**Lesson 3**
**Data Displays**

Outline
Frequency Distributions
Grouped Frequency Distributions
-class interval and frequency
-cumulative frequency
-relative percent
-cumulative relative percent
-interpretations
Histograms/Bar Graphs

## Frequency Distributions

We often form frequency distributions as a way to abbreviate the values we are dealing with in a distribution. With frequency distributions we will simply record the frequency or how many values fall at a particular point on the scale.

For example, if I record the number of trips out of town (X) a sample of FSU students makes, I might end up with the following data:

0 2 5 3 2 4 3 1 0 2 6 0 4 7 0 1 2 4 3 5 4 3 1 6 1 0 5 3

Instead of having a jumbled set of numbers, we can record how many of each value (f) there are for the entire x-distribution. Below is a simple frequency distribution where the X column represents the number of trips, and the corresponding value for "f" indicates how many people in the sample gave us that particular response.

| X | f |
|---|---|
| 0 | 5 |
| 1 | 4 |
| 2 | 4 |
| 3 | 5 |
| 4 | 4 |
| 5 | 3 |
| 6 | 2 |
| 7 | 1 |

From the graph we can see that five people took no trips out of town, four people took one trip out of town, four people took two trips out of town, and so on.

It is important not to confuse the f-value and the x-value. The f-values are just a count of how many. So, you can reverse the process as well. It might also be helpful in some examples to go from a frequency distribution back to original data set, especially if it causes confusion.

In the following example I start with a frequency distribution and go backward to find all the original values in the distribution.

| X | f |
|---|---|
| 0 | 2 |
| 1 | 3 |
| 2 | 4 |
| 3 | 3 |
| 4 | 2 |

What is the most frequent score? The answer is two because we will have four ‘twos’ in our distribution:

0 0 1 1 1 2 2 2 2 3 3 3 4 4

**Grouped Frequency Distributions**

The above examples used discreet measures, but when we measure a variable it is often on a continuous scale. In turn, there will be few values we measure that are at the exact same point on the scale. In order to build the frequency distribution we will group several values on the scale together and count any of measurements we observe in that range for the frequency.

For example, if we measure the running time of rats in a maze we might obtain the following data. Notice that if I tried to count how many values fall at any single point on the scale my frequencies will all be one.

3.25 3.95 4.61 5.92 6.87 7.12 7.58 8.25 8.69 9.56 9.67 10.24 10.95 10.99 11.34

11.59 12.34 13.45 14.13 14.46

We will begin by forming the class interval. This will be the range of value on the scale we include for each interval. There are many rules we could use to determine the size of the interval, but for this course I will always indicate how big the interval should be. In the end, we want to construct a display that has between 5 and 15 intervals. Thus:

| Class Interval |
|---|
| 0-2 |
| 3-5 |
| 6-8 |
| 9-11 |
| 12-14 |

Once we have the class interval, we will count how many values fall within the range of each interval. Since there is a gab in each class interval, we will be actually counting any values that would get rounded down or up into a particular interval. For example, with the above data the value 8.26 would be rounded down into the 6-8 class interval. The value 8.69 would be rounded up into the 9-11 class interval. We will include a column to indicate the real limits of the class interval. These are the limits of the interval, including any rounded values.

| Real Limits | Class Interval | f |
|---|---|---|
| -.5-2.5 | 0-2 | 0 |
| 2.5-5.5 | 3-5 | 3 |
| 5.5-8.5 | 6-8 | 5 |
| 8.5-11.5 | 9-11 | 7 |
| 11.5-14.5 | 12-14 | 5 |

Notice that my real limits cover half the distance of the gap between each class interval. Most of the time this value will be 0.5 since most scales will have one unit values and 0.5 is half the distance. So, real limits have no gap, but the class intervals do. If a value falls exactly on one of the real limits we could randomly choose its group.

Cumulative Frequency

Once we have formed the basic grouped frequency distribution above, we can add more columns for more detailed information. The first of these is the cumulative frequency column. With this column we will keep a running count of the frequency column as we move down the class interval.

| Real Limits | Class Interval | f | Cum. f |
|---|---|---|---|
| -.5-2.5 | 0-2 | 0 | 0 |
| 2.5-5.5 | 3-5 | 3 | 3 |
| 5.5-8.5 | 6-8 | 5 | 8 |
| 8.5-11.5 | 9-11 | 7 | 15 |
| 11.5-14.5 | 12-14 | 5 | 20 |

So, at the first interval we have zero frequency, so cumulatively we have zero values. For the second interval we have three, so cumulatively we have three. For the third interval we have five values, so cumulatively we have 8. That includes the five for the third interval, plus the three from the previous intervals. We continue this process until the last interval. Notice that when we reach the last interval we have all the values in the distribution represented. So, the bottom cumulative frequency is N or the total number of values in the distribution (20 here).

Relative Percent

Another column will tell us the proportion of total values that fall at each interval. That is, we will express the frequency (column) as a percentage of the total. To convert the frequency to a percentage take the frequency (f) and divide by the number of values (N). This will give us the proportion of values for that particular interval. Move the decimal over two places (or multiply by 100) to change the proportion into a percent. Thus:

| Real Limits | Class Interval | f | Cum. f | Rel % |
|---|---|---|---|---|
| -.5-2.5 | 0-2 | 0 | 0 | 0 |
| 2.5-5.5 | 3-5 | 3 | 3 | 15 |
| 5.5-8.5 | 6-8 | 5 | 8 | 25 |
| 8.5-11.5 | 9-11 | 7 | 15 | 35 |
| 11.5-14.5 | 12-14 | 5 | 20 | 25 |

Cumulative Relative Percent

For a final column we will keep a running count of the relative percent column in the same way we did with the cumulative frequency. Keep in mind we are counting relative percentages now as we move down the display.

| Real Limits | Class Interval | f | Cum. f | Rel % | Cum. Rel. % |
|---|---|---|---|---|---|
| -.5-2.5 | 0-2 | 0 | 0 | 0 | 0 |
| 2.5-5.5 | 3-5 | 3 | 3 | 15 | 15 |
| 5.5-8.5 | 6-8 | 5 | 8 | 25 | 40 |
| 8.5-11.5 | 9-11 | 7 | 15 | 35 | 75 |
| 11.5-14.5 | 12-14 | 5 | 20 | 25 | 100 |

Notice that we can keep a running count of the relative percent column, but we could also obtain the same numbers by computing the percentage for each cumulative frequency as well.

Interpretations

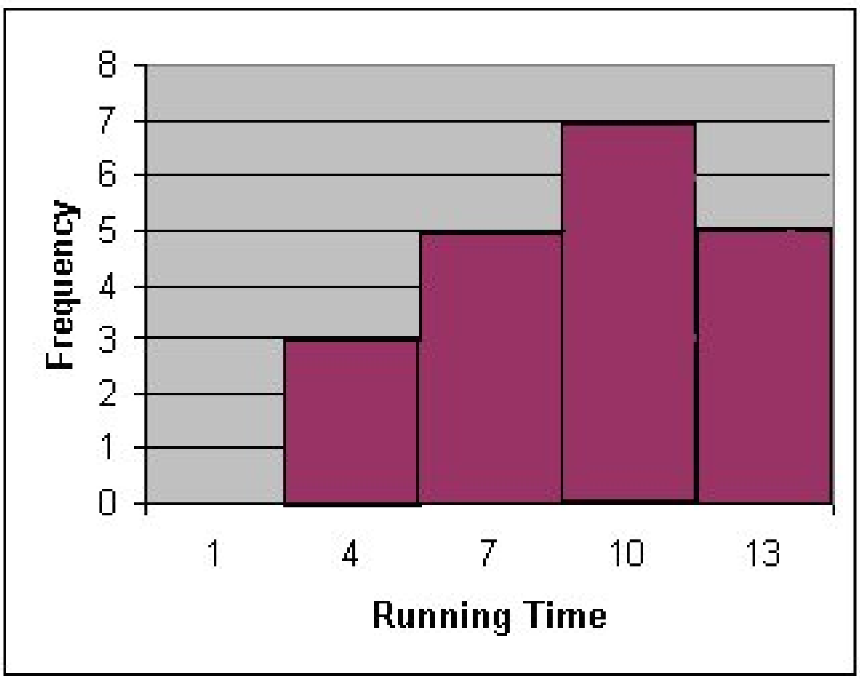
The data display gives a good deal of information about where values in the sample fall. One good piece of information is about percentiles. A percentile is the percentage at or below a certain score. You often get percentile information when you get your SAT or ACT test scores back. Percentile information is found in the cumulative relative percentage column. Each value in that column tells us the percentage of the distribution at that point or less on the scale. Since we will be rounding values down into a certain interval based on the real limits, then we will indicate where on the scale a certain percentile is based on its corresponding upper real limit. For example, what score corresponds with the 75$^{th}$ percentile? The answer is 11.5 because any values of 11.5 or less are within the bottom 75% of the distribution. Similarly, what percentile is associated with a score of 8.5? We would use the cumulative relative percent that corresponds to 8.5, which is 40%. So, the score 8.5 corresponds with the bottom 40% or 40$^{th}$ percentile of the distribution.

Other interpretations from the table can be made as well. For example, we might be interested in how many people fall at a particular interval, or at or below a certain interval. How many scored between 3 and 5? The answer is a found in the frequency column, or three. How many scored 8.5 or less? The answer for this question is in the cumulative frequency column, or eight.

**Histograms/Bar Graphs**

We can also take the frequency information in our frequency or grouped frequency distribution and form a graph. In the graph we will form a simple x-y axis. On the x-axis we will place values from our scale, and on the y-axis we will plot the frequency for each point on the scale. For grouped frequency distributions, we will use the midpoint of each interval to indicate different points on the scale. We will continue with our previous example, but notice I have created a new column that indicates the center or midpoint of each interval. We will use this value to graph the display.

| Real Limits | Class Interval | MP | f | Cum. f | Rel % | Cum. Rel. % |
|---|---|---|---|---|---|---|
| -.5-2.5 | 0-2 | 1 | 0 | 0 | 0 | 0 |
| 2.5-5.5 | 3-5 | 4 | 3 | 3 | 15 | 15 |
| 5.5-8.5 | 6-8 | 7 | 5 | 8 | 25 | 40 |
| 8.5-11.5 | 9-11 | 10 | 7 | 15 | 35 | 75 |
| 11.5-14.5 | 12-14 | 13 | 5 | 20 | 25 | 100 |




Note that the bars are touching. The bars touch like this when we are dealing with continuous data rather than discreet data. When the scale measures discreet values we call it a bar graph, and the lines do not touch. For example, if I measured the number of democrats, republicans, and independents in a sample, we would use a bar graph if we wanted to create a data display.