Unit 7
Statistics

AFM
Mrs. Valentine

7.1 Samples and Surveys

- *Obj.*: I will understand the different methods of sampling and studying data. I will be able to determine the type used in an example, and select the best choice for collecting/studying data.

- **Vocabulary**
  - Statistics
  - Population
  - Convenience Sample
  - Self-Selected Sample
  - Systematic Sample
  - Random Sample
  - Bias
  - Observational Study
  - Controlled Experiment
  - Survey

7.1 Samples and Surveys

- Statistics is the study, analysis, and interpretation of data.
- Analyzing Sampling Methods
  - Population – members of a set

7.1 Samples and Surveys

- Samples
  - Convenience Sample – members are readily available
  - Self-Selected Sample – volunteers only
  - Systematic Sample – selected from regular intervals in ordered set
  - Random Sample – all members equally likely to be chosen

7.1 Samples and Surveys

- Bias
  - Over/underrepresented part of population
  - Systematic error introduced by sampling method
- Data from non-random samples may be true, but will likely be suspect.

7.1 Samples and Surveys

- Analyzing Study Methods
  - Observational Study – observe but do not affect the study
  - Controlled Experiment – two groups: treatment imposed on one but not the other. Results of the two groups are compared.
  - Survey – ask every member a set of questions.

7.2 Central Tendency

- *Obj.*: I will be able to determine the measures of central tendency for a set of data.

- **Vocabulary**
  - Central Tendency
  - Mean
  - Median
  - Mode
  - Outliers
  - Quartiles
  - Interquartile Range
  - Percentile

7.2 Central Tendency

- Measures of Central Tendency
  - Mean – average of the data
  - Median
    - Odd # data points: middle value of the data
    - Even # data points: mean of two middle values
  - Mode – most frequent values
    - Bimodal (two modes)
    - 3+ modes → not statistically relevant

7.2 Central Tendency

- Outliers
  - Misleading data, different from the rest
  - Affect measures of central tendency
  - Use differences between adjacent values in ordered data to identify

7.2 Central Tendency

- Comparing Data Sets
  - Range of data set – difference between greatest and least values
  - Quartiles – if data is divided into four parts by medians, the values separating the four parts are quartiles.
  - Interquartile range – difference between first and third quartiles
- Percentile
  - Number from 0 – 100 that shows the percent of data less than or equal to x.

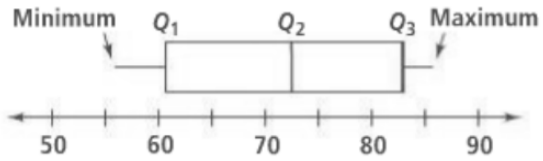
7.3 Interpreting Graphical Data

- *Obj.*: I will be able to interpret graphical data to find measures of central tendency.

- **Vocabulary**
  - Box-and-Whisker Plots
  - Stem-and-Leaf Display
  - Dot Plot
  - Histogram
  - Class Interval

7.3 Interpreting Graphical Data

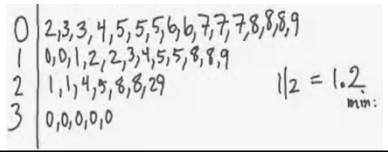
- Box-and-Whisker Plots
  - Quartiles bound center box.
  - Minimum and maximum values are used to form whiskers.



  - In calculator: STAT, enter data in L1, enter window values, draw box-and-whisker plot.
  -

7.3 Interpreting Graphical Data

- Stem-and-Leaf Display
  - Choose the leading digit or digits to be the stems. Arrange the stems vertically from lowest to highest value from top to bottom.
  - The last digit is the leaf. Record a leaf for each data value on the same horizontal line as its corresponding stem. Arrange the leaves from lowest to highest value from left to right.
  - Provide a key to show what the data means, including units.



7.3 Interpreting Graphical Data

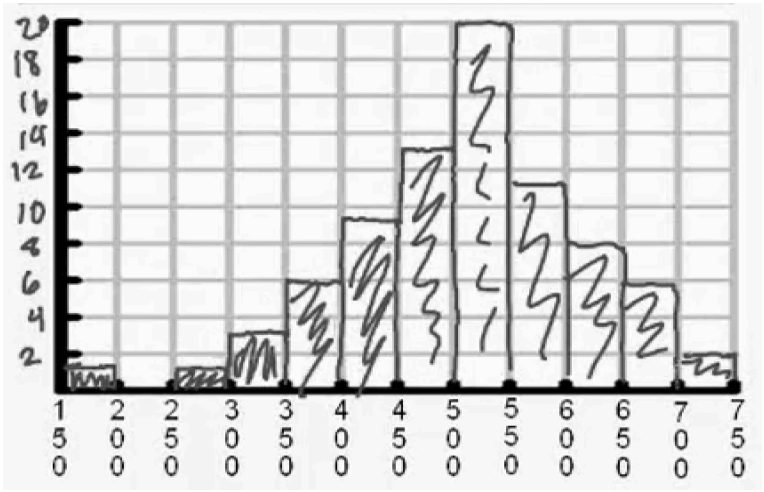
- Dot Plot
  - A graphical display of data using dots.
  - A number line is provided with the full set of possible values.
  - Each dot above a value represents a single occurrence of that value.
  - Shows relative frequency and distribution of data.

7.3 Interpreting Graphical Data

7.3 Interpreting Graphical Data

- Histogram
  - Similar to a bar graph.
  - Divide the range of the data into classes of equal width, so that each data value is in exactly one class.
  - The width of these intervals is called the class interval.
  - Draw a horizontal axis and indicate the first value in each class interval.
  - Draw a vertical scale and label it with either frequencies or relative frequencies.
  - Draw rectangles with a width equal to the class interval and height equal to the frequency of the data within each interval.

7.3 Interpreting Graphical Data



7.4 Standard Deviation

- *Obj.*: I will be able to calculate standard deviation for a set of data. I will be able to calculate z-scores of values in the set.

- **Vocabulary**
  - Standard deviation
  - Variance
  - Z-scores

7.4 Standard Deviation

- Standard Deviation and Variance
  - Measures showing deviation from mean
    - σ (sigma) = standard deviation
    - σ^2 (sigma squared) = variance

7.4 Standard Deviation

- Finding standard deviation
  - Find the mean $\overline{x}$
  - Find the difference between each value and the mean: $x - \overline{x}$
  - Square each difference $(x - \overline{x})^2$ $\sigma^2 = \frac{\sum(x-\bar{x})^2}{n}$
  - Find the mean of the squares (= variance)
  - Take the square rot of the variance (= standard deviation) $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$

7.4 Standard Deviation

- Using the Calculator
  - STAT → EDIT → Enter data in L1
  - STAT → CALC → 1-Var Stats
  - $\overline{x}$ = mean
  - σX = standard deviation

7.4 Standard Deviation

- Describing Data with Standard Deviation
  - Written a range
  - Example: $x=50$ $\sigma=10$, then a value x with $40 \leq x \leq 60$ is within one standard deviation of the mean.
- Z-scores
  - Z-scores are the number of standard deviations there are between a given value and the mean of the set.

$$z-score = \frac{value - mean}{standard\ deviation} = \frac{x - \bar{x}}{\sigma}$$

  - Can be used to find outliers; if a data point has a z-score > |3|, it is an outlier

7.5 Normal Distribution

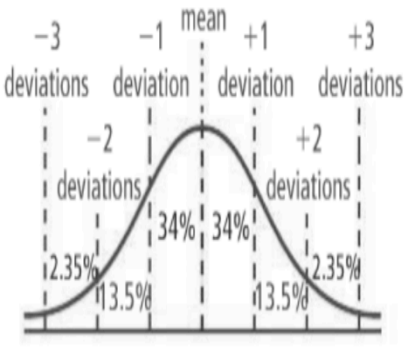
- *Obj.*: I will be able to read, create, and use normal distributions. I will be able to determine probability from a normal distribution.

- **Vocabulary**
  - Normal Distribution
  - Skewed Data

7.5 Normal Distribution

- Normal Distributions
  - Discrete probability distribution – finite number possible events
  - Continuous probability distribution – any value in an interval of real numbers (usually large data sets)
  - Normal distribution – data that vary randomly from the mean.

7.5 Normal Distribution



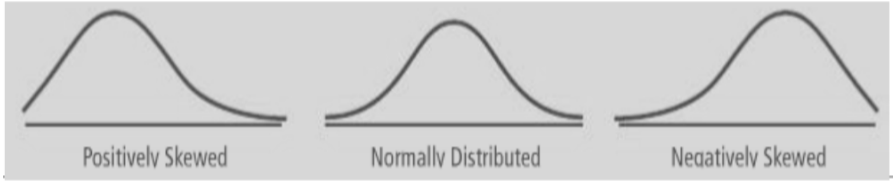
68% of data fall within 1 standard deviation

95% of data fall within 2 standard deviations

99% of data fall within 3 standard deviations

7.5 Normal Distribution

- Skewed data (do not vary predictably from mean)



7.5 Normal Distribution

- Sketching a Normal Curve
  - Use the symmetry of a normal distribution to help draw the curve.
  - Find the mean and standard deviation of the population.
  - Multiply the standard deviation by 1, 2, and 3
  - Draw vertical lines at the mean ± these values.
  - Sketch the normal curve.
- Analyzing Normal Distribution

7.5 Normal Distribution

- Analyzing Normal Distribution
  - The area between the curve and the interval on the x-axis represents probability.

7.6 Binomial Distribution

- *Obj.*: I will be able to find probabilities for binomial distributions and related them to normal distributions

- **Vocabulary**
  - Binomial Distribution
  - Correction for Continuity
  - Delimiting value

7.6 Binomial Distribution

- Binomial Distribution
  - A binomial distribution is very different from a normal distribution, but it can have a similar shape if the sample size is large enough.
  - A binomial distribution is discrete, not continuous.

7.6 Binomial Distribution

- Requirements for a binomial distribution:
  - The random variable of interest is the count of successes in n trials.
  - The number of trials (or sample size), n, is fixed
  - Trials and independent, with fixed value p = P(success on a trial) and q = P(failure on a trial)
  - There are only two possible outcomes on each trial, called "success" and "failure."

7.6 Binomial Distribution

- Notation: B(n, p)
- Binomial Probability Formula: ${}_nC_x p^x (1-p)^{n-x}$
- To find the relative frequency distribution for set X,
  - Use keystrokes 2nd DISTR ALPHA A to pull up binompdf(
  - Enter values for n, p, and X → binompdf(n,p,X)

7.6 Binomial Distribution

- Select ENTER to see the relative frequency for that value of X.
- Repeat for all X values to get the relative frequency distribution

- Create a histogram of the relative frequencies of X to see the graph of a binomial distribution.

7.6 Binomial Distribution

- Relating Binomial Distribution to Normal Distribution
  - Approximate binomial probabilities resemble normal curve areas when the sample size n is very large.
  - The area under a normal curve that would be determined to approximate the given binomial probability can be found by using a correction for continuity.

7.6 Binomial Distribution

- Subtract 0.5 from the bottom value of the range and add 0.5 to the top value of the range to find the lower and upper delimiting values.
- Approximate with Normal Distribution when:
  - $0 \leq \mu \pm 3\sigma \leq n$
  - $np(1-p) \geq 10$

7.6 Binomial Distribution

- The formula for the mean is $\mu = np$
- The formula for the standard deviation is

$$\sigma = \sqrt{np(1-p)}$$

- Find the z-score for X, and find the probability of the z-value.

$$z = \frac{(x + 0.5\hat{}) - \mu}{\sigma}$$

7.6 Binomial Distribution

- Use key strokes 2^{nd} DISTR 2 to pull up normalcdf(
- Then use the range -9999, z. It should look like: normalcdf(-9999,z).