Name: _____

Summer 2018

Problem 1: Explain the difference between a census and a sample.

Problem 2: Explain the difference between statistical significance and practical significance. Give an example.

Problem 3: Does correlation imply causation? Give an example.

Problem 4: Explain what is wrong here: the magazine *Hired* performed a phone survey of 561 people. Of the people surveyed, 82% of individuals stated that they believed your outfit you wore during an interview affected whether or not you were hired.

Problem 5: Identify whether the following survey methods are random, systematic, convenient, stratified, or clustered.

(a) A survey of SU professors is done by selecting every 10th professor listed in the SU telephone directory.

(b) An education researcher randomly select 3 middle schools in New York State and interviews all the teachers at each of these selected schools about their salary satisfaction.

(c) A market researcher selects 500 people from each of all 50 American states.

(d) An education researcher randomly selects 10 universities and interviews all the teachers at each university.

(e) Four different classes from Cornell were randomly selected, and all of their students were surveyed about their satisfaction of the campus snow removal.

(f) A sample consists of every 5th student alphabetically from a class of 163 students.

(g) A professor was asked to estimate the average GPA of students in his department. The professor took an average from 10 different large lecture classes she had taught.

(h) A university wants to sample their incoming freshmen. They put all their names into a spreadsheet and use a computer to randomly select 50 names.

(i) A teacher in a large lecture writes every name on a slip of paper, places them into a large covered container, reaches in, and grabs 10 names.

(j) A marketing company has a computer which randomly generates phone numbers and then dials those numbers and talks to whomever answers.

Problem 6: Explain the difference between a parameter and a statistic.

Problem 7: Determine whether the following data is quantitative or categorical; the underlined word indicates the information considered:

(a) My zip code is 13206.

(b) My height is 132 cm.

(c) His gender is male.

(d) That shirt is blue.

(e) The temperature is 92°F.

Problem 8: Determine whether the given information is at the nominal, ordinal, interval, or ratio level of measurement. The underlined word indicates the information considered.

(a) My car is a Toyota.

(b) My temperature is 98.7°F.

(c) I was born in the year 1978.

(d) I have walked 500 miles.

(e) My volleyball team placed 6th in the tournament.

(f) The house is yellow.

(g) This textbook costs $800.

Problem 9: Determine whether the given data is a statistics or a parameter.

(a) A survey by the CDC included 5,283 individuals from the U.S..

(b) The mean atomic weight of all the elements on the periodic table is 134.355 amu.

(c) In a study of average IQ of Boston residents, a group of researchers found an average IQ of 93.3 among a group of 127 individuals in Boston.

(d) According to the Census Bureau, the total number of Hispanics in the U.S. is 52,232,121 persons.

Problem 10: What does it mean for a study to be blinded? What does it mean for a study to be double-blinded? Give an example.

Problem 11: Explain the difference between a random sample and a simple random sample. Give an example.

Problem 12: Determine whether the following are a random sample, a simple random sample, or both.

(a) In a large auditorium, the 8 individuals with the highest exam score are selected.

(b) In a large auditorium, a group of students is selected by randomly lining up the students and taking every 5th student.

(c) In a large auditorium, a group of students is selected by choosing a person at random from each row.

Problem 13: Give an example of a cross-sectional, retrospective, and prospective study.

Problem 14: What is a matched-pairs experimental design? Given an example.

Problem 15: Compute the following:

(a) $\dfrac{153 - 147}{6} =$

(b) $\dfrac{1211 - 1006}{52} =$

(c) $\dfrac{13 - 27}{3.1} =$

(d) $\dfrac{134 - 122}{\dfrac{\sqrt{10.1}}{11}} =$

Problem 16: Consider the following frequency distribution:

Range	Frequency	Cumulative Frequency
20–29	30	
30–39	34	
40–49	17	
50–59	5	
60–69	2	
80–89	1	

(a) Identify the class boundaries.
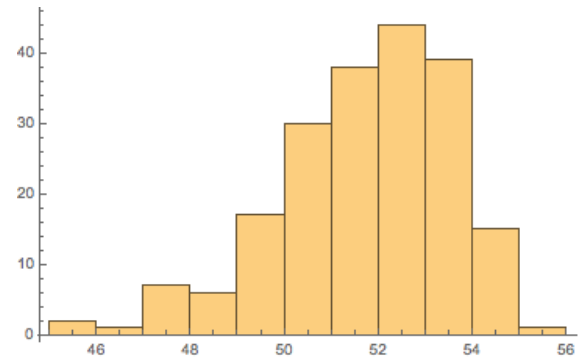
(b) Identify the class midpoints.

(c) Identify the class ranges.

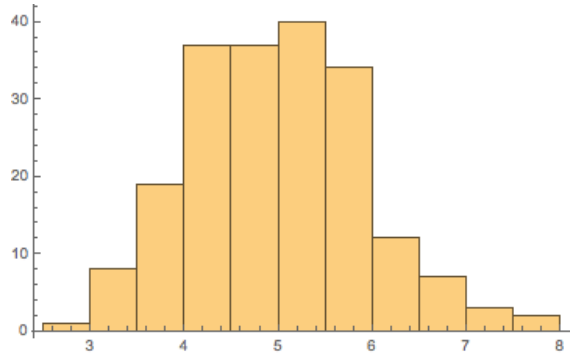(d) Complete the cumulative frequency column of the table.

Problem 17: Determine if the following distributions are symmetric or skewed. If it is skewed, indicate whether the distribution is right skewed or left skewed.
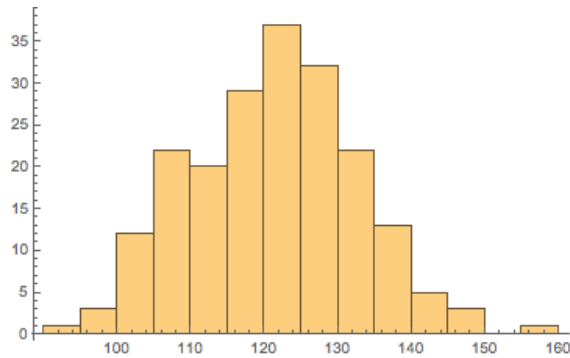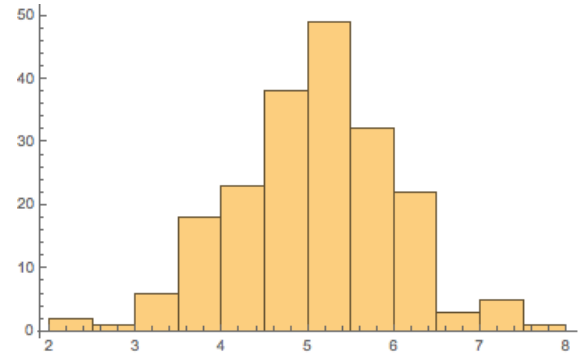

(a)

(b)

(c)

(d)

Problem 18: A statistician performs a linear regression between years of education (beyond high school), x, and average starting salary (in thousands of dollars), y. A model of $\hat{y} = 6.4\hat{x} + 15$ is obtained with correlation coefficient $r = 0.834$.

(a) Is data used 'mostly' linear? Explain.

(b) What 'percent' of the data is explained by the model?

(c) Is there a positive or negative correlation between x and y? Explain.

(d) Interpret the slope for the model.

(e) Interpret the y-intercept for the model.

(f) Using the model, predict the average starting salary of a person who attended 4 years of college.

(g) Using the model, predict how many years of college would one have to attend to have an average starting salary of $53,000.

Problem 19: The weights of a certain type of watermelon are normally distributed with mean 5 kg and standard deviation of 0.7 kg. What percent of this type of watermelon weights less than 4.3 kg?
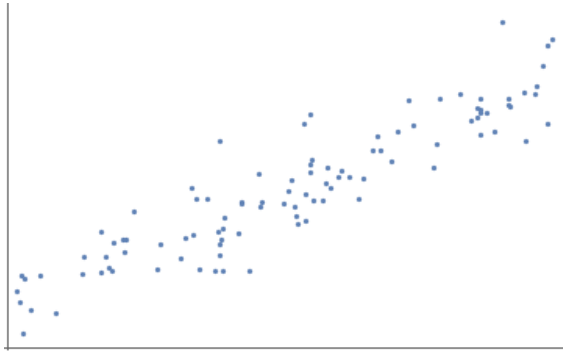
Problem 20: Consider the following data:

x	y
0.2	12.84
1.7	6.69
2.3	5.49
3.0	5.0
3.6	5.36
4.2	6.44

Performing a linear regression on this data gives a Pearson correlation coefficient, r, of 0.617. True or False: this r value implies there is not very much correlation between x and y. Explain.
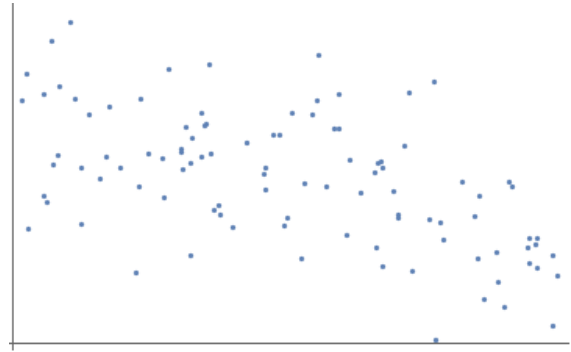
Problem 21: Construct a stem-and-leaf plot of the following data:

$$1.4, \quad 2.3, \quad 4.4, \quad 3.6, \quad 3.5, \quad 4.5, \quad 3.2, \quad 3.7, \quad 4.0, \quad 5.1, \quad 4.2, \quad 5.3$$
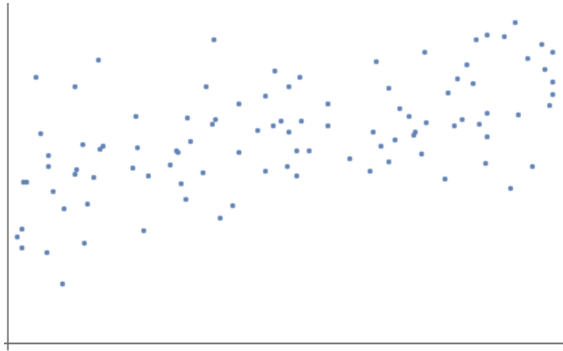
Problem 22: Match each of the following scatter plots with the correlation coefficients of its linear regression. On each plot, sketch a possible line of best fit for the data. Note that the scale on each axes is the same.
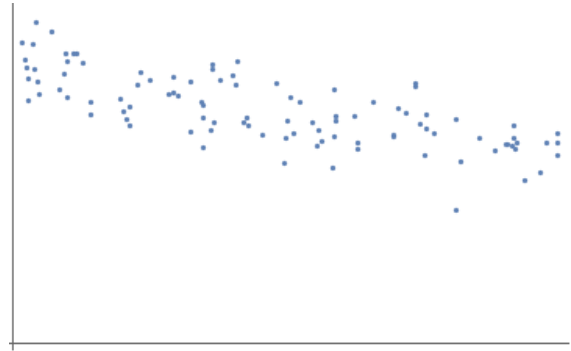
(a)

(b)

(c)

(d)

_____ : $r = 0.912$

_____ : $r = -0.855$

_____ : $r = -0.273$

_____ : $r = 0.546$

Problem 23: Is the median a robust measure of center? Explain with an example.

Problem 24: Consider the following sample of data values:

$$17, \ , 12, \ 18, \ 13, \ 9, \ 24, \ 30, \ 14, \ 32, \ 39, \ 20, \ 22, \ 15, \ 21, \ 23.$$

(a) Find the 5-number summary for the data.

(b) Find the percentile of the data value 21.

(c) Find P_{60}, the value of the 60th percentile.

Problem 25: David Hilbert is 70 in in height. If the mean height of all males is 64.5 in and the standard deviation is 2.5 in, compute Hilbert's z-score.

Problem 26: Determine if the following distributions are symmetric or skewed. If it is skewed, indicate whether the distribution is right skewed or left skewed.
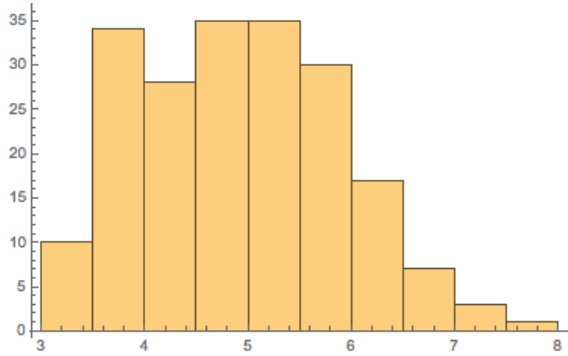
(a)
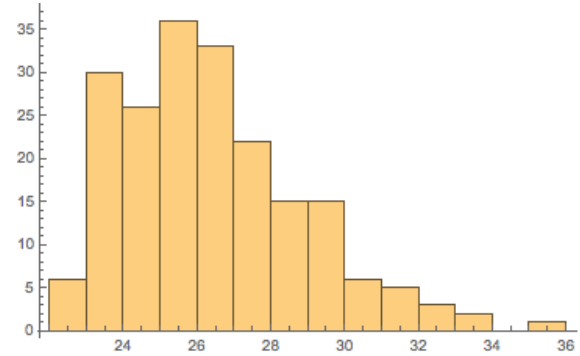
(b)

(c)
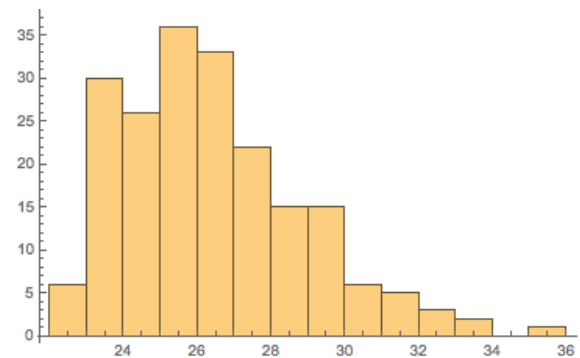
(d)

Problem 27: Is the mean a robust measure of center? Explain with an example.

Problem 28: The table below classifies a group of voters according to sex and political affiliation.

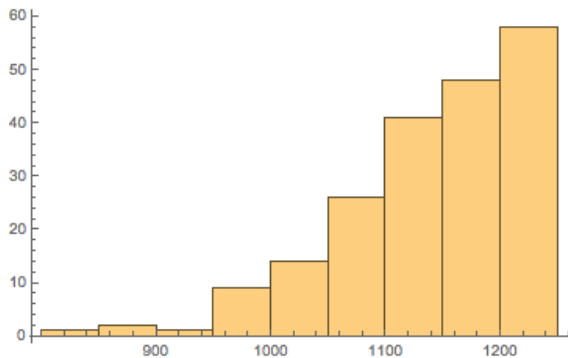|	Democrat	Republican	Independent
Male	115	271	53
Female	299	142	47

(a) How many voters were male?

(b) How many voters were there?

(c) How many voters were Republican?

(d) What percent of voters were Democrats?

(e) How many voters were Independent and female?

(f) How many voters were Republican or female?

(g) What percent of independent voters were female?

(h) What proportion of males were independent?

Problem 29: Consider the following sample of data values:

$$10, \quad 9, \quad 7, \quad 12, \quad 11, \quad 9, \quad 12, \quad 8, \quad 13, \quad 10$$

(a) What is the mean?

(b) Find P_{27}.

(c) Find P_{70}.

Problem 30: Find the mean, median, mode, and midrange for the following list of data:

$$5, \quad 5, \quad 5, \quad 6, \quad 8, \quad 9, \quad 9, \quad 11, \quad 13, \quad 15$$

Problem 31: Given the following data values, construct a frequency histogram. Begin at 3 and use bin width 1.

$$3, \ 5, \ 6, \ 4, \ 4, \ 3, \ 4, \ 5, \ 6, \ 6, \ 3, \ 4$$

Problem 32: Is the midrange a robust measure of center? Explain with an example.

Problem 33: Explain the difference between a box plot and a modified boxplot? Use an example for demonstration.

Problem 34: Consider the following list of data.

$$2, \ 6, \ 7, \ 9, \ 10, \ 13, \ 15, \ 17, \ 18, \ 19, \ 22, \ 25$$

(a) Find the mean and standard deviation for the data.

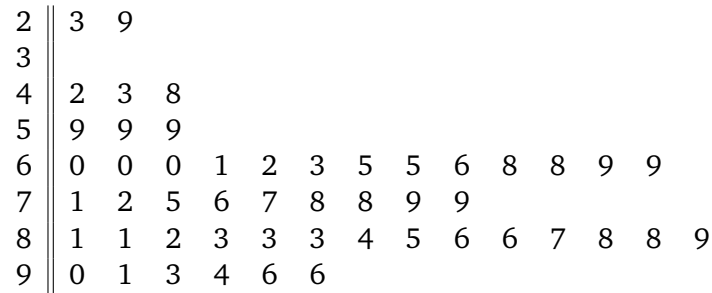(b) Give the 5-number summary.

(c) Are there any outliers?

(d) Find the percentile of the data value 18.

(e) Find the value of the 20th percentile, P_{20}.

Problem 35: The salaries at a large technology company are approximately normal with mean $72,000 and standard deviation $5,000. What proportion of people at the company earn between $67,000 and $77,000?

Problem 36: The following stem plot displays the proportion (in the form of a percentage) of non-Hispanic whites for the fifty states of the U.S. and the District of Columbia (*USA Today, August 15, 2006*). In the stem plot, the stem represents the tens digit and there are 51 data values.

```
2 ‖ 3  9
3 ‖
4 ‖ 2  3  8
5 ‖ 9  9  9
6 ‖ 0  0  0  1  2  3  5  5  6  8  8  9  9
7 ‖ 1  2  5  6  7  8  8  9  9
8 ‖ 1  1  2  3  3  3  4  5  6  6  7  8  8  9
9 ‖ 0  1  3  4  6  6
```

(a) Give the 5-number summary for the data.

(b) The distribution is (right skewed, left skewed, approximately symmetric)

(c) The level of measurement of the proportions is (nominal, ordinal, interval, ratio)

(d) The proportions presented in the stem plot are (categorical, quantitative) data.

Problem 37: For the following numbers, compute the mean, median, and standard deviation:

$$-2, \quad 1, \quad 4, \quad 5, \quad 7, \quad 8$$

Problem 38: Exam A had a mean of 30 and standard deviation of 3 while Exam B had a mean score of 270 and standard deviation 27. If Sophie scored a 28.3 on Exam A and Emmy scored a 260 on Exam B, who did better?

Problem 39: For the following list of numbers, compute the mean, median, and sample standard deviation.

$$7, \quad 9, \quad 13, \quad 15, \quad 19$$

Problem 40: The table below gives a classification of the grades earned by 44 students on a 50-point exam.

Grade	Frequency	Class Midpoint
21–25	4	
26–30	5	
31–35	9	
36–40	11	
41–45	9	
46–50	6	

(a) Construct a frequency histogram for the distribution of the grades, labeling it appropriately. Use the class boundaries for the horizontal axis.

(b) Determine the class midpoints and enter them in the chart above.

(c) Estimate to 1 decimal place the mean of the exam grades.

Problem 41: For the following list of numbers, find the mean, median, mode, and midrange.

$$4, \ 4, \ 4, \ 5, \ 6, \ 7, \ 9, \ 10, \ 11, \ 11$$

Problem 42: Person A scored 70 on a test where the mean was 60 and the standard deviation was 4. Person B scored 80 on a test where the mean was 50 and the standard deviation was 15.

(a) Compute the z-score for each person.

(b) Which person scored relatively higher?

Problem 43: Make a stem plot of the following data:

$$57, \ 58, \ 42, \ 56, \ 54, \ 60, \ 43, \ 65, \ 46, \ 54, \ 72$$

Also, make a standard frequency table for this data. Use a class width of 7 and start at a lower class limit equal to 40.

Problem 44: On a 4.3 grade scale, the following letter grades receive the following weight:

Letter	Scale
A+	4.3
A	4.0
A−	3.7
B+	3.3
B	3.0
B−	2.7
C+	2.3
C	2.0
D	1.0
F	0

If a student took a class worth 4 credits and received an A+, a class worth 3 credits and received a C, a class worth 4 credits and received a B+, and a class worth 3 credits and received an A−, what was their GPA that semester?

Problem 45: If a distribution of heights in a classroom is approximately normal with mean 5 ft with standard deviation of 1 ft, what percent of the class is between 4 ft tall and 7 ft tall?

Problem 46: Consider the following dataset:

$$2, \quad 5, \quad 11, \quad 11, \quad 13, \quad 15, \quad 15, \quad 16, \quad 17, \quad 19, \quad 19, \quad 28$$

(a) Find the 5-number summary for the data.

(b) Construct a box plot of the data.

(c) Are there any outliers?