Chapter 1 Part 2 Notes: Sections 1.3 – 1.4 and Excel Lab 2

Section 1.3: Frequency, Frequency Tables, and Levels of Measurement

The way a set of data is measured is called its level of measurement.

Data can be classified into four levels of measurement. They are (from lowest to highest level):

- Nominal scale level: Data that is measured using a nominal scale is qualitative.
    - Data that is measured using a **nominal scale** is **qualitative**.
        - *Categories, colors, names, labels and favorite foods along with yes or no responses are examples of nominal level data.*
    - Nominal scale data are not ordered.
        - *For example, trying to classify people according to their favorite food does not make any sense. Putting pizza first and sushi second is not meaningful.*
    - Nominal scale data cannot be used in calculations.
- Ordinal scale level
    - Data that is measured using an **ordinal scale** is similar to nominal scale data, but **the ordinal scale data can be ordered.**
        - *An example of ordinal scale data is a list of the top five national parks in the United States. The top five national parks in the United States can be ranked from one to five but we cannot measure differences between the data.*
    - Like the nominal scale data, ordinal scale data cannot be used in calculations.
- Interval scale level
    - Data that is measured using the **interval scale** is similar to ordinal level data because it has a definite ordering but **the differences between interval scale data can be measured though the data does not have a starting point.**
        - *Temperature scales like Celsius (C) and Fahrenheit (F) are measured by using the interval scale. In both temperature measurements, 40° is equal to 100° minus 60°. Differences make sense. But 0 degrees does not because, in both scales, 0 is not the absolute lowest temperature. Temperatures like -10° F and -15° C exist and are colder than 0.*
    - Interval level data can be used in calculations, but one type of comparison cannot be done. 80° C is not four times as hot as 20° C (nor is 80° F four times as hot as 20° F). There is no meaning to the ratio of 80 to 20 (or four to one).
- Ratio scale level
    - Data that is measured using the **ratio scale** takes care of the ratio problem and gives you the most information. Ratio scale data is like interval scale data, **but it has a 0 point and ratios can be calculated.**
        - *For example, four multiple choice statistics final exam scores are 80, 68, 20 and 92 (out of a possible 100 points). The exams are machine-graded. The data can be put in order from lowest to highest: 20, 68, 80, 92.*
        - Ratios can be calculated. The smallest score is 0. So 80 is four times 20. The score of 80 is four times better than the score of 20.

Types of Frequency:

- A **frequency** is the number of times a value of the data occurs. The sum of the values in the frequency column represents the total number of values included in the sample.
- A **relative frequency** is the ratio (fraction or proportion) of the number of times a value of the data occurs in the set of all outcomes to the total number of outcomes. To find the relative frequencies, divide each frequency by the total number of data values. Relative frequencies can be written as fractions, percents, or decimals.
- **Cumulative relative frequency** is the accumulation of the previous relative frequencies. To find the cumulative relative frequencies, add all the previous relative frequencies to the relative frequency for the current row. Cumulative frequencies should sum to 1.

<u>Example</u> Use the frequency table to answer the following questions:

HEIGHTS (INCHES)	FREQUENCY	RELATIVE FREQUENCY	CUMULATIVE RELATIVE FREQUENCY
59.95–61.95	5	$\frac{5}{100} = 0.05$	0.05
61.95–63.95	3	$\frac{3}{100} = 0.03$	0.05 + 0.03 = 0.08
63.95–65.95	15	$\frac{15}{100} = 0.15$	0.08 + 0.15 = 0.23
65.95–67.95	40	$\frac{40}{100} = 0.40$	0.23 + 0.40 = 0.63
67.95–69.95	17	$\frac{17}{100} = 0.17$	0.63 + 0.17 = 0.80
69.95–71.95	12	$\frac{12}{100} = 0.12$	0.80 + 0.12 = 0.92
71.95–73.95	7	$\frac{7}{100} = 0.07$	0.92 + 0.07 = 0.99
73.95–75.95	1	$\frac{1}{100} = 0.01$	0.99 + 0.01 = 1.00
	Total = 100	**Total = 1.00**	

Table 1.12 Frequency Table of Soccer Player Height

a) The percentage of heights that are from 67.95 to 71.95 inches is:

b) The percentage of heights that are from 67.95 to 73.95 inches is:

c) The percentage of heights that are more than 65.95 inches is:

d) The number of players in the sample who are between 61.95 and 71.95 inches tall is:

e) What kind of data are the heights?

f) Describe how you could gather this data (the heights) so that the data are characteristic of all male semiprofessional soccer players.

Excel Lab 2

What is the symbol for locking row/column references?

Using the cell reference A2, how would you lock the row?

Using the cell reference B5, how would you lock the column?

How do were use this locking feature when computing relative frequency?

How do we use this locking feature when computing cumulative frequency?

How can we turn on data analysis in Excel?

Explain how to use data analysis to create a frequency table from a set of data.

Example Use Excel to find relative and cumulative frequency, then answer the following questions.

Number of Years of Service	Number of Federal Employees
24	2
25	1
26	3
27	0
28	4
29	6
30	11
31	12
32	7
33	8
34	6
35	10

a) What is the cumulative frequency for years of service between 30 and 35 (inclusive)?

b) What is the relative frequency for 30 years of service?

c) What is the relative frequency for 30 years of service or less?

d) What is the relative frequency for 25 years of service or more?

Section 1.4: Experimental Design and Ethics

- The purpose of an experiment is to investigate the relationship between two variables.
- When one variable causes change in another, we call the first variable the **explanatory variable**. The affected variable is called the **response variable**.
- In a randomized experiment, the researcher manipulates values of the explanatory variable and measures the resulting changes in the response variable.
- The different values of the explanatory variable are called **treatments**. An **experimental unit** is a single object or individual to be measured.

What should you consider if you are trying to run an experiment and why? This is an open-ended question, share your thoughts.

What is a lurking variable?

How must a researcher account for lurking variables?

What is the power of suggestion?

Explain the difference between a blinded versus a double-blinded study:

How do ethics play a role in statistics?