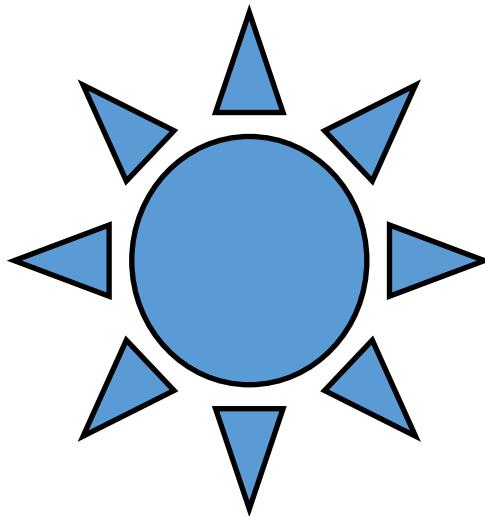


Introduction to Descriptive Statistics



By: Catherine Napier

Descriptive Statistics vs. Inferential Statistics

Descriptive statistics is the action of summarizing a sample of scores making them easier to understand whereas inferential statistics is when you draw conclusions based on the data or scores collected during research.

To be able to understand descriptive statistics, we first have to go over some vocabulary.

- ♦ **Values:** the possible number or category that a score can have.
- ♦ **Score:** a individual person's value on a certain variable.
- ♦ **Variable:** a characteristic that can have a variety of values.

Important Fact: You cannot draw conclusions from descriptive statistics. You can only say the information that is observable.

Two important differences in variables that we will see in this manual are nominal and numeric variables. **Nominal variables** are when the values are made up of categories such as spring or winter. **Numeric variables** are when the values are numbers. To remember the difference, think of nominal variables as name and numeric variables as number.



◆◆◆◆ How do we organize the data? ◆◆◆◆

There are multiple ways to organize our data but for this manual we will focus on two simple ways; Frequency tables and histograms.

Frequency Tables show how frequently a score was used and makes patterns of numbers much easier to see.

How to Make a Frequency Table

1. Make a list down the page of each possible value from lowest to highest.
2. Go through the each score, making a tally next to its matching value.
3. Make a table showing the values and how many times each value was used.
4. Figure out the percentage of scores for each value.

Example #1

Variable: Gender

Scores: (Male=1, Female=2) 2,2,1,2,2,1,1,2,2,2,2,2,1,2,1

Values: Male, Female

Table 1.

Gender

Gender	Frequency	Percent	Valid Percent	Cumulative Percent
Male	5	33.3	33.3	33.3
Female	10	66.7	66.7	100.0
Total=	15	100.0	100.0	

Example #2

Variable: Political Affiliation

Scores: (Moderate=1, I Don't Know=2, Conservative=3, Other=4, Liberal=5, Very Liberal=6) 1,2,3,1,5,1,5,2,5,6,1,6,2,2,1

Values: Moderate, I Don't Know, Conservative, Other, Liberal, Very Liberal

Table 2.

Political Affiliation

Political Affiliation	Frequency	Percent	Valid Percent	Cumulative Percent
Moderate	5	33.3	33.3	33.3
I Don't Know	4	26.7	26.7	60.0
Conservative	1	6.7	6.7	66.7
Liberal	3	20.0	20.0	86.7
Very Liberal	2	13.3	13.3	100.0
Total=	15	100.0	100.0	

Example #3

Variable: Season

Scores: (Spring=1, Summer=2, Fall=3, Winter=4)

3,1,3,1,1,1,1,3,1,3,1,4,3,3,1

Values: Spring, Summer, Fall, Winter

Table 3.

Season

Season	Frequency	Percent	Valid Percent	Cumulative Percent
Spring	8	53.3	53.3	53.3
Fall	6	40.0	40.0	93.3
Winter	1	6.7	6.7	100.0
Total=	15	100.0	100.0	

Histograms are bar-like graphs that make large groups of data even easier to see than frequency tables.

How to Make a Histogram

1. Make a frequency table.
2. Put the values along the bottom of page from lowest to highest.
3. Make a scale of frequencies on the left side of the page starting from zero to the highest frequency for a value.
4. Make a bar for each value going up to the matching frequency.

Example #1

Variables: Age

Scores: 20,31,23,24,36,27,21,25,26,29,23,28,24,24,21

Values: 20-36

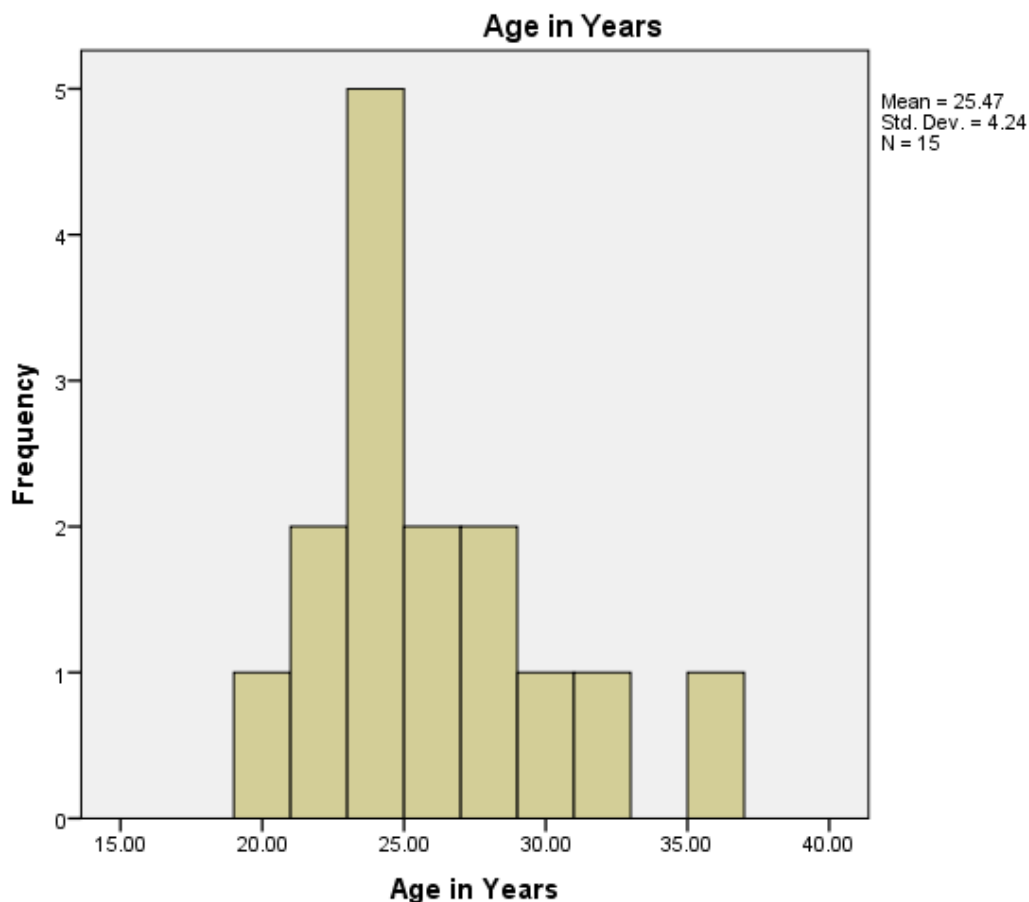


Figure 1. Age in years. Histogram based on data from PSY 230 class.

Example #2

Variable: Number of Credits Taken

Scores: 6,7,13,10,3,3,9,12,3,12,6,9,6,13,9

Values: 3-13

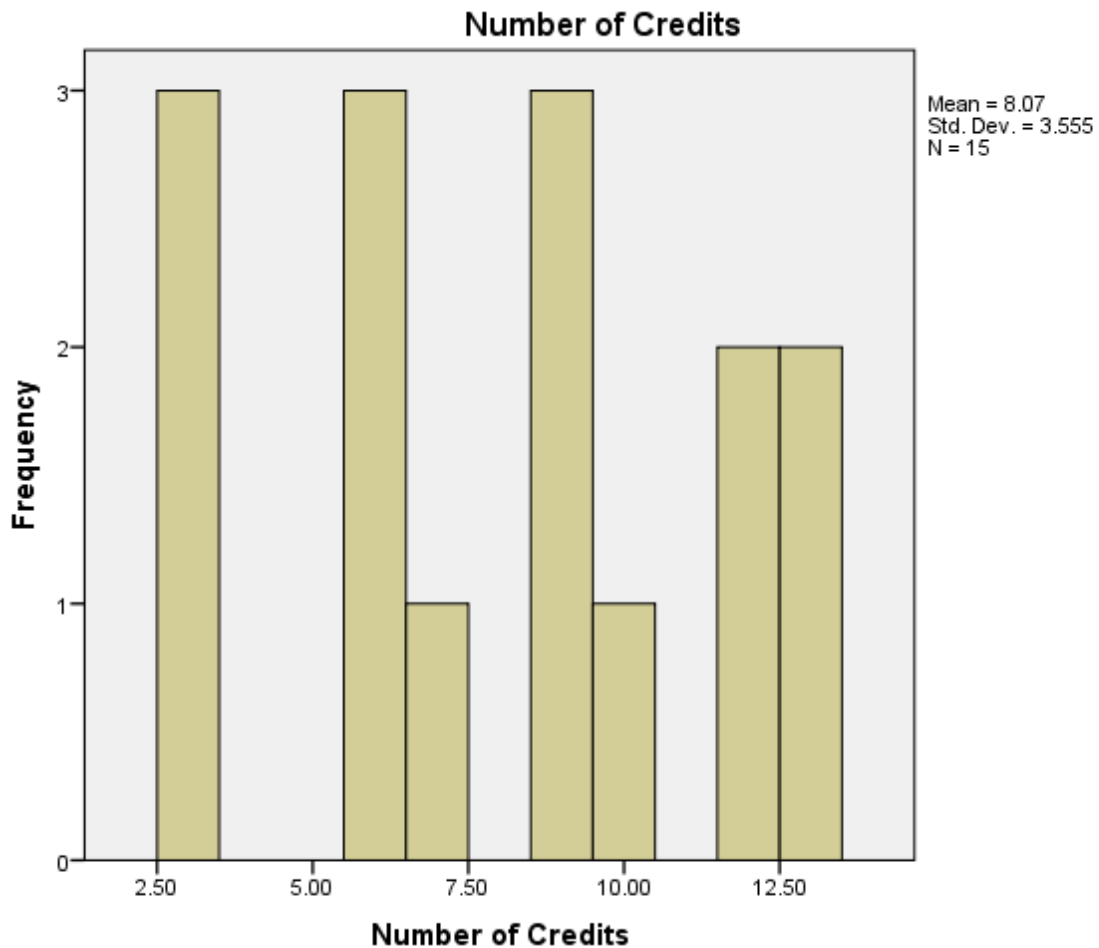


Figure 2. Number of credits. Based on data from PSY 230 class.

Example #3

Variable: Number of Romantic Relationships

Scores: 0,4,1,2,3,2,2,3,3,5,2,4,1,1,2

Values: 0-5

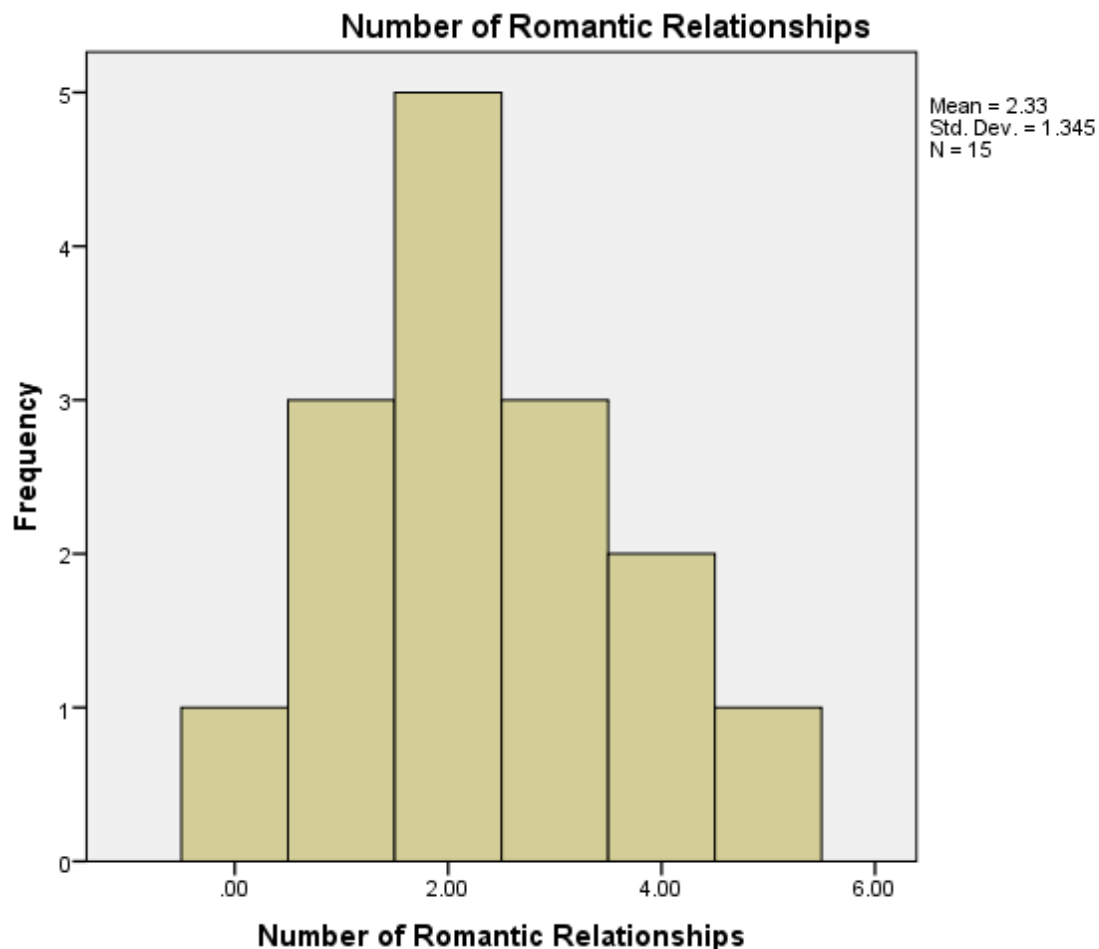


Figure 3. Romantic Relationships. Based on data from PSY 230 class.

Measures of Central Tendency

Measures of central tendency looks at the middle of a group of scores. You may have already been introduced to mean, median, and mode and those are the three measures of central tendency. Each one measures a group of scores differently but are still looking at the middle. Mean is the most commonly used measure and plays an important role in other statistical

Mean is the average of a group of scores and is found by adding up all the scores and dividing them by the total number of scores.

Example: $1+3+2+4= 10$ $10/4=2.5$

Mode is the value with the highest frequency. So all you have to do to find the mode is to find which value is used the most.

Median is the middle score when you line up the scores from lowest to highest.

Example: 1,2,3,4,5,6,7,8,9,10 Median is 5.

Variability

Variability is looking at the how spread out the scores are in a distribution. The two measures of variability are variance and standard deviation.

Variance is the average of a score's difference from the mean, squared. There are four steps to find variance.

1. Subtract the mean from each score. This makes a deviation score which is how far away the score is from the mean.
2. Square each deviation score.
3. Add all the squared deviation scores.
4. Divide the sum of the squared deviation scores by the number of scores.

$$SD^2 = \frac{\Sigma(X - \bar{X})^2}{N}$$

Where:

SD^2 = the variance

Σ = the sum of

X = the obtained score

\bar{X} = the mean score of the data

N = the number of scores

Thus, $\Sigma(X - \bar{X})^2$ = the sum of squared deviations

Standard deviation is used to describe how spread of scores and is the square root of the variance.

1. Figure out the variance.
2. Take the square root of the variance.

$$S = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N}}$$

where S = the standard deviation of a sample,

Σ means "sum of,"

X = each value in the data set,

\bar{X} = mean of all values in the data set,

N = number of values in the data set.

Now let's analyze a set of data using the knowledge and vocabulary we have learned.

Variable: Number of Romantic Relationships

Values: 0-5

Scores: 0,4,1,2,3,2,2,3,3,5,2,4,1,1,2

First make a frequency table.

Romantic	Frequency	Percent	Valid Percent	Cumulative
.00	1	6.7	6.7	6.7
1.00	3	20.0	20.0	26.7
2.00	5	33.3	33.3	60.0
3.00	3	20.0	20.0	80.0
4.00	2	13.3	13.3	93.3
5	1	6.7	6.7	100.0
Total=	15	100.0	100.0	

Based on the frequency table we can find the mode by looking at value that got the highest frequency. The mode is 2 since 2 got a frequency of 5.

Second make a histogram.

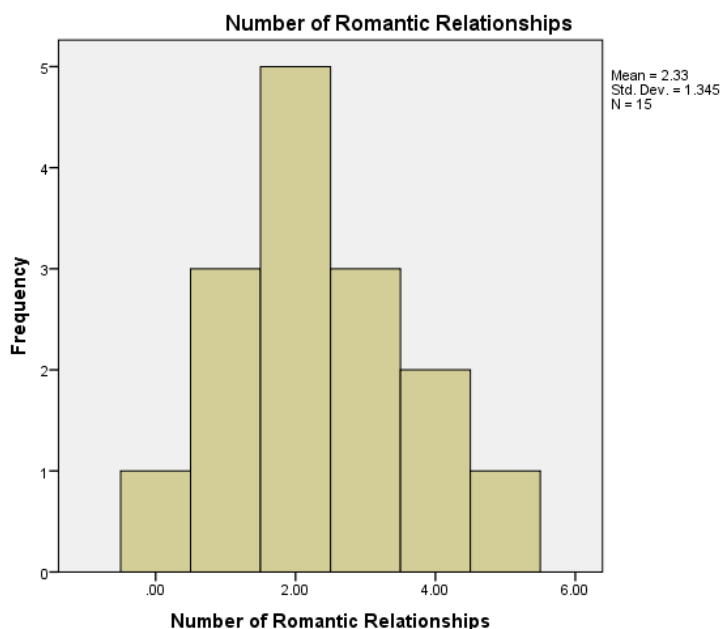


Figure 1. Romantic Relationships. Based on data from PSY 230 class.

To find the mean we have to add up all the scores then divide them by the number of scores.

$$0+4+1+2+3+2+2+3+3+5+2+4+1+1+2=37 \quad 37/15=2.46 \quad \text{Mean}=2.46$$

To find the median we have to line up the scores from lowest to highest and find the middle number.

$$0,1,1,1,2,2,2,2,2,3,3,3,4,4,5 \quad \text{Median}=2$$

Next we need to find the variance for each score using excel.

0	-2.33333	5.444444	
4	1.666667	2.777778	
1	-1.33333	1.777778	
2	-0.33333	0.111111	
3	0.666667	0.444444	
2	-0.33333	0.111111	
2	-0.33333	0.111111	
3	0.666667	0.444444	
3	0.666667	0.444444	
5	2.666667	7.111111	
2	-0.33333	0.111111	
4	1.666667	2.777778	
1	-1.33333	1.777778	
1	-1.33333	1.777778	
2	-0.33333	0.111111	
2.333333	(X-M)	(X-M)^2	1.688889
Average		25.33333	Variance
		Sum	

We will also find the standard deviation using excel.

0	-2.33333	5.444444		
4	1.666667	2.777778		
1	-1.33333	1.777778		
2	-0.33333	0.111111		
3	0.666667	0.444444		
2	-0.33333	0.111111		
2	-0.33333	0.111111		
3	0.666667	0.444444		
3	0.666667	0.444444		
5	2.666667	7.111111		
2	-0.33333	0.111111		
4	1.666667	2.777778		
1	-1.33333	1.777778		
1	-1.33333	1.777778		
2	-0.33333	0.111111		
2.333333	(X-M)	(X-M)^2	1.688889	1.299572579
Average		25.33333	Variance	Standard Deviation
		Sum		

Final Analysis of Data: With the information we collected, we can see that the mean, median, and mode were all pretty close around the number 2 with the mean being 2.33 and both the median and mode at 2. The data wasn't extremely spread out with a variance of 1.68 and a standard deviation of 1.29. Since this is an example of descriptive statistics, we cannot draw any conclusions about these numbers or the data.