Descriptive Statistics and Measurement Scales

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Descriptive statistics are typically distinguished from inferential statistics. With descriptive statistics you are simply describing what the data shows. With inferential statistics, you are trying to reach conclusions that extend beyond the immediate data alone. For instance, we use inferential statistics to make judgments of the probability that an observed difference between groups is a dependable one and is unlikely to have happened by chance in the study. Thus, we use inferential statistics to make inferences from our data to more general conditions; we use descriptive statistics simply to describe what's going on in our data.

One way to summarize data is to report the number of people who obtained given scores on a scale. In such a case we would be tallying the responses and reporting frequencies. In some cases, frequency is more meaningfully expressed as percentages or as proportions. For example, if I tell you that last semester 20 of my students scored an A or higher on the first exam I am giving you the frequency of A grades in that class. This frequency would indicate very different things if my class were comprised of 25 students compared to 200 students. In one case 80% of the students earned As; in the other case only 10% earned As. Percentages and proportions adjust a frequency and report it as though the sample size is 100 or 1 respectively.

Another way of summarizing data involves reporting a measure of central tendency. Measures of central tendency define the most representative value of the data set. There are three common ways of describing central tendency, Mean, Median and Mode. Each of these measures uses different criteria for determining the most representative score. Which measure should be used depends on the nature of the measurement scale being reported and the meaning that you wish to convey. A second group of descriptive statistic that is useful in psychology is the measures of dispersion. These include range, variance and standard deviation.

Measures of Central Tendency

Measures of central tendency, attempt to quantify the "typical" or "average" score in a data set. The concept is extremely important and we encounter it frequently in daily life. For example, we often want to know before purchasing a car its average gas mileage. Or before accepting a job, you might want to know what a typical salary is for people in that position so you will know whether or not you are going to be paid what you are worth. Statistics geared toward measuring central tendency all focus on this concept of "typical" or "average." As we will see, we often ask questions in psychological science revolving around how groups differ from each other "on average". Answers to such a question tell us a lot about the phenomenon or process we are studying.

Mode. By far the simplest, but also the least widely used, measure of central tendency is the mode. The mode in a distribution of data is simply the score that occurs most frequently. One way of describing a distribution is in terms of the number of modes in the data. A unimodal distribution has one mode. In contrast, a bimodal distribution has two.
The biggest weakness of a mode is that it is not very stable. Small changes in the distribution can produce large changes in the mode. For example, if in this class 20 students have an A; 19 students have a B; 13 students have a C and 3 students have less than a C, the modal grade would be A (yeah!!!!!). If one student who previously held an A, slipped into the B category, this would change the mode from A to B. Change in one data point can produce large changes in this measure of central tendency. The biggest strength of this measure, on the other hand, is that it can be used with both categorical and continuous measures (see below).

**Median.** Technically, the **median** of a distribution is the value that cuts the distribution exactly in half, such that an equal number of scores are larger than that value as there are smaller than that value. The median is by definition what we call the 50th percentile. The median is most easily computed by arranging the data set from smallest to largest. The median is the "middle" score in the distribution. Suppose we have the following scores in a data set: 5, 7, 6, 1, 8. Arranging the data, we have: 1, 5, 6, 7, 8. The "middle score" is 6, so the median is 6. Half of the (remaining) scores are larger than 6 and half of the (remaining) scores are smaller than 6. The biggest strength of this measure is that it is not sensitive to "outliers" (extreme scores). For example, compare the following distribution of scores

Set A  2,5,6,6,7,8,9   Median = 6

Set B  2,5,6,6,7,8,100    Median = 6.

The limitation of this measure is that it can only be used with continuous scales (see below).

**Mean.** The **mean** is the most widely used measure of central tendency. The mean is defined as the arithmetic average (sum of all the data scores divided by the number of scores in the distribution). In a sample, we often symbolize the mean with a letter with a line over it. If the letter is "X", then the mean is symbolized as $\overline{X}$, (pronounced "X-bar"). If we use the letter X to represent the variable being measured, symbolically the mean is defined as

$$\overline{X} = \sum_{n}^{X}$$

Where $\Sigma$ indicates sum of and n represents the number of data points in the set.

The mean can be used only with continuous scales of measurement. A second problem with the mean is that it is affected by outliers. Looking at the two distributions used above, the mean of set A is 6.14 whereas the mean of Set B is 19.14. This difference in the mean is produced by a difference a single extreme score.
Measurement Scales

**Categorical (Discrete) Variables**

Categorical variables are defined as unique, clear-cut grouping of outcomes. When we define a variable in categorical terms we treat each outcome which falls into a category as an equal outcome. We might categorize people into categories such as sex. A person is either categorized as male or female. There are no middle values between the categories and there are no distinctions between outcomes included within a category. Categorical variables have the advantage of allowing us to describe our outcomes according to an organizing structure with clearly defined criteria for falling into a category. Nominal and Ordinal scales are by definition categorical variables.

**Nominal Scales.** Numbers are often used to label categories which differ in type or kind. The magnitudes of the numbers are not meant to indicate that one category is better than another. No rank ordering of categories is intended. A common example of a nominal category is sex. We might enter the number one to indicate that a subject is male and the number two to indicate that a subject is female. The number should only be considered a label (name) and should not be viewed as a value. The only measurement property this scale has is that of difference (i.e., it defines that different numbers on the scale indicate different qualities of the property being measured.

We might think of nominal scales as the weakest form of measurement. By this we mean that there is a limited number meaningful ways of summarizing the results. For example, I might ask 120 people to indicate their favorite color. When I enter my data into my data file for analysis can differentiate between color categories by assigning them each a different number. The number label however is meant simply as a name of the category and should not be thought of as having properties we usually associate with numbers. The table below represents the results of my survey of a group of 120 participants who were asked to report their favorite color.

<table>
<thead>
<tr>
<th>Color</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Blue</td>
<td>45</td>
</tr>
<tr>
<td>2. Green</td>
<td>28</td>
</tr>
<tr>
<td>3. Yellow</td>
<td>25</td>
</tr>
<tr>
<td>4. Red</td>
<td>10</td>
</tr>
<tr>
<td>5. Other</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>120</td>
</tr>
</tbody>
</table>

Because I have assigned numerical values to these categories it would be possible to calculate the mean of these scores. If I totaled each score and divided by the total number of scores I could report the mean as 2.3, but this would not be an accurate summary of the data because the assignment of numbers to categories is arbitrary and if I had assigned the numbers to the categories in a different order the mean would be different. Further, I could put all of the numbers in order and determine the median of the scores. In this case it would be 2, but again had I assigned the numbers differently the median would be different. The only measure of central tendency which is meaningful for nominal data is the mode. In this case the modal response was one, and from that I can say that the most frequently given response is Blue.
**Ordinal Scales.** Numbers in this case are meant to indicate an order to the levels of the variable being measured. For example, education level could be measured on an ordinal scale. We might enter a zero to indicate that a subject has not completed a high school degree, a one to indicate that a person’s highest level of education is a high school degree, a two to indicate an associate degree, a three to indicate a B.A. or B.Sc, a four to indicate an MA or M.Sc, and a five to indicate a Doctorate. In this case, the higher the number is the more education a person has. So an Ordinal scale defines not only a difference between categories but also assigns an order to the variables. The intervals, however, on an Ordinal scale are not equal. By this I mean that you could not logically argue that a MA (labeled 4), is as much greater than a BA (labeled 3), as a Doctorate (labeled 5) is greater than an MA. Because the numbers on this scale do not represent equal intervals the **mode** is the only measure of central tendency that can meaningfully be used.

**Continuous scales**

Continuous scales do not limit score to defined categories. On a continuous scale differences in numbers indicate both order and changes that are of equal intervals. An example of the difference between a continuous and a categorical score is percentage exam scores compared to letter grades. A percentile score can take on any value between zero and 100. Letter grades can take on only one of 5 values. A, B, C, D of F. Continuous scales have the advantage of providing more precise information. For example, a student who obtains a 65 would be categorized in the same group as a student who obtained a 74. Continuous scales can always be translated into categorical scores simply by defining boundaries for categories. When we do this, however, we lose precision.

Because continuous scales have the properties of difference, order and equal intervals we can meaningfully calculate means and determine medians. Although modes can be determined for continuous variables the large range of values that data can take on often produces a mode which is not really a summary of the distribution. For example, when recording exam scores for 25 students, the modal score of the distribution might simply reflect a chance coincidence of two students obtaining the same grade.

**Interval and Ratio scales** are two types of continuous scales:

**Interval Scales.** Numbers on this scales indicate differences and order, however, they also have equal intervals. The magnitude of the categories indicates difference, order and degree of difference. An example of an interval scale is temperature measured in either Fahrenheit or Celsius. In this case we can say that 50 degrees is as much higher than 25 degrees as 75 degrees is higher than 50 degrees. Thus, the intervals indicated by the numbers are equal. An interval scale, however, does not have an absolute zero. In our example, zero does not indicate an absence of temperature; therefore, we cannot say that 50 degrees is twice as hot as 25 degrees. Since these scales have equal intervals, they are continuous in nature (rather than discrete categories and therefore the mean median and mode may all be used as measures of central tendency.

**Ratio Scales.** Numbers on this scale indicate differences, order, equal intervals and also have an absolute or naturally falling zero point. For example, time measured in minutes is measured on a ratio scale. Zero indicates no time. The importance of an absolute zero point is that it allows us to logically state not only that the difference between a person who finishes an exam in 5 minutes is as much faster that a person who takes 10 minute as a person who takes 10 minutes is faster than a person who takes 15 minutes to complete the task, but we can also say that a person who takes 5 minutes to complete the task completed it twice as fast as a person who took 10 minutes. In other words, we can discuss the differences in scores as ratios. This scale
also is continuous in nature and therefore central tendency can be defined as Mean, Median or Mode.

**Measures of Dispersion**

The central score in a distribution is important in many research contexts. So too is another set of statistics that quantify how spread out (or "how dispersed") the scores tend to be. Do the scores vary a lot, or do they tend to be very similar or near each other in value?

The simplest measure of dispersion is the **range**. The range defines the difference between the largest and smallest score in the data set. One advantage of this measure is that it can be used with any measurement scale that has the property of **order** (e.g., continuous, interval and ratio scales). A weakness is that it is strongly affected by extreme scores. For example, Set A and Set B have hugely different ranges even though they differ by only one score.

The two most commonly used measures of dispersion; **Variance** ($s^2$) and **Standard Deviations** ($S$) are calculated by finding the average amount that the scores in a distribution differ from the mean. One might think that the way to calculate this is to take each score and minus the mean (calculate deviation scores), then add these up and divide by the number of scores. One thing that is important to notice is that the mean deviation from the mean is 0. This will always be the case because, by definition the mean is the point at which the sum of scores above the mean and the sum of scores below the mean are exactly equal to each other. The negative numbers thus cancel out the positive numbers.

We use a very simple mathematical trick to get around this in calculating **variance**. We simply square all the deviation scores and add them up also referred to as sum of squares). Recall that the square of any number is a positive number; therefore, the sum of all the squared deviation scores will be a positive number. We simply divide by the number of scores to get the average squared deviation from the mean. Since most of us do not think in terms of squared distances, the most commonly use statistic for describing dispersion is the square root of variance. This statistic is called the **standard deviation** and it is **simply the average distance of score from the mean of the distribution**. The larger the standard deviation, the more spread out the distribution is.
Shape of the Distribution

Another important thing to know about a distribution is whether it is **symmetric or skewed**. A distribution curve is symmetric if when folded in half the two sides match up. If a curve is not symmetric it is skewed. When a curve is positively skewed, most of the scores occur at the lower values of the horizontal axis, and the curve tails off towards the higher end. When a curve is negatively skewed, most of the scores occur at the higher value and the curve tails off towards the lower end of the horizontal axis. Note that the shape of the distribution is names after the tail. A negatively skewed distribution has its mean pulled down by the extreme scores, whereas a positively skewed distribution has its mean pulled up by the extreme scores.

| Skewed Positively | Skewed Negatively |
|-------------------|--|---|
| ![Skewed Positively](image1.png) | ![Skewed Negitively](image2.png) |

If a distribution is a unimodal symmetrical curve, the mean, median and mode of the distribution will all be the same value. When the distribution is skewed the mean and median will not be equal. Since the mean is most affected by extreme scores, it will have a value closer to the extreme scores than will the median.

Why does the measure of Central Tendency Reported matter??

Consider a country so small that its entire population consists of a queen (Queen Cori) and four subjects. Their annual incomes are

<table>
<thead>
<tr>
<th>Citizen</th>
<th>Annual Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queen Cori</td>
<td>1,000,000</td>
</tr>
<tr>
<td>Subject 1</td>
<td>5,000</td>
</tr>
<tr>
<td>Subject 2</td>
<td>4,000</td>
</tr>
<tr>
<td>Subject 3</td>
<td>4,000</td>
</tr>
<tr>
<td>Subject 4</td>
<td>2,000</td>
</tr>
</tbody>
</table>

Queen Cori might boast that this is a fantastic country with an “average” annual income of $203,000. Before rushing off to become a citizen, you might want to be wise and find out what measure of central tendency Queen Cori is using! The mean is $203,000, so she is not lying, but this is not a very accurate representation of the incomes of the population. Money is a continuous (ratio) variable, but in this case the median ($4,000.00) or the mode (also $4,000.00) would be a more reprehensive value of the “average” income. The point to be made here is that the appropriate measure of central tendency is affected not only by the type of measurement but also by the distribution of income. The mean of a distribution is strongly affected by extreme scores.
Standard Normal Curve

The frequency polygon which is most useful in psychology is the **normal curve**. Many measures of human ability, attitudes and behaviors are **normally distributed in the population**. When we say that they are normally distributed we mean:

1. They are symmetrically distributed so that mean = median = mode.
2. The **Standard Normal Curve** is a normal curve with a **mean of zero** and a **standard deviation of one**. This curve can be used as a measuring device that allows us to determine the percentage of scores that fall at or between any two values in any normal distribution.
3. The Standard Normal Curve is **asymptotic**, which means that its tails approximate (approach) but never actually reach zero. This is the mathematicians’ way of recognizing that in an infinite universe (such as ours) – NOTHING is impossible. This is important because when we use this curve to determine probabilities, you can never measure from the end – you must always measure from the middle (the mean).
4. Mathematicians have been able to show that in a normal distribution, exactly 68.26% of scores fall between one standard deviation above and below the mean. You do not need to know why this true, it is just the magic of the standard deviation that it can be used as a measuring unit. In fact, if we know the mean and standard deviation of a normally distributed variable, we can determine the percentage of scores that fall above or below that point by determining how many standard deviations away from the mean the score is. This is called a **z-score**. The X (horizontal) axis of a **standard normal curve** represents z-scores.