# Chapter 1

## 1.1 - Population, Samples, and Statistics

### Population

**Definition**

A set of all objects of interest in a statistical study

**Example**

The GPA of all SJSU students

### Sample

**Definition**

Any subset of a population

**Example**

The GPA of 4 random SJSU students

### Variable

**Definition**

Any characteristic whose value may change from one object to another in a statistical study

**Note**

Use uppercase letters to name variables and lowercase letters to represent actual values of the variables

**Example**

$x = 5.2(lb)$

## Discrete Variable

**Definition**

A numerical variable where its set of possible values either is finite or can be listed in an infinite sequence (one in which there is a first number, second number, and so on)

**Example**

The number of pets in a household

## Continuous Variable

**Continuous Definition**

A numerical value where its possible values consist of an entire interval on the number line

**Example**

Hair length

## Collecting Data

**Warning**

Data should be properly collected

### Sampling Techniques

- Simple Random Sampling
- Stratified Sampling
- Cluster Sampling
- Convenience Sampling

- Systematic Sampling

# 1.2 - Pictorial and Tabular Methods in Descriptive Statistics

## Stem and Leaf Plots

## Dot Plot

**Definition**

An attractive summary of numerical data when the data set is reasonably small an there are relatively few distinct values

- Each observation is represented by a dot above the corresponding location on a number line for each occurrence
- Gives information about shape and various indicators

## Distribution and Histogram for Discrete Data

**Frequency**

The number of times that the value of a discrete variable occurs in the set

**Frequency Distribution**

Lists data values along with their corresponding frequencies or counts

**Histogram**

A bar graph based on the frequency distribution of data

# 1.3 - Measures of Location

**Categorical Data**

# 1.4 - Measures of Variability

**Sample Variance**

$s^2 = \frac{\sum_{i=1}^{n}(x_1-x)^2}{n-1} = \frac{s_x x}{n-1}$ where $S_x x$ is called the sum of squares

### Sample Standard Deviation

$s = \sqrt{s^2}$

## Finding the Sample Standard Deviation Using the Definition

- Find the sample mean $x$
- Compute the deviations $(x_1 - x)$
- Square the deviation $(x_1 - x)^2$
- Add the squares of deviations $S_x x = \sum_{i=1}^{n}(x_1 - x)^2$
- Divide the result by the sample size n - 1
- Take the square root of the resulting number

### Shortcut Formula

*see notes*

### Population Variance

$\alpha^2 = \frac{\sum(x_1-\mu)^2}{N}$, where $\mu$ is the population mean and $N$ is the size of the population

### Population Standard Deviation

$\alpha = \sqrt{\alpha^2}$

### Properties of Sample Variance

1. If $y_1 = x_1 + c, y_2 = x_2 + c, \ldots y_n = x_n + c$, then $s^{2_y =} s_x^2$
2. If $y_1 = cx_1, \ldots y_n = cx_n$, then $s_y^2 = c^2 s_x^2, s_y = |c| s_x$
   Where $s_x^2$ is the sample variance of the x's ad $s_y^2$ is the sample variance of the y's

## Quartiles

- Divide an ordered data set (arranged in increasing order) into 4 groups with about 25 percent of the values in each group

- The second quartile $Q_2$ is the median of the data set
- The median of the lower half is $Q_1$ (lower fourth)
- The median of the upper half is $Q_3$ (upper fourth)
- Even observations - average the two values at each quartile split
- Odd observations - include median in both halves, the middle of each half becomes the fourth
- Interquartile Range (fourth spread)
  - $IQR$ or $f_s$
  - $IQR = f_s = Q_3 - Q_1$
- Five Number Summary
  - $Q_1, Q_2, Q_3$
  - Minimum value
  - Maximum Value
- Outliers
  - A mild outlier is if any observation is farther than $1.5f$ from the closest fourth
  - An extreme outlier is if any observation is farther than 3f from the nearest fourth

# Box Plot

1. Draw a number line
2. Plot the quartiles
3. Draw a box next to the number line that has a line at the $IQR$
4. Plot min and max
5. Draw "whiskers" from min/max (excluding outliers, if any) to box
6. Draw an asterisks to represent outlier

## Distribution Shape

- Rotate box plot $90$ degrees clockwise if vertical
- Match to shape of histogram (excluding outliers on box plot)