Liberian Mathematics Teacher Training Program 2023–2024

Andrew Obus[1]

CUNY Baruch College

September 29, 2023

[1]This program is partially supported by NSF CAREER Grant DMS-2047638

HW Exercise 1

Suppose a survey is given where people are asked about the highest level of education they have completed: Primary School, Secondary School, University, or Graduate School. What do you think would be the most appropriate way to visually present the resulting data?

HW Exercise 2

Make a stem and leaf plot for the following data: Liberia: Population by county (2008): Bomi: 82036, Bong: 328919, Gbarpolu: 83758, Grand Bassa: 224839, Grand Cape Mount: 129055, Grand Gedeh: 126146, Grand Kru: 57106, Lofa: 270114, Margibi: 199689, Maryland: 136404, Montserrado: 1144806, Nimba: 468088, Rivercess: 65862, River Gee: 67318, Sinoe: 104932. Try to choose the stem and the leaves so that the plot you get is meaningful!

Numerical data

- Stem and leaf plots
- Histograms and frequency tables (TODAY!)
- Cumulative frequency curves
- Box and whisker plots

Frequency distributions and histograms

- A histogram is visually similar to a bar graph, but is used when the data are numerical.
- Histograms work well whether or not the data set is large or small, and whether or not there are lots of different values.
- One website you can use to make histograms is https://www.socscistatistics.com/descriptive/histograms/
- Histograms can also be made in Microsoft Excel
- It is also worth learning how to make them by hand.

Histograms, continued

- To make a histogram, we start with a *frequency table* for our data. A frequency table simply has two columns, one listing each measurement and the other listing the number of times that measurement occurs in the data.
- When we make a histogram, we need to choose a horizontal and a vertical scale. The vertical axis usually represents the *frequency* of the measurements, while the horizontal axis represents the *values* of the data.
- When the data takes on very few values, one marks these values along the horizontal axis, and makes a column with each value at the center. The *height* of the column represents the number of times that data value occurs.

Histogram example

- Raccoons in Texas: mating behavior. Number of partners for each female: 1 3 2 1 1 4 2 4 1 1 1 3 1 1 1 1 2 2 1 1 4 1 1 2 1 1 1 1 3.
- We will make an absolute and a relative frequency table, and then construct a histogram by hand.
- Remember that

$$\text{relative frequency} = \frac{\text{absolute frequency}}{\text{total number of measurements}}.$$

Histograms using intervals

- If the data take lots of values, it can be more convenient to build a frequency table where each row represents a *range* of values across an interval, rather than a *particular* value.
- These ranges are sometimes called *class intervals*.
- One should generally make each of the class intervals equal width.
- The corresponding histogram should have one bar for each interval.

Interval histogram example

Average number of minutes used per month (cell phone usage)

Minutes	Relative Frequency
0 to $<$ 200	.585
200 to $<$ 400	.180
400 to $<$ 600	.115
600 to $<$ 800	.120

Creating an interval histogram from discrete data

- Suppose we are not given a frequency table, and instead we are given a list of data, say, exam scores from a class.
- There are too many different scores for a discrete frequency table to be useful. We have to use class intervals.
- So we decide on what class intervals to use, and then we build a frequency table using those intervals. After we do this, we can draw (or have a computer draw) the histogram.

- One rule of thumb is for the number of intervals to be roughly the square root of the total number of observations.
- This is not a hard-and-fast rule. But it often leads to useful histograms that give insight into the overall distribution of the data.
- Note that the *number* of intervals and the *width* of the intervals are related by the formula

$$\text{NUMBER of intervals} \times \text{WIDTH of intervals} = \text{range of data.}$$

Example

Suppose a class has the following scores on an exam: 53, 56, 61, 62, 68, 68, 70, 72, 73, 73, 75, 77, 78, 78, 79, 80, 81, 84, 85, 86, 86, 87, 87, 87, 89, 90, 92, 92, 96, 98. We will create a frequency table and histogram based on 5 class intervals.

Homework Exercise 1

The following is a frequency distribution for commuting times to work (in minutes) for working adults in a county in California.

Time	Frequency
0 to $<$ 5	5200
5 to $<$ 10	18200
10 to $<$ 15	19600
15 to $<$ 20	15400
20 to $<$ 25	13800
25 to $<$ 30	5700
30 to $<$ 35	10200
35 to $<$ 40	2000
40 to $<$ 45	2000

Add a column to this table for relative frequency, and then create a histogram for this data using relative frequency on the vertical axis.

Homework Exercise 2

The following data represents total length of roads (in miles) in various subdivisions in the US: 1280, 360, 3350, 450, 1850, 3150, 5320, 3330, 540, 2250, 2460, 1890, 4390, 3380, 3870, 2320, 5850, 510, 2100, 340, 1250, 2400, 2700, 240, 1240, 1000, 2400, 3150, 2730, 3060. The data was used to help determine whether it would make more sense to build power lines above ground or underground.

Construct a stem and leaf plot for this data using the thousands digit as a stem. Then, contruct a histogram for the data using the classes 0 to < 1000, 1000 to < 2000, 2000 to < 3000, and so on.

Thank you for your attention. There is NO meeting next week (October 6). Our next meeting will be on October 13 at 3:30 PM (note the time change)! We will discuss cumulative frequency tables, plots, and percentiles.