Midterm Review
(Chapters 1, 2, 3, 4, 5, 8, 9, 12, 13, intro to 15)

INTRO AND CHAPTER 1 (Picturing Distributions with Graphs)

Qualitative vs Quantitative Research
Divide in Social Science
Individual (unit of analysis): People, animals or things.
Variable – description/measurement of an individual.
Descriptive statistics – summarizing the data
Inferential statistics – generalizing findings.
Statistic/Sample
Parameter/Population
Randomization and Bias

Variable Types - Categorical and Quantitative (nominal, ordinal, continuous/interval) (dichotomous).

univariate/bivariate/multivariate

PICTURING DISTRIBUTIONS
Display Categorical: Bar Graph/Pie Graph
Display Quant: Histogram/Stemplot

CHAPTER 2 (Describing Distributions with Numbers

DESCRIBING DISTRIBUTIONS
Quantitative Variable Description: Center, Spread, and Shape

Center: Mean, Median, Mode
Spread/Variability:
Standard Deviation vs. 5 # summary
Calculating the mean and st dev (and variance)
Finding the 5 # summary and box plot
Overall Range and Interquartile Range
Shape: Symmetric or Skewed –
Unimodal, Bi-Modal, Multimodal (Bell Shaped)
Any outliers?

When to use Mean vs Median

CHAPTER 3 (The Normal Distributions)

Normal Distributions
Uni-modal, no outliers, symmetric, bell-shaped (mean/median are the same)
Examples?
Using the 68, 95, 99.7 rule
Be able to draw & label a normal distribution

Density Curve = 1 or 100 percent.
Area under the curve=100%

Calculating a Standard Score/z-score
-Using formula to calculate z-score
-Using Table A to convert to percentiles

CHAPTER 4 (Scatterplots and Correlation)

Bivariate Analysis (relationship between 2 variables)

When comparing two quantitative variables…
(Are two variables associated/related?)

Relationship Qualities: Form/Direction/Strength
Form: (Linear, Curvilinear, Diffuse)
Direction: (Positive or Negative)
Strength: (Weak/Moderate/Strong)

Display relationship with…Scatterplot
Explanatory Variable (x-axis): (independent/causal)
Response Variable (y-axis): (dependent/effect)

CHAPTER 5 (Regression and Relationships)

Describe relationship with…
Correlation Coefficient (be able to interpret)
(ranges from -1 to +1)
Regression (trend line) (be able to interpret)
(line that is closest to all points – shows the trend/relationship) y=a+bx
Error terms/residual (random error)
Define and interpret **slope** and **intercept**.
Beware of **extrapolation**

Coefficient of Determination (R-squared)
(the percent of y explained by x)

Correlation vs Causation (association is not causation)

Lurking Variables (often a third variable explains variation in the other two)

CHAPTER 8 (Sampling and Inference)

Where data come from…
Observational studies:
Sampling and Surveying a population.

Inference: Can sample results be generalized to the population?

Good samples: **Probability Samples**
Gold Standard: **Simple Random Sample**

Random selection (every individual in the population has equal chance of being selected for sample)
Know how to use **random number table**

Bad samples: convenience/**voluntary response (nonresponse)**/ snowball

Random error vs Bias (random error can be accounted for)

Bias – three types
1. Sampling error/bias (bad sampling/non response bias etc)
2. Respondent error/bias (bias from invalid responses – respondents misrepresent, don't understand, answer too quickly/guess, etc.)
3. Measurement error/bias (bias from bad surveys – biased questions, bad measurement)

CHAPTER 9 (Experiments)

Strength – isolating a cause and effect
Weakness – may not represent real world
Classical Experimental Design:
1. **Explanatory** variable (**treatment**) and **Response** variable (**outcome**).
2. Randomly assigns subjects to **treatment** and **control** groups
3. **Pre-test & Post-test** on response variable.

Importance of using a **placebo** for control group
Importance of conducting a **double blind** exp.

Hawthorne Effect: The act of observing can affect the outcome (bias).

Chapters 12 and 13 (Probability)

Random behavior (unpredictable in the short term, predictable in the long term)

Personal probability (gut feeling/not scientific)

Probability model (possible outcomes with probability of each)

Probability Rules:
1. All probabilities must be between 0 and 100
2. All outcomes must add to 100 percent
3. Addition rule for disjoint events:
P(A or B)= P(A)+P(B)
4. Multiplication rule for independent events
P(A and B)=P(A)*P(B)

Intro to Chapter 15 (Sampling)

Central Limit Theorem: Why sampling works for inference: The sampling distribution has a mean that approximates the population mean.
Law of large numbers:
As a sample size increases, the sample estimate (statistic) comes closer to the actual population value (parameter).
Example: The more times you flip a coin, the closer your percentage heads/tails will come to 50/50 (unpredictable in the short run/predictable in the long run).

The larger the sample size, the more accurate your estimate/prediction of the actual population value (usually the mean).

Standard Error: Std. Dev. / sq root of n
Confidence Intervals: Consist of confidence level and margin of error.