• CH1: Picturing Distributions With Graphs
  1. Types of Variable - Categorical - Quantitative
  2. Representations of Distributions
     (a) Categorical - Pie Chart - Bar Graph
     (b) Quantitative - Stem/Leaf plot - Histogram

• CH2: Numerical Summaries of Distributions
  1. Mean
  2. Median
  3. Quartiles
  4. Box-Plot
  5. Outliers
  6. Variance/Standard Deviation

• CH3: Normal Distributions
  1. Definition of a Density
  2. 68-95-99.7 Rule
  3. Finding Normal Probabilities/ Proportions (Standardizing)
  4. Finding Normal Percentiles (z-scores) - Ex: Find c so that
     \[ P(N(3, 5) \leq c) = .12 \]

• CH4: Scatterplots and Correlation
  1. How to make a scatter plot
  2. Association in Data (Linear, Positive, Negative)
  3. Correlation Coefficient

• CH5: Regression
  1. Least Squares Regression Line
     \[ \hat{y}(x) = (r \frac{s_y}{s_x})x + (\bar{y} - \bar{x}(r \frac{s_y}{s_x})) \]
  2. Purpose of Least Squares Regression Line
3. Prediction using Least Squares Regression line

- CH8/9: Producing Data Samples
  1. Definition: Simple Random Sample
  2. Sampling Issues
     Examples: Bias Generated By Voluntary Response, etc
  3. Def: Double Blind Experiments
  4. Lurking Variables

- CH10: Introduction to Probability
  1. Probability Model
     (a) Sample Space- $S$
     (b) Set of Events- $S$
     (c) Assignment of Probabilities- $P$
  2. Rules of Probability:
     If $A$ and $B$ are disjoint events (don’t share any outcomes):
     (a) $0 \leq P(A) \leq 1$
     (b) $P(S) = 1$
     (c) $P(A$ happens OR $B$ happens) = $P(A) + P(B)$
     (d) If all outcomes in the sample space are equally likely then
        $$P(A) = \frac{\text{Number of outcomes in } A}{\text{Total Number of outcomes}}$$

- CH11: Sampling Distributions
  1. Definition: Parameter/Statistic
  2. If $X_1, ..., X_n \sim N(\mu, \sigma)$ are normal observations with mean $\mu$ and
     standard deviation $\sigma$ then
     $$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$
  3. Law of Large Numbers
     As $n$ gets larger and larger, $\bar{X}$ gets closer and closer to $\mu$.
  4. Central Limit Theorem: If $X_1, ..., X_n$ come from a distribution with mean
     $\mu$ and standard deviation $\sigma$, then
     $$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$
     for $n$ large.
• CH14: Introduction to Inference

1. Definition of a $\alpha$ level confidence interval.

\[ \text{estimate} \pm \text{error} \]

2. Interpretation of a $\alpha$ level confidence interval: We are $\alpha$ % sure that the true value of the parameter lies in the confidence interval.

3. Confidence interval for $\mu$ when $\sigma$ is known:

\[ \bar{X} \pm Z_{\alpha} \frac{\sigma}{\sqrt{n}} \]

4. Definition of p-value

5. Hypothesis Testing: Generic Method
   (a) Write down definition of p-value in terms of the test statistic
   (b) Standardize the test statistic
   (c) Recognize the distribution after standardization
   (d) Look up the probability in the appropriate table or use software.
   (e) Check p-value against significance level and decide whether or not to reject. If p is smaller than significance: Reject $H_0$; Otherwise: Cannot reject.

• CH 15: Thinking about Inference

1. Be familiar with how confidence Intervals behave.
2. Know the difference between Type 1 and Type 2 error.
3. Be able to find the Type 2 error given the power of the test.
4. Sample size for desired margin of error.

\[ n = \frac{z^*\sigma}{m} \]

5. Performing hypothesis test for $\mu$ when $\sigma$ is known.

• CH18: One Sample t procedures: Inference for the mean when $\sigma$ is unknown.

1. Conditions for t test:
   (a) Observations have a normal distribution with mean $\mu$ and st. dev $\sigma$, which are unknown.
(b) Observations come from a distribution with mean \( \mu \) and st. dev \( \sigma \), which are unknown, the distribution is roughly symmetric without outliers.

2. One sample t statistic:
\[
\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)
\]

3. Confidence interval of size \( \alpha \) for \( \mu \) when \( \sigma \) is unknown:
\[
\bar{X} \pm t^* \frac{S}{\sqrt{n}}
\]

4. Hypothesis Tests for \( \mu \) when \( \sigma \) is unknown.

- **CH19**: Two Sample t procedures.
  1. Inference for \( \mu_x - \mu_y \)
  2. Two-Sample t statistic:
\[
t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \sim t(\min(n-1, m-1))
\]
  3. Conf Interval for \( \mu_x - \mu_y \):
\[
\bar{X} - \bar{Y} \pm t^* \sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}
\]
   \( t^* \) is the \( \alpha \) Critical value from the \( t(\min(n-1, m-1)) \)
  4. Hypothesis tests for \( \mu_x = \mu_y \) i.e. \( \mu_x - \mu_y = 0 \).

- **CH20**: Inference About population proportions.
  1. Definition of \( \hat{p} \)
\[
\hat{p} = \frac{\text{Number of Successes}}{\text{Sample Size}}
\]
  2. Distribution of \( \hat{p} \sim N(p, \sqrt{\frac{p(1-p)}{n}}) \)
  3. Conf. Interval for \( p \)
\[
\hat{p} \pm Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}
\]

- **CH23**: \( \chi^2 \) test
1. Expected Counts: \[
\begin{array}{c}
\text{RowTotal x ColumnTotal} \\
\text{Sample Size} \\
\text{npi,0}
\end{array}
\]

2. \(\chi^2\) Statistic

\[\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected} \sim \chi^2((r - 1)(c - 1))\]

Where the sum is taken over all cells in the two way table.

\[\chi^2 = \sum \frac{(Observed_i - np_{i,0})^2}{np_{i,0}} \sim \chi^2(k - 1)\]

Where the sum is taken over all \(k\) cells in the distribution table.

- CH25: ANOVA
  1. Know how and when to use the ANOVA test.
  2. Know how to read an ANOVA computer output.
  3. Know how to make confidence intervals using the pooled standard deviation.