

## Statistics 311 Learning Objectives

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

### Data Collection and Surveys:

- A1. Given a study, identify population, sample, parameter, sampling frame and statistic.
- A2. Given a survey sample, determine whether the sample is a simple random sample (SRS), a voluntary response sample, a convenience sample, or has other forms of sampling bias.
- A3. Given a study, recognize typical forms of biases such as potential undercoverage, nonresponse, and response bias.
- A4. Given a study, describe the potential problem of selecting a sample with severe undercoverage, nonresponse, and response bias.
- A5. Given a study, decide if undercoverage, nonresponse, and response bias are likely to cause significant problems in inference.
- A6. Given a study, determine whether a SRS, stratified random sample, cluster sample, or systematic sample was selected.
- A7. Given a study's objective, explain the advantage of using a stratified random sample, cluster sample, or systematic sample over a SRS.
- A8. Given a study's objective, decide when to use a stratified random sample, cluster sample, systematic sample, 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 charts and pie graphs with appropriate software.
- 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, dot plot or stemplot by hand or with appropriate software.
- B8. Given a histogram, dot plot, or stemplot, describe the distribution's shape (skewed left, skewed right, symmetric, or multimodal), center, and spread.
- B9. Given a histogram, dot plot, or stemplot identify values that would be considered outliers.
- B10. Given a graphical summary, propose an explanation of the distribution of the data.
- B11. 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 or 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 is the same.
- C4. Explain the purpose of measures of center and measures of spread. Contrast how measures of spread and measures of center differ.
- C5. 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.
- C6. 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).
- C7. Given a histogram, be able to determine the approximate location of the median and quartiles.
- C8. Match given histograms, dot plots, or boxplots to given sets of appropriate summary statistics. (For example, mean, median, standard deviation and quartiles).
- C9. Explain the impact of outliers on summary statistics such as mean, median and standard deviation.
- C10. Given a set of raw data or five number summary, create boxplot. Given a boxplot determine the five number summary for that data.
- C11. Given a boxplot, determine the shape of the distribution (skewed right, skewed left or symmetric).
- C12. Given side-by-side boxplots, contrast key features of the groups represented by the boxplots.

### **The Normal Distribution:**

- D1. Explain that the normal distribution is a model for a bell-shaped histogram.
- D2. List the key characteristics of the normal distribution.
- D3. 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.
- D4. Given a mean  $\mu$ , standard deviation  $\sigma$ , and observed value  $x$ , calculate the standardized value (z-score). Describe the characteristics of a standard score.
- D5. Given a z-score, use a normal table to find the corresponding probability.
- D6. Given a mean  $\mu$  and standard deviation  $\sigma$ , find a specified percentile of the normal distribution. (e.g. Given a probability find the corresponding value of  $x$ .)

### Sampling Distributions:

- E1. Describe the sampling distribution of a statistic and define the standard error of a statistic.
- E2. Given a study, describe the sampling distribution of  $\bar{x}$  as specifically as possible. This involves stating whether this distribution is at least approximately normal.
- E3. Given a population standard deviation ( $\sigma$ ), calculate the standard deviation of the sample mean  $\bar{x}$ , using the formula  $\sigma/\sqrt{n}$ .
- E4. Given a population mean ( $\mu$ ), standard deviation ( $\sigma$ ), sample size ( $n$ ) and sample mean, calculate the standardized value (z-score) for a sample mean.
- E6. Given a population proportion ( $p$ ), calculate the standard deviation of the sample proportion,  $\hat{p}$ , using the formula  $\sqrt{p(1-p)/n}$ .
- E7. Given a study, describe the sampling distribution of the sample proportion ( $\hat{p}$ ) as specifically as possible. This involves stating whether this distribution is at least approximately normal.
- E8. Given a population proportion ( $p$ ), sample size ( $n$ ) and sample proportion  $\hat{p}$ , calculate the standardized value (z-score) for a sample proportion.
- E9. Calculate the standard error of the sample proportion  $\hat{p}$  using the formula  $\sqrt{\hat{p}(1-\hat{p})/n}$  or the sample mean  $\bar{x}$  using the formula  $s/\sqrt{n}$ .

### Confidence Intervals:

- F1. Given a study, describe what role statistical inference plays in terms of the population and sample.
- F2. Given a study, determine whether the study meets the conditions under which inferences on a population proportion may be performed. (For example, requiring a simple random sample).
- F3. Given a confidence level  $C$ , determine the critical value ( $z^*$ ) from the standard normal table needed to construct the confidence interval.
- F4. Explain that confidence intervals are random quantities which vary from sample to sample and that they may miss the true population parameter and explain that the confidence level is that proportion of possible samples for which the confidence interval will capture the true parameter.
- F5. Construct a confidence interval for a population proportion using the formula  $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ .
- F6. Given a study, interpret the result of a confidence interval in the context of the problem.
- F7. Given a study and confidence interval for the population proportion, describe how the following will affect the width of the confidence interval.
  - a. Increasing the sample size
  - b. Increasing the confidence level  $C$
- F8. Identify the favorable conditions for inferences on a mean. (For example, having data based on a SRS from a normal population.) Further explain how inferences based on the t-distribution are robust.

- F9. Explain the purpose of the t-distribution and when it is used.
- F10. Given a sampling situation, determine the appropriate degrees of freedom associated with the t-distribution.
- F11. 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.)
- F12. Describe how the shape of a t-distribution is affected by changing the degrees of freedom.
- F13. Construct and interpret a one sample confidence interval for the mean based on the t-distribution using the formula  $\bar{x} \pm t \frac{s}{\sqrt{n}}$  (This includes using the t-table to find the required “t\*” critical value based on the confidence level “C”.)
- F14. Given a study and confidence interval for the mean, describe how the following will affect the width of the confidence interval.
- Increasing the sample size
  - Increasing the confidence level C
- F15. Describe how varied sample sizes affect the requirements on which t-procedures may be considered.

### **Tests of Hypothesis:**

- G1. Given a study objective, determine whether significance testing is appropriate.
- G2. Given a study objective, choose appropriate null and alternative hypotheses, including determining whether the alternative should be one-sided or two-sided.
- G3. 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 than what was seen in the sample given, that the null hypothesis is true.
- G4. Given a test statistic, calculate a p-value based on the standard normal distribution or t-distribution as appropriate.
- G5. Given a study, interpret the results of a test of significance in context.
- G6. Given a study objective, significance level ( $\alpha$ ) and summary statistics, conduct a formal test of significance on a population mean (and a population proportion) 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.)
- G7. Explain the relationship between a confidence interval and a two-sided hypothesis test.
- G8. Given results from a hypothesis test, comment on the impact of sample size and the practical importance.

### **Inference for Means of Paired Differences:**

- H1. Identify a matched pairs design and when inference is appropriate in this situation.
- H2. Given a study, describe the sampling distribution of the sample mean of paired differences as specifically as possible.
- H3. Conduct a statistical inference (confidence interval or significance test) based on matched pairs data.

## Correlation, Scatterplots, Introduction to Regression:

- J1. Given a study, distinguish between explanatory and response variables.
- J2. Given a set of raw data, make a scatterplot by hand or using software.
- 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. Explain the relationship between the slope of the regression line and the correlation coefficient.
- J9. Given the least squares line and a value of  $x$ , calculate the predicted value of  $y$ .
- J10. Identify situations in which it is not appropriate to summarize the relationship between variables using a least squares line.
- J11. Given standard regression output, interpret the estimated regression coefficients and the  $r^2$  value.
- J12. 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.
- J13. Given a study, interpret the value of the square of the correlation coefficient ( $r^2$ ). 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.
- J14. Given a least squares line and an observation  $(x,y)$ , calculate the residual for that observation.
- J15. Identify situations where the correlation coefficient would not do a good job of summarizing the relationship between two variables.
- J16. Given a scatterplot, identify values that are outliers.
- J17. Given a scatterplot, contrast the influence of different outliers on the least squares regression or correlation coefficient.
- J18. 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.
- J19. 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.

**Inference for Regression:**

- K1. Given a study explain what is meant by the true regression line and describe the sampling distribution of the sample slope.
- K2. State the assumptions of inference about the regression model.
- K3. Explain the practical reason for testing that the slope is zero.
- K4. Given a study objective, significance level ( $\alpha$ ) and summary statistics, conduct a formal test of significance on a slope based on the t-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.)
- K5. Given standard regression output, interpret the results of the test of hypothesis about the slope.

**Experiments:**

- L1. Given a study, determine whether it is an observational study or an experiment.
- L2. Given a study, identify subjects and treatments.
- L3. Given a study objective, describe an appropriate comparative experiment.
- L4. Given a study, determine whether a comparative experiment was used.
- L5. Given a study objective, explain the advantage of a comparative experiment over a non-comparative experiment.
- L6. Given a study, identify whether a placebo and/or control group were used.
- L7. Given a study objective, describe what a placebo and/or control group would consist of.
- L8. Given a study, determine whether a completely randomized design was used.
- L9. Given a study objective, describe how to implement a completely randomized design.
- L10. Given a study, explain why randomization should be used.
- L11. Given a set of subjects and treatments, randomly assign subjects to treatments in a completely randomized design.
- 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 blinded or double-blinded.
- 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 objective, describe if and how a matched pairs experiment could be used.
- L19. Given a study, determine whether a matched pairs experiment was used.
- L20. Given a study objective, explain the advantage of using a matched pairs design.
- L21. Given a study, determine whether a blocking design was used and describe the blocks.
- L22. For a set of subjects divided into blocks and a set of treatments, randomize subjects in a blocking design.
- L23. Given a study objective, explain the advantage of using a blocking design.
- L24. Given a study objective, decide whether a blocking design should be used.
- L25. Given a study objective, describe an appropriate comparative experiment appropriately using the principles of randomization, replication, control, blocking, double-blinding, placebo, and control group.