

Module 2a – Describing Distributions with Numbers

Reviewing Module 1

- **Two main variable types (picturing distribution)**
 - Categorical (Bar Graphs/Pie Charts)
 - Quant./Continuous (Histogram/Stemplot/Boxplot)
 - [How to create a stemplot](#) (*note the resulting “shape” – similar to a histogram but vertical instead of horizontal*)
- **We can describe the distribution of a quant/continuous variable**
 - Shape/Center/Spread

Stat Procedure Diagram – Where we are

		Descriptive Statistics (Describing Pops or Samples)		Inferential Statistics (from Samples)
	Variable Types	Display	Describe	Estimation
Univariate	categorical (nominal or ordinal)*	Bar Graph/Pie Chart	Counts/Percentages	When binary/dichotomous: Confidence interval for proportions
	quantitative/continuous	Histogram/Stem & Leaf Box Plot	Mean/St Dev (normal) Median/Min, Q1, Q3, Max (skewed)	Confidence interval for means
		Display	Describe	Significance Tests/Hypothesis Tests
Bivariate	2 categorical	Tables or Bar Graphs	Two-way tables/Crosstabulation	Chi-square test (for goodness of fit)
	1 categorical, 1 quant.	Bar Graphs	Comparison of means/averages	T-test (one sample/group, two samples/groups) ANOVA (two or more samples/groups)
	2 quant.	Scatterplot	Correlation Coef. (Coef. of determ)/ Regression Line	T-test for correlation
		Display	Describe	Significance Tests/Hypothesis Tests
Multivariate	Response Variable is Quant.	-	Ordinary Least Squares Regression (OLS)	F-test for overall model T-tests for each explanatory variable
	Response Variable is categorical (dichotomous)	-	Logistic Regression	Chi-square tests of significance

NOTE: Items highlighted in yellow are covered in this course.

*When a categorical variable has two categories, it is called dichotomous.

Describing distributions w/ numbers

- **Categorical**

- Counts (or frequency) and Percentages (rate) in each category
 - *We tend to prefer **rates** over counts*

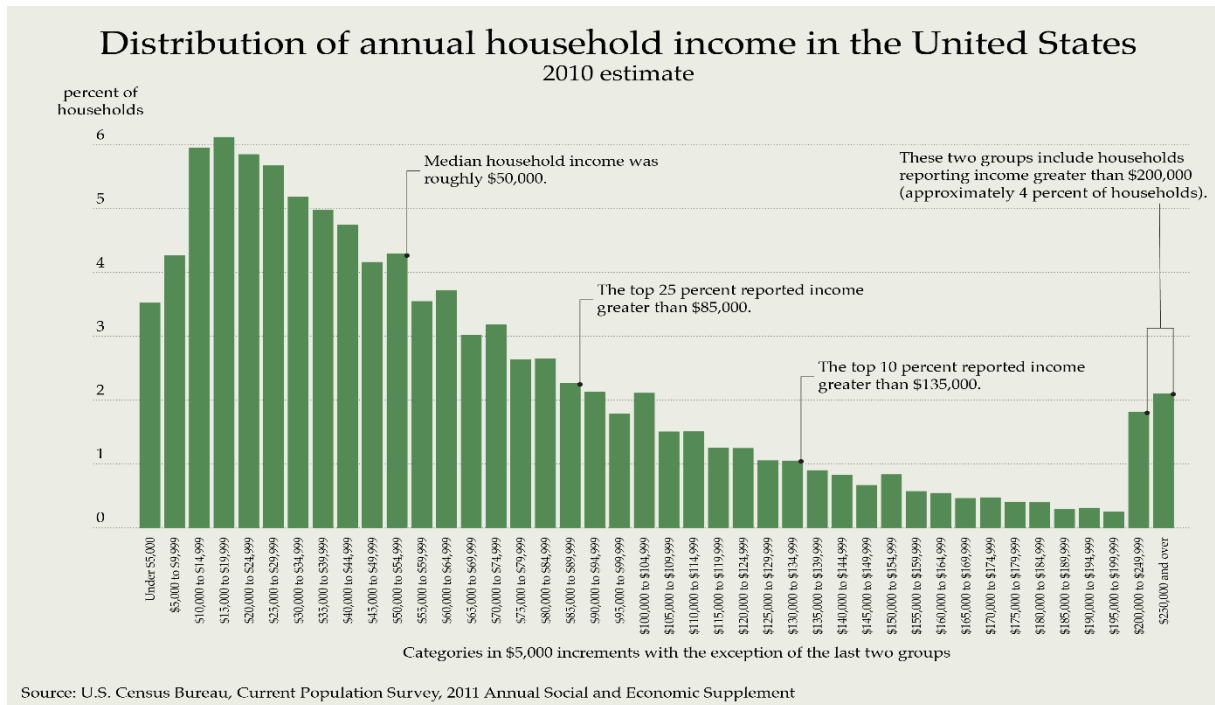
- **Quantitative/Continuous**

- Measures of *center* and *spread/variation*
- Symmetric Distributions - **Mean & Standard Deviation**
- Skewed Distributions – **Median & (min, quartile 1, median, quartile 3, max)** – *a.k.a Five number summary*

- [Watch “Against All Odds” Unit 4, Measures of Center](#)

Example: Skewed vs Symmetric Distribution

(Beware of the mean – it is susceptible to outliers)



- In skewed distributions, the mean is pulled toward the tail, giving a false sense of center. We use the five number summary to represent center and spread.
- Example: skewed right/high

- In relatively symmetric distributions, the mean and median are close together. We use Mean and Standard deviation to represent the center and spread.
- $\text{Mean} > \text{Median}$ = skewed right;
- $\text{Mean} < \text{Median}$ = skewed left;
- $\text{Mean} = \text{Median}$ = symmetric

