Measures of Spread (Section 2.2)

1. Consider again the high temperatures in 36 US cities on July 20, 2012 (source: http://www.nws.noaa.gov/xml/tpex/scs.php).

CITY	HI
ABILENE TX	93
AKRON CANTON	92
ALBANY NY	94
ALBUQUERQUE	100
ALLENTOWN	95
AMARILLO	92
ANCHORAGE	60
ASHEVILLE	86
ATLANTA	88
ATLANTIC CITY	95
AUSTIN	90
BALTIMORE	98
BATON ROUGE	89
BILLINGS	73
BIRMINGHAM	87
BISMARCK	72
BOISE	82
BOSTON	97
BRIDGEPORT	92
BROWNSVILLE	87
BUFFALO	85
BURLINGTON VT	95
CARIBOU	85
CASPER	72
CHARLESTON SC	86
CHARLESTON WV	94
CHARLOTTE	93
CHATTANOOGA	91
CHEYENNE	68
CHICAGO	95
CINCINNATI	90
CLEVELAND	91
COLORADO SPGS	77
COLUMBIA SC	91
COLUMBUS GA	89
COLUMBUS OH	93

   a. What is the range of the data?
   b. Find the mean, $\bar{x}$ of the sample:
   c. Find the standard deviation, s, of the data (assuming it is a sample):
   d. Calculate $\bar{x} - s =$
   e. Calculate $\bar{x} + s =$

f. Count the number of data points between the two values you found in d and e: i.e. the number of temperatures in the range $\bar{x} - s$ to $\bar{x} + s$. What percent of the total is that?

g. Calculate $\bar{x} - 2s =$

h. Calculate $\bar{x} + 2s =$

i. Count the total number of data points between the two values you found in g and h: i.e. the number of temperatures in the range $\bar{x} - 2s$ to $\bar{x} + 2s$. What percent of the total is that?

j. Calculate $\bar{x} - 3s =$

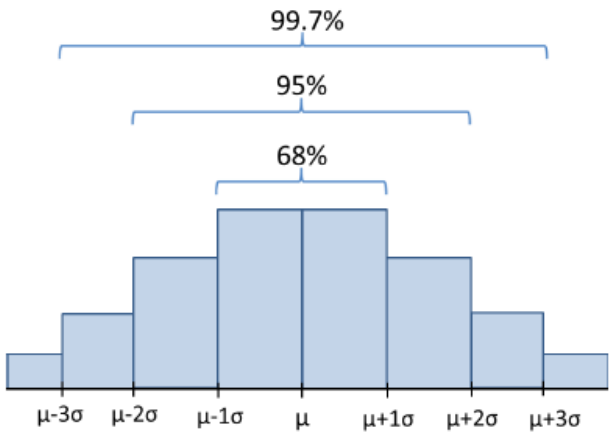
k. Calculate $\bar{x} + 3s =$

l. Count the total number of data points between the two values you found in j and k: i.e. the number of temperatures in the range $\bar{x} - 3s$ to $\bar{x} + 3s$. What percent of the total is that?

The Empirical Rule: (see page. 19 of your book)

Empirical Rule: If the data is approximately **normally distributed**, the following are true.

- About **68%** of all values fall within **1** standard deviation of the mean.
- About **95%** of all values fall within **2** standard deviations of the mean.
- About **99.7%** of all values fall within **3** standard deviations of the mean.



2. Does the data in problem 1 seem approximately normally distributed? Explain.

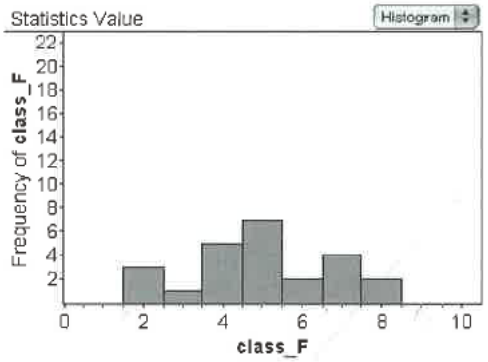
3. Do problems 13 and 14 on page 19 of your book.

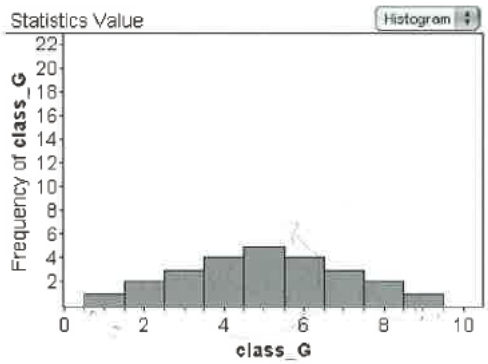
4. Are there any “unusual” values in the data in problem 1? Explain

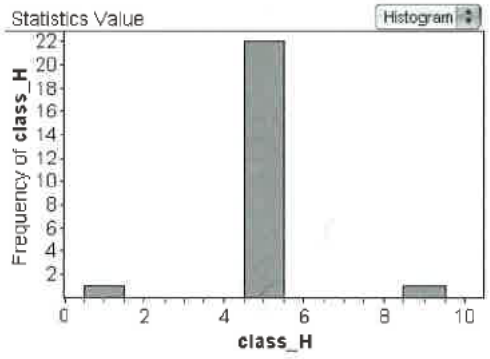
5. The following problem was taken from *Workshop Statistics 3rd Ed.* by Rossman, Chance and Lock:

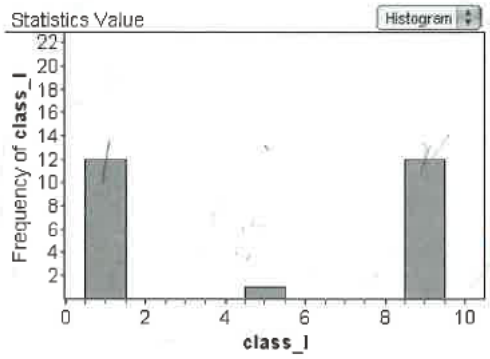
   Consider the following hypothetical ratings from five classes of the value of statistics on a 1-9 scale. The data are given in the following frequency table and displayed in the following histograms:

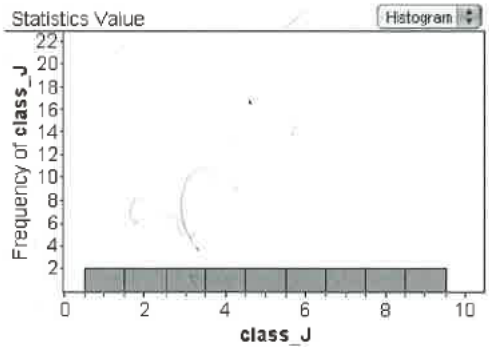
Rating	1	2	3	4	5	6	7	8	9
Class F Count	0	3	1	5	7	2	4	2	0
Class G Count	1	2	3	4	5	4	3	2	1
Class H Count	1	0	0	0	22	0	0	0	1
Class I Count	12	0	0	0	1	0	0	0	12
Class J Count	2	2	2	2	2	2	2	2	2











a. Judging from the tables and histogram, make a prediction as to which classes' ratings have more variability: class F or class G. Explain your reasoning.

b. Judging from the tables and histograms, which class has the most variability in ratings among classes H,I and J? Which class has the least variability in ratings? Explain your reasoning.

Most variability: Least variability: