

Math 1100: Week 3: The concept of a distribution

- The **distribution** of a variable is a record of the frequency with which each of the possible values occurs within the reference class. (*Comment:* You can't begin to think about the distribution of a variable with being clear in your mind about what the reference class is and what the possible values of the variable are. This is why I emphasized these two basic ideas last week.)
- A distribution may be displayed by means of a **frequency table**.
 - *Example.* Consider the variable “score on the first quiz in Math 1100 section 3, fall 2002”. The reference class for this variable consists of all the people who took the quiz. When the quiz was graded, I decided not to give fractional credit. The possible values of the variable were the numbers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. We can describe the distribution of the variable by means of a “frequency table”, as follows:

score	0	1	2	3	4	5	6	7	8	9	10
number of students	0	0	1	1	4	1	5	4	1	5	12

On the second quiz in Math 1100, section 3, fall 2002, I gave fractional credit, but only in multiples of $1/2$ of a point. As a result, the possible scores on the quiz were:

0, 0.5, 1.0, 1.5, 2.0, 2.5, . . . , 9.0, 9.5, 10.0

The actual scores were:

7.5, 8, 5, 9, 10, 7, 4, 9, 6, 10, 9.5, 10, 6, 4, 9.5, 9.5, 9, 7.5, 5,
9, 9.5, 10, 9, 7.5, 10, 7.5, 5.5, 5.5, 8, 9.5, 6, 9.5, 6, 9.5, 9.5

- *Exercise.* Make a frequency table for quiz 2.
- In making a frequency table for a continuous variable, one often uses **bins**. You divide up the range of possible values into intervals and record how many times the variable takes on values in each of the intervals. You must decide on the size of the intervals. Bins may also be used with a discrete variable to lump together nearby values.

◦ *Example.* In a class of 60 students the scores on a 100-point test were:

54.7	55.2	56.9	60.4	61.0	65.2	66.4	66.4	66.7	66.7
67.0	67.0	67.1	67.7	68.1	68.7	69.7	70.3	70.5	70.6
70.8	70.8	70.9	71.6	71.9	72.4	73.3	73.5	74.3	74.8
75.6	75.7	75.9	75.9	76.3	77.3	77.4	77.4	77.7	77.8
77.8	79.0	79.2	79.4	79.8	79.8	80.8	80.9	81.3	81.4
82.5	84.5	85.1	85.8	86.2	89.2	91.6	96.2	98.8	99.9

If we decide to make bins for the scores in the 50s, 60s, 70s, 80s and 90s, then the “binned frequency table” looks like this:

score range	50 – 59.9	60 – 69.9	70 – 79.9	80 – 89.9	90 – 99.9
number of students	3	14	29	10	4

We could make smaller bins if for some reason we wanted. This time, we'll write the table vertically:

score range	number of students
50.0 – 54.9	1
55.0 – 59.9	2
60.0 – 64.9	2
65.0 – 69.9	12
70.0 – 74.9	13
75.0 – 79.9	16
80.0 – 84.9	6
85.0 – 89.9	4
90.0 – 94.9	1
95.0 – 99.9	3

- If we are interested in the characteristics of a reference class that are reflected in a given variable, then the **relative frequencies** of the different values of that variable are more important than the absolute frequencies. The relative frequency of a particular value of variable is

$$\frac{\text{number of occurrences of the value}}{\text{number of units in the reference class}}$$

For example, the entries in the previous table may be expressed as relative frequencies. One can use fractions, decimals or percents.

score range	number of students	relative frequency	rel. freq. (decimal)	rel. freq. (per cent)
50.0 – 54.9	1	1/60	0.0166	1.66
55.0 – 59.9	2	2/60	0.0333	3.33
60.0 – 64.9	2	2/60	0.0333	3.33
65.0 – 69.9	12	12/60	0.2000	20.00
70.0 – 74.9	13	13/60	0.2166	21.66
75.0 – 79.9	16	16/60	0.2666	26.66
80.0 – 84.9	6	6/60	0.1000	10.00
85.0 – 89.9	4	4/60	0.0666	6.66
90.0 – 94.9	1	1/60	0.0166	1.66
95.0 – 99.9	3	3/60	0.0500	5.00

- Bar graphs, dot plots and various other kinds of charts or graphs may be used to display a distribution. The kind of chart you use is mostly a matter of convenience.

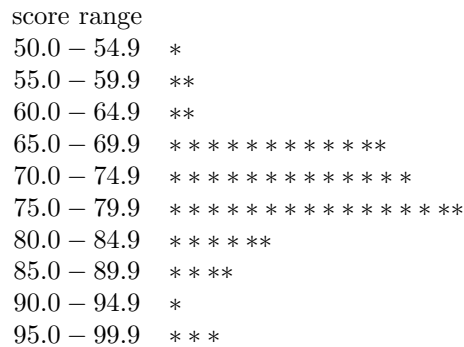


Figure 1. Is this a bar graph, dot chart or a stem and leaf plot? (Who cares?)

- **CAUTION!** Not every bar graph depicts a distribution. (In fact, depictions of distributions are not very common in newspapers and magazines.) To know whether or not a chart is showing you a distribution, you must ask yourself, “What is the variable? What are the possible values? Is this chart showing me the frequencies (or relative frequencies) with which the different values occur within the reference class?”
- **Histograms** and **stem-and-leaf plots**. See text, pages 104–5. Note that a histogram uses binning.
- **Pie charts** are an especially good way of showing relative frequencies for a qualitative variable that has a small number of possible values, *e.g.*, political party. Pie charts are almost always used to show distributions, so if you want to find a distribution in a newspaper, look for a pie chart.

Review/Homework.

- Page 93: review questions 1,2,3.
- Page 94: 1–8, 11, 12, 20.
- Page 109: review questions 1–5.
- Page 110: 7, 9, 11, 15, 16, 17, 18.