Math 214 – Introductory Statistics **Summer 2008**
6-6-08 Class Notes

Sections 2.2, 2.3
2.2: 3-15 odd
2.3: 1-15 odd

Frequency Distributions

Suppose we have the following data (TV viewing times for 50 students in hours per week):

3	8	17	7	7	1	12	13	4	15
9	15	10	10	20	18	3	6	5	16
4	17	11	9	9	18	5	14	12	5
7	8	11	19	16	3	5	12	9	14
14	18	4	7	11	10	10	9	13	5

The frequency distribution is:

Class No.	Hours per week	Number of students
1	1-4	7
2	5-8	12
3	9-12	15
4	13-16	9
5	17-20	7
		50

Definition: The ***lower class limits*** are the smallest values that can actually belong to a class. The ***upper class limits*** are the largest values that can actually belong to a class. If there is a gap between the upper class limit for one class and the lower class limit for the next class, we call the midpoint between the two numbers the ***class boundary***. The ***class width*** is the difference between lower class limits. This should be constant in a frequency distribution.

For the above example, the lower class limits are 1, 5, 9, 13, and 17, the upper class limits are 4, 8, 12, 16, and 20, the class boundaries are 0.5, 4.5, 8.5, 12.5, 16.5, and 20.5, and the class width is 4.

Definition: The ***range*** is the difference between the lowest data value and the highest data value. For each class, the ***class midpoint*** is the average of the class limits.

In our previous example the class midpoints were 2.5, 6.5, 10.5, 14.5, and 18.5, and the range was 19.

Guidelines for constructing a frequency distribution:

- Arrange your data in order from lowest to highest.
- Decide on how many classes you want to have. There should be several, but not too many. Of course, this is an individual choice, but generally speaking, you should have 5 to 20 classes. Avoid empty classes (if possible). There is actually a formula to help make the decision as to how many classes to have, called ***Sturges' Rule:***

$$\#\text{ of classes} \approx 1 + 3.3\log(n)$$

(where n is the number of data values)

- Choose your class width. It should be approximately equal to the range of data divided by the number of classes (if you don't get an integer, round up!).

Example: Construct a frequency distribution for the data below (which are grades on a Math 214 exam).

data						sorted data					
53	87	76	73	62	99	49	53	59	62	63	65
78	93	82	69	65	93	65	69	73	73	76	77
92	92	78	82	89	65	78	78	82	82	82	85
63	49	88	87	94	73	87	87	88	89	92	92
85	77	98	59	93	82	93	93	93	94	98	99

Class No.	Grade	Number of students
1	41-50	1
2	51-60	2
3	61-70	5
4	71-80	6
5	81-90	8
6	91-100	8
		30

A couple of properties of the classes: they must be *exhaustive* (every data value must fit somewhere) and *mutually exclusive* (every data value must fit in only one class).

There are a few other columns we can have in a frequency distribution. In addition to the frequency for each class, we can have the ***cumulative frequency*** (the frequency of all classes up to an including the chosen class), the ***relative***

frequency (frequency of the class divided by the total number of data values), and the ***cumulative relative frequency*** (the relative frequency of all classes up to and including the chosen class). These are added below:

Class	Freq.	Cum. Freq.	Rel. Freq.	Cum. Rel. Freq.
41-50	1	1	$\frac{1}{30} \approx .033$	.033
51-60	2	3	$\frac{2}{30} \approx .067$	.100
61-70	5	8	$\frac{5}{30} \approx .167$	.267
71-80	6	14	$\frac{6}{30} \approx .200$	.467
81-90	8	22	$\frac{8}{30} \approx .267$	.733
91-100	8	30	$\frac{8}{30} \approx .267$	1.000
	30		1	

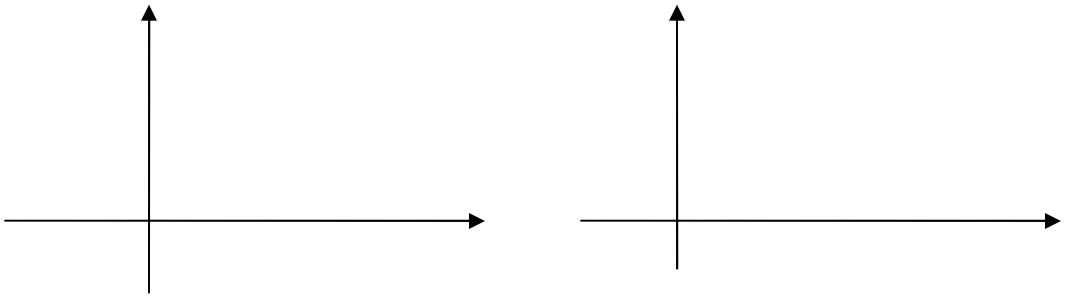
Graphical Representations

Definition: A ***histogram*** is a bar graph where the horizontal axis represents the classes (specifically the class boundaries) and the vertical axis represents the frequencies or relative frequencies.

Example: Let's use the data from our last example and graph it.

Definition: A ***frequency polygon*** is a piecewise linear graph for which the horizontal axis represents the class midpoints and the vertical axis represents the frequencies or relative frequencies. The ***ogive*** is a piecewise linear graph for which the horizontal axis represents the class upper limits and the vertical axis represents the cumulative frequencies or cumulative relative frequencies.

The frequency polygon and the ogive for our data set are as follows:



Finally, let's draw all three graphs using the relative frequencies.