Final Review Ch. 1

(1A)
- Two Types of Data
  - Categorical (words)
  - Quantitative (# measure units)

(1B)
- Methods of Collecting Data
  - Census (entire Pop.) } Good!
  - Random Sample
  - Convenience } Bad
  - Voluntary Response

(1C)
Bias data (does not reflect Population)

5 types of Bias
1. Sampling Bias (method is bad)
2. Question Bias
3. Response Bias (lie, inaccurate)
4. Non-Response Bias (refuse to give data)
5. Deliberate Bias (falsify data, deleting data, falsify report, leaving out groups from population)

(1D)
Experiments
- Prove Cause & Effect
- Controlling Confounding variables
- Two or more groups similar in all confounding variables
  "Random Assignment"

(1E)
Categorical Data Analysis

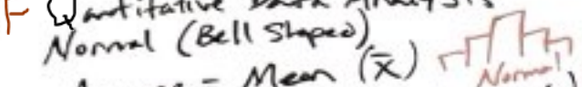
$\text{Proportion} = \frac{\text{amount}}{\text{total}}$ (decimal equiv. of %)

Percent = Proportion × 100%

Proportion = Percent ÷ 100 (remove "%")

Bar Chart (Statkey)

(1F)
Quantitative Data Analysis
Normal (Bell Shaped)



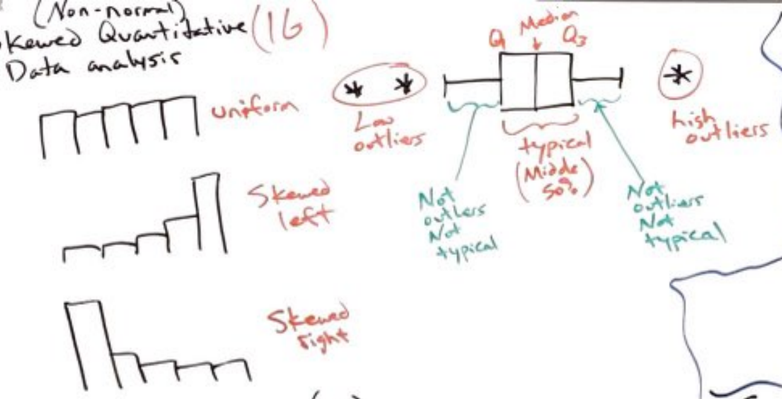
Average = Mean ($\bar{x}$)
Spread = Standard Deviation (s)
Typical = within 1 stnd dev of mean
(between $\bar{x}-s$ and $\bar{x}+s$)
(middle 68%)

High outliers (unusual high): 2 or more standard deviations above the mean
$\geq \bar{x}+2s$
(top 2.5%)

Low outliers (unusually low): (bottom 2.5%)
2 or more standard deviations below the mean
$\leq \bar{x}-2s$

(Empirical Rule %)

(Non-Normal) Skewed Quantitative Data analysis (1G)



- Average = Median (Q_2)
- Spread = IQR = $Q_3 - Q_1$
- Typical: Between Q_1 and Q_3 (middle 50%)
- High outliers (unusual high) $\geq Q_3 + (1.5 \times IQR)$ stars on right side of box plot
- Low outliers (unusual low) $\leq Q_1 - (1.5 \times IQR)$ stars on left side of box plot

Letters (2A)

Sample Size: n

Sample correlation Coefficient: r

Sample mean: $\bar{x}$

Population mean: μ

Sample standard deviation: s

Sample proportion: $\hat{p}$

Population proportion: π or p

Confidence Intervals (2D, 2E, 2F)

Confidence Levels: 90%, 95%, 99%

Confidence Interval Sentence: "I am 95% confident that Population Parameter is between # and #."

Margin of Error = Critical Value (T or Z) × Standard Error

Lower Limit = Sample Statistic − Margin of Error

Upper Limit = Sample Statistic + Margin of Error

Two-Population Confidence Intervals

(−, −) Pop. 1 sig lower than Pop 2

(+, +) Pop 1 sig. higher than Pop 2

(−, +) No sig difference between Populations (Close)

Sampling Distributions Central Limit Theorem (2B & 2C)

Sampling Variability: Random Samples Stats are usually different

Sampling Distribution: Lots of random sample stats on same graph

* Standard Error: Standard Deviation of Sampling Distribution

Population Parameter: Center of sampling distribution

Sample Size Increases → Less Error → Sampling Distribution More Normal

Sample Size Decreases → More Error → Sampling Distribution Less Normal

* Bootstrapping: Taking random samples with replacement from a sample.