Review - Exam 1
Ch 1 - 5

Variables
1) Numeric
   a) discrete
   b) continuous

2) Categorical
   a) ranked
   b) unranked
Tables / Graphical Displays

1) Dot Plot
2) Stem and Leaf Plot
3) Box Plots
4) Bar Graph
5) Pie Chart
6) Frequency Distribution
7) Histogram

Misusing Statistics
Calculations

\[ \sum x \]
\[ \sum x^2 \]
\[ (\sum x)^2 \]
\[ \sum (x - \bar{x}) \]
\[ \sum (x - \bar{x})^2 \]
\[ \frac{n!}{m! (n-m)!} \]
\[ b^n \]
Averages - Centers of Data

Mean: \( \bar{X} = \frac{\sum X}{n} \)

Median: \( \bar{X} = \frac{1}{2} (n+1) \)th ranked observation

Mode: most frequent observed value
Sample Standard Deviation

\[ S = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} \]

or

\[ S = \sqrt{\frac{\sum x^2 - (\sum x)^2}{n}} \cdot \frac{1}{n-1} \]

Interpretation
Interpretation

Value of $s$ indicates how "spread out" the data is.

1) $s = 0 \rightarrow$ No variation in the data; values all the same.

2) $s$ "small" $\rightarrow$ the data values are not widely dispersed.

3) $s$ "large" $\rightarrow$ the data values are widely dispersed.
$$\text{IQR} = Q_3 - Q_1$$

$$\text{Skewness} = \frac{3(\bar{X} - \overline{X})}{S}$$

Symmetric Dist.
Pos. Skewed "
Neg. Skewed "

5-Number Summary
Boxplots
Chapter 3 - Correlation

Pearson Correlation Coefficient.

\[ r = \frac{SS(xy)}{\sqrt{SS(x) \cdot SS(y)}} \]

where

\[ SS(xy) = \sum xy - \frac{(\sum x)(\sum y)}{n} \]

\[ SS(x) = \sum x^2 - \frac{(\sum x)^2}{n} \]

\[ SS(y) = \sum y^2 - \frac{(\sum y)^2}{n} \]
Interpretation
1) Positive Corr.
2) Negative Corr.
3) Zero Corr.
4) $-1 \leq r \leq 1$
5) Cause and Effect
Regression Analysis

Regression (Prediction) Equation

\[ y = b_0 + b_1 x \]

\[ b_1 = \frac{SS(xy)}{SS(x)} \]

\[ b_0 = \bar{y} - b_1 \bar{x} \]

Predicting Values of \( y \)

Point Estimator: \( \hat{y} = b_0 + b_1 x \)
Probability - Relative Frequency

Definition

Def: Suppose an experiment consists of $n$ trials, and $k$ of these trials result in event $E$. Then

$$\hat{P}(E) = \frac{k}{n} = \frac{\# \text{ successful repetitions}}{\text{total \# repetitions}}$$

Note: This is called the empirical probability of an event or the relative frequency of the event.
Probability - Equally Likely Outcomes

Def: Suppose an experiment can result in one of \( m \) equally likely outcomes. Suppose that \( r \) of these outcomes result in event \( A \) occurring. Then the theoretical probability of event \( A \) is

\[
P(A) = \frac{r}{m}
\]

\( = \frac{\text{# outcomes in event } A}{\text{total # possible outcomes}}\)

Note: For each outcome in S.S.

\[
P(\text{outcome}) = \frac{1}{\text{total # possible outcomes}}
\]
A discrete probability distribution is a list (or description) of the values the random variable can have, along with the associated probabilities.

We can do this using a probability tree.
Rules

The probability of an event $E$ is always between 0 and 1, inclusive:

$$0 \leq P(E) \leq 1$$

$P(E) = 0 \quad \rightarrow \quad$ event $E$ cannot occur

$P(E) = 1 \quad \rightarrow \quad$ event $E$ must always occur
2) The probability of event $A$ is equal to the sum of the probabilities of the outcomes in event $A$

$$P(A) = \sum_{\text{all outcomes in } A} P(\text{outcome})$$

**Complementary Event**

Def: Suppose $A$ is an event. The complement of event $A$, denoted "not $A$", is the event "$A$ does not occur".

**Rule of Complementary Events**

$$P(\text{not } A) = 1 - P(A)$$