

- LO 1.** Use scatterplots for describing the relationship between two numerical variables making sure to note the direction (positive or negative), form (linear or non-linear) and the strength of the relationship as well as any unusual observations that stand out.
- LO 2.** When describing the distribution of a numerical variable, mention its shape, center, and spread, as well as any unusual observations.
- LO 3.** Note that there are three commonly used measures of center and spread:
- center: mean (the arithmetic average), median (the midpoint), mode (the most frequent observation).
 - spread: spread: standard deviation (variability around the mean), range (max-min), interquartile range (middle 50% of the distribution).
- LO 4.** Identify the shape of a distribution as symmetric, right skewed, or left skewed, and unimodal, bimodal, multimodal, or uniform.
- LO 5.** Find the 5 number summary and IQR, and draw a box plot with outliers shown.
- LO 6.** Create stem-and-leaf plots, dot plots, and histograms to visualize the distribution of a numerical variable.
- LO 7.** Use the above three graphs along with box plots to identify the shape, center, and spread of numerical distributions.
- LO 8.** Calculate, interpret, and apply Standard Deviation as a measure of spread about the mean.
- LO 9.** Define a robust statistic (e.g. median, IQR) as measures that are not heavily affected by skewness and extreme outliers, and determine when they are more appropriate measures of center and spread compared to other similar statistics.
- LO 10.** Recognize when transformations (e.g. log) can make the distribution of data more symmetric, and hence easier to model.
- * *Reading: Section 2.1 of Advanced High School Statistics*
 - * *Videos:*
 - *Mean, median, and mode, Dr. Çetinkaya-Rundel (1:16)*
 - *Visualizing distributions of numerical variables, Dr. Çetinkaya-Rundel (6:00)*
 - *Reading box plots, KhanAcademy (3:18)*
 - * *Test yourself:*
 1. *Describe what is meant by robust statistics and when they are used.*
 2. *Describe when and why we might want to apply a log transformation to a variable.*
- LO 11.** Use frequency tables and bar plots to describe the distribution of one categorical variable.
- LO 12.** Use contingency tables and segmented bar plots or mosaic plots to assess the relationship between two categorical variables.
- LO 13.** Use back-to-back stem-and-leaf plots, side-by-side box plots, or pairs of dot plots/histograms for comparing the distribution of a numerical variable across multiple groups (e.g. male/female).

* *Reading: Section 2.3 of Advanced High School Statistics*

* *Video: Exploring relationships between categorical variables, Dr. Çetinkaya-Rundel (4:37)*

* *Test yourself:*

1. *You collect data on 100 classmates, 70 females and 30 males. 10% of the class are smokers, and smoking is independent of gender. Calculate how many males and females would be expected to be smokers. Sketch a bar graph of this scenario.*

LO 14. Note that an observed difference in sample statistics suggesting dependence between variables may be due to random chance, and that we need to use hypothesis testing to determine if this difference is too large to be attributed to random chance.

LO 15. Set up null and alternative hypotheses for testing for independence between variables, and evaluate data's support for these hypotheses using a simulation technique.

* *Reading: Section 2.4 of Advanced High School Statistics*

* *Test yourself: Explain why difference in sample proportions across two groups does not necessarily indicate dependence between the two variables involved?*