the confidence level is that proportion of possible samples for which the confidence interval will capture the true parameter.
G6. Construct a confidence interval for a normal population mean when the population standard deviation is known using the formula $\bar{x} \pm \frac{\sigma}{\sqrt{n}}$.
G7. Construct a confidence interval for a population proportion using the formula $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
G8. Given a study, interpret the result of a confidence interval in the context of the problem.
G9. Given a study and confidence interval for the population mean, describe how the following will affect the width of the confidence interval.
a. Increasing the sample size
b. Increasing the confidence level C
c. A larger population standard deviation

G10. Given a study and confidence interval for the population proportion, describe how the following will affect the width of the confidence interval.

d. Increasing the sample size
e. Increasing the confidence level C

G11. Given a value of σ and a confidence level, use the formula $n = \left(\frac{z^* \sigma}{m} \right)^2$ to determine the sample size needed to obtain a desired margin of error *m* for a confidence interval for μ.

G12. Given an estimate of a population proportion *p* and a confidence level, use the formula $n = \left(\frac{z^*}{m} \right)^2 p^*(1 - p^*)$ to determine the sample size needed to obtain a desired margin of error *m* for a confidence interval for *p*.

Tests of Hypothesis:

H1. Given a study objective, determine whether significance testing is appropriate.

H2. Given a study objective, choose appropriate null and alternative hypotheses, including determining whether the alternative should be one-sided or two-sided.

H3. Given a study and p-value, explain in context that p-value is a probability of getting a sample statistic as extreme or more extreme that what was seen in the sample given, that the null hypothesis is true.

H4. Given a test statistic, calculate a p-value based on the standard normal distribution.

H5. Given a study, interpret the results of a test of significance in context.

H6. Given a study objective, significance level (α) and summary statistics, conduct a formal test of significance on a population mean (and a population proportion) based on the normal distribution by conducting the appropriate steps. (This includes choosing and stating hypotheses, calculating a test statistic, calculating and interpreting the p-value and interpreting the conclusion of the test in context.)

H7. Given a study, determine the result of a hypothesis test. Explain the relationship between a confidence interval and a two-sided hypothesis test.

Inference for the Mean:

I1. Identify the favorable conditions for inferences on a mean. (For example, having data based on a SRS.)

I2. Define the standard error of a statistic.

I3. Calculate the standard error of a sample mean.

I4. Explain the purpose of the t-distribution and when it is used.

I5. Given a sampling situation, determine the appropriate degrees of freedom associated with the t-distribution.

I6. Explain the differences and similarities between the normal and t-distributions. (For example, the t-distribution is more variable but approaches normality as *n* increases.)

I7. Given a test statistic, calculate an appropriate p-value using a t-table.

I8. Describe how the shape of a t-distribution is affected by changing the degrees of freedom.

I9. Construct and interpret a one sample confidence interval for the mean based on the t-distribution. (This includes using the t-table to find the required "t*" critical value based on the confidence level "C".)
I10. Conduct a formal test of significance based on the t-distribution and draw the relevant conclusions using the language of the problem. (This includes calculating a one-sample t-test statistic.)
I11. Explain the reason a t-distribution may be used in a test of significance.
I12. Identify a matched pairs design.
I13. Conduct a statistical inference (confidence interval or significance test) based on matched pairs data.
I14. Explain how inferences based on the t-distribution are robust.
I15. Describe how varied sample sizes effect the requirements on which t-procedures may be considered. (For example, for sample sizes fifteen or less, the t-procedures should only be used if the data appear close to "normal".)

Correlation and Scatterplots:

J1. Given a study, distinguish between explanatory and response variables.
J2. Given a set of raw data, make a scatterplot by hand..
J3. Given a scatterplot, identify patterns such as positive and negative associations, non-linear patterns and outliers.
J4. Given a correlation coefficient, determine if it is legitimate. That is, determine if its value is between -1 and 1.
J5. Given two variables and their correlation coefficient, describe how the correlation changes if the units of either variable are changed.
J6. Given two variables (x and y), describe the correlation you would expect to find between x and y.
J7. Match given scatterplots with possible values of the correlation coefficient.
J8. Identify situations where the correlation coefficient would not do a good job of summarizing the relationship between two variables.

Introduction to Regression:

K1. Given the means, standard deviations and correlation of variables x and y, calculate the slope and intercept of the regression line by hand.
K2. Explain the relationship between the slope of the regression line and the correlation coefficient.
K3. Given the least squares line and a value of x, calculate the predicted value of y.
K4. Identify situations in which it is not appropriate to summarize the relationship between variables using a least squares line.
K5. Given standard regression output, interpret the estimated regression coefficients and the r^2 value.
K6. Given a study, explain in context that the regression method is used to estimate the average value of y when you know x and that individual values will vary around the predicted value.

K7. Given a study, interpret the value of the square of the correlation coefficient. That is, explain that it measures the proportion of the variance of one variable that can be explained by straight-line dependence on the other variable.
K8. Given a least squares line and an observation (x,y), calculate the residual for that observation.
K9. Given a scatterplot, identify values that are outliers.
K10. Given a scatterplot, contrast the influence of different outliers on the least squares regression.
K11. Given a study, explain why it might not be a good idea to use a least squares line to predict beyond the range of data that were used to create the line.
K12. Given a study, explain why correlation or association does not imply causation. That is, explain that the association may be due to common response, confounding or unusual events.

Experiments:

L1. Given a study, identify subjects, factors and treatments.
L2. Given a study objective, describe an appropriate comparative experiment.
L3. Given a study, determine whether a comparative experiment was used.
L4. Given a study objective, explain the advantage of a comparative experiment over a non-comparative experiment.
L5. Given a study, identify whether a placebo and/or control group were used.
L6. Given a study objective, describe what a placebo and/or control group would consist of.
L7. Given a study, determine whether a completely randomized design was used.
L8. Given a study objective, describe how to implement a completely randomized design.
L9. Given a study, explain why randomization should be used.
L10. Given a set of subjects and treatments, randomly assign subjects to treatments in a completely randomized design using a random number table or generator.
L11. (opt) Given a study objective, sketch a randomized comparative experiment.
L12. Given a study objective, explain how randomization, replication and control could be applied in a comparative experiment.
L13. Given a study objective, describe the advantages of using randomization, replication and control.
L14. Given a study, determine whether the experiment was double-blind.
L15. Given a study objective, describe how the experiment could be made double-blind.
L16. Given a study objective, explain the advantage of using a double-blind experiment.
L17. Given a study objective, decide whether the experiment should be double-blinded.
L18. Given a study, describe why "lack of realism" is a potential weakness of the experiment.
L19. Given a study objective, describe if and how a matched pairs experiment could be used.
L20. Given a study, determine whether a matched pairs experiment was used.
L21. Given a study objective, explain the advantage of using a matched pairs design.
L22. Given a study, determine whether a blocking design was used and describe the blocks.
L23. For a set of subjects divided into blocks and a set of treatments, randomize subjects in a blocking design using a random number table or generator.
L24. Given a study objective, explain the advantage of using a blocking design.
L25. Given a study objective, decide whether a blocking design should be used.

L26. (opt) Given a study, sketch the blocking design.

L27. Given a study objective, describe an appropriate comparative experiment appropriately using the principles of randomization, replication, control, blocking, double-blinding, placebo, and control group.