1.3 Frequency, Frequency Tables, and Levels of Measurement

Levels of Measurement

Definition	Examples Characteristic (Categories)
• The ***nominal level of measurement*** is the classification of data into mutually exclusive (nonoverlapping), exhausting categories in which no order or ranking can be imposed on the data.	• Eye color (brown, blue, green, hazel, other) • Gender (male, female, other, decline to state) • Sport • Ethnicity
• The ***ordinal level of measurement*** is the classification of data into categories that can be ranked; however, precise differences between the ranks do not exist.	• Rating of instructor (poor, average, excellent) • Competition results (1^{st}, 2^{nd}, 3^{rd}) • Quality of health (poor, average, good) • Income bracket (low, middle, high)
• The ***interval level of measurement*** is the classification of data into ordered ranks with precise units of measure existing, however, there is no meaningful zero.	• Temperature in degrees Fahrenheit. • SAT scores • IQ scores • Dates
• The ***ratio level of measurement*** is the classification of data into ordered ranks that include precise units of measure and a meaningful zero.	• Height • Weight • Income • Time

Summary of Levels of Measurement Characteristics

	nominal	***ordinal***	***interval***	***ratio***
• data is classified into nonoverlapping, exhausting categories	x	x	x	x
• ordered		x	x	x
• includes precise units of measure • addition and subtraction of data are meaningful			x	x
• includes a meaningful zero • multiplication and division of data are meaningful				x

Illustrative Examples of the 4 Levels of Measurement

I. The first (and weakest) level of data is called **nominal** level data. There is no standard ordering scheme to this data.

Ex. The colors of M&M candies is an example of nominal level data. This data is distinguished by name only. There is no agreed upon ordering of this data, although we each may have an opinion about which should be listed first. I may be partial to brown and feel that brown should always be listed first, but you may like blue and feel it should go first.

II. The **ordinal** level data is also distinguished by name, but there is an ordering scheme.

Ex. Movies may be classified as 2 thumbs up, 1 thumb up, or 0 thumbs up. There is an order here. A movie that receives 2 thumbs up is better that a movie that receives 1 thumb up (supposedly anyway). How much better is a movie that receives 2 thumbs up than a movie that receives 1 thumb up? Is it 1 thumb better? What exactly does that mean?

Ex. Voters are classified as low-income, middle-income, or high-income. This is an example of ordinal level data. We do know that people in the low-income bracket earn less than the people in the middle-income bracket, who in turn earn less than the people in the high-income bracket. So there is an ordering scheme to this data. The thing that ordinal level data lacks is that you can't measure the difference between two pieces of data. We know that high-income people earn more than low-income people, but how much more.

III. The **interval** level data has a definite ordering scheme, but the differences between data is meaningful and can be measured.

Ex. The boiling temperatures of different liquids are listed. This is an example of interval level data. We can tell whether a temperature is higher or lower than another, so we can put them in an order. Also, if water boils at 212 degrees and another liquid boils at 284 degrees, the second temperature is 72 degrees higher than the first. So the differences between data are measurable and meaningful.
The one thing that interval data lacks is a zero starting point. Is 0 degrees the absolute lowest temperature? Because there is no zero starting point, ratios between 2 data values are meaningless. Is 75 degrees three times as hot as 25 degrees? No, because the ratio of 75 to 25 (i.e. 3 to 1) is meaningless here. Think about the following cooking example.

Ex. A brownie recipe calls for the brownies to be cooked at 400 degrees for 30 minutes. Would the results be the same if you cooked them at 200 degrees for 60 minutes? How about at 800 degrees for 15 minutes? I think we would get 3 different types of brownies : just right, awful gooey, and awful crunchy. The problem is that 200 degrees is not half as hot as 400 degrees, and 800 degrees is not twice as hot as 400 degrees.

IV. The highest level of measurement is the **ratio** level data which is just like interval level data, except that ratios make sense.

Ex. Four people are randomly selected and asked how much money they have with them. Here are the results : \$21, \$50, \$65, and \$300.
Is there an order to this data? Yes, \$21 < \$50 < \$65 < \$300.
Are the differences between the data values meaningful? Sure, the person who has \$50 has \$29 more than the person with \$21.
Can we calculate ratios based on this data? Yes because \$0 is the absolute minimum amount of money a person could have with them. The person with \$300 has 6 times as much as the person with \$50.

Demonstration Problems	*Practice Problems*
Classify each as nominal, ordinal, interval, or ratio levels of measurement	Classify each as nominal, ordinal, interval, or ratio levels of measurement
1. (a) the year of each of the Olympic games	**1. (b)** the temperature at which butter melts
2. (a) customer satisfaction with smartphone service (very, somewhat, not at all)	**2. (b)** income bracket of patients at a free health clinic (low, middle, high)
3. (a) position played by a member of the USA soccer team	**3. (b)** smartphone service provider
4. (a) the number of hours a student spends studying for an exam	**4. (b)** the number of pounds a newborn elephant weighs
Answers: **1. (b)** interval **2. (b)** ordinal; **3. (b)** nominal; **4. (b)** ration	

Summary of Characteristics of Data

Qualitative		**Quantitative**		
Nominal	***Ordinal***	***Interval***	***Ratio***	
			Discrete	*Continuous*
• Eye color • Gender • Sport • Ethnicity	• Rating of instructor • Competition results • Quality of health • Income bracket	• Temperature in degrees Fahrenheit. • SAT scores • IQ scores • Dates	• Number of children in a family • Number of students in a classroom	• Height • Weight • Income • Time

Frequency Tables

Study the list of ages at the inaugurations of the 44 U. S. presidents. Look for generalizations that you can make about the ages in the list.

Ages of U. S. Presidents at Inauguration

		Age			Age
1	George Washington	57	23	Benjamin Harrison	55
2	John Adams	61	24	Grover Cleveland	55
3	Thomas Jefferson	57	25	William McKinley	54
4	James Madison	57	26	Theodore Roosevelt	42
5	James Monroe	58	27	William Taft	51
6	John Quincy Adams	57	28	Woodrow Wilson	56
7	Andrew Jackson	61	29	Warren Harding	55
8	Martin Van Buren	54	30	Calvin Coolidge	51
9	William Harrison	68	31	Herbert Hoover	54
10	John Tyler	51	32	Franklin Roosevelt	42
11	James Polk	49	33	Harry Truman	60
12	Zachary Taylor	64	34	Dwight Eisenhower	62
13	Millard Fillmore	50	35	John Kennedy	43
14	Franklin Pierce	48	36	Lyndon Johnson	55
15	James Buchanan	65	37	Richard Nixon	56
16	Abraham Lincoln	52	38	Gerald Ford	61
17	Andrew Johnson	56	39	Jimmy Carter	52
18	Ulysses Grant	46	40	Ronald Reagan	69
19	Rutherford Hayes	54	41	George H.W. Bush	64
20	James Garfield	49	42	William Clinton	46
21	Chester Arthur	51	43	George W. Bush	54
22	Grover Cleveland	47	44	Barack Obama	47
			45	Donald Trump	70

And now study the following frequency table for generalizations of their ages. It is clear that organization of data produces much more useful information at a glance than raw data.

Age at Inauguration of U. S. Presidents	Tally	Frequency
40 - 44	xxx	3
45 – 49	xxxxxxx	7
50 - 54	xxxxxxxxxxx	12
55 - 59	xxxxxxxxxxx	12
60 – 64	xxxxxxx	7
65 - 69	xxx	3
70 - 74	x	1

We will see in this section how to organize raw data into a meaningful format called a frequency distribution.

A ***frequency distribution*** is a grouping of data into non-overlapping classes showing the ***frequency*** (number of observations) of data in each class. Other data may be calculated for each class such as *percent* (the percent of all data that lies in each class), *relative frequency* (the proportion of all data that lies in each class, or *cumulative frequency* (the cumulative total of data that lies in each successive class).

Examples

A ***categorical frequency distribution*** is a distribution in which the data is strictly categorical

Blood type of 25 army inductees	Tally	Frequency	Percent
A	xxxxx	5	20
B	xxxxxxx	7	
O	xxxxxxxxx		
AB	xxxx		

A ***grouped frequency distribution*** is a frequency distribution that categorizes data by intervals.

Number of miles 50 surveyed students drive to school	Tally	Frequency	Relative frequency
1–3	xxxxxxxxxx	10	0.20
4–6	xxxxxxxxxxxxx	14	
7–9	xxxxxxxxxx		
10–12	xxxxxx		
13–15	xxxxx		
16-18	xxxxx		

An ***ungrouped frequency distribution*** is a frequency distribution that categorizes data by single units.

Number of cars registered in each of 20 households surveyed	Tally	Frequency	Cumulative frequency
0	xxxx	4	4
1	xxxxxx	6	10
2	xxxxx	5	
3	xxx		
4	xx		

Categorical Frequency Distribution

The categorical frequency distribution is used for data that can be placed in specific categories, such as nominal level or ordinal level data. Make a frequency distribution for the results of a survey of 10 people of their overall health.

poor	average
average	excellent
excellent	average
excellent	poor
excellent	excellent

Health (Class)	Tally	Frequency	Relative Frequency	Cumulative Frequency

Grouped Frequency Distribution

Vocabulary

A ***class*** is a category into which data is distributed.

Class limits are the lower and upper limits of each class in a grouped frequency distribution.

Class boundaries are an extension of class limits that fills the gaps occurring between the upper limit of a class and the lower limit of the next class higher.

Class width is the difference between the lower limits of successive classes.

A ***class midpoint*** is the average of the lower and upper limits of each class.

In a grouped frequency distribution the data is categorized into intervals determined by ***class limits***. In our example on the preceding page of the number of miles each student surveyed drives to school, where would we place a tally mark for a distance of 6.2 miles? There are gaps between the class limits, so we establish ***class boundaries*** to fill these gaps. Class boundaries have one decimal place more than the values of the data set.

Let's find the class boundaries of the example below.

1. The first class lower boundary is 0.5 less than the first class lower limit.
2. The next class boundaries can be determined by averaging each class upper limit with the next higher class lower limit.
3. And the final class upper boundary is 0.5 higher than the final class upper limit.

Number of miles 45 surveyed students drive to school (Class limits)	Number of miles (Class boundaries)	Tally	Frequency	Cumulative frequency
1–3		xxxxxxxxxx	10	10
4–6		xxxxxxxxxxxxxx	14	24
7–9		xxxxxxxxxx	10	34
10–12		xxxxxx	6	40
13–15		xxxxx	5	45

In the construction of grouped frequency distributions in general,

1. There should be between 5 and 20 classes.
2. The class width should be an odd number.
3. The classes must be mutually exclusive.
4. The classes must be continuous.
5. The classes must be equal in width.

Consider the following data set of 16 test scores on a statistics exam worth 85 points:

72, 76, 53, 68, 72, 85, 46, 77, 36, 81, 49, 73, 68, 65, 70, 71

Construct a frequency distribution for the data set using 4 classes.

1. Find the range (maximum – minimum)

2. Find the class width (maximum – minimum)/4

3. Determine the class lower limits (start with minimum data value)

4. Determine the class upper limits (one less than each successive lower limit)

5. Determine the class boundaries

6. Complete the Tally, Frequency, and Cumulative Frequency columns

Scores (Class limits)	Scores (Class boundaries)	Tally	Frequency	Cumulative frequency

Ungrouped Frequency Distribution

An ungrouped frequency distribution is constructed when the range and the number of distinct data values are small. In this case, the classes are single values rather than intervals of values.

Construct an ungrouped frequency distribution for the following data collected by a kindergarten teacher of his students' ages.

5	5	5	4
5	4	6	6
5	5	5	5
5	6	5	5
6	6	5	5

Age (Class)	Age (Class boundaries)	Tally	Frequency	Cumulative frequency

Why do we construct frequency distributions?

1. To organize the data in a meaningful, intelligible way.
2. To enable the reader to determine the nature or shape of the distribution.
3. To facilitate computational procedures for measures of average and spread.
4. To enable the researcher to draw charts and graphs for the presentation of data.
5. To enable the reader to make comparisons among different data sets.