Math 10 - Homework 2
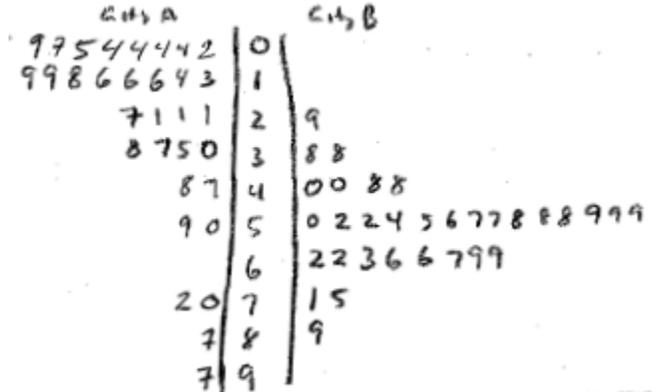
1. The following average daily commute time (minutes) for residents of two cities.

City A	2	4	4	4	4	5	7	9	13	14	16	16	16	18	19	19
	21	21	21	27	30	35	37	38	47	48	50	59	70	72	87	97

City B	29	38	38	40	40	48	48	50	52	52	54	55	56	57	57	58
	58	58	59	59	59	62	62	63	66	66	67	69	69	71	75	89

a. Construct a back-to back stem and leaf diagram and interpret the results.

City A   City B
9 7 5 4 4 4 4 2 | 0 |
9 9 8 6 6 6 6 4 3 | 1 |
7 1 1 1 | 2 | 9
8 7 5 0 | 3 | 8 8
8 7 | 4 | 0 0 8 8
9 0 | 5 | 0 2 2 4 5 6 7 7 8 8 8 9 9 9
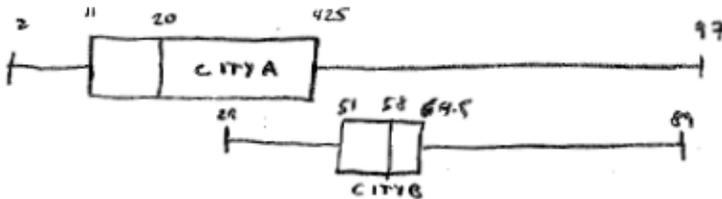 | 6 | 2 2 3 6 6 7 9 9
2 0 | 7 | 1 5
7 | 8 | 9
7 | 9 |

Result for A heavily skewed
" " B symmetric

City A more narrowish,
City B continuous but longer

b. Find the quartiles and interquartile range for each group.
c. Calculate the 80th percentile for each group.

	Q1	Med	Q3	IQR	80P
CITY A	11	20	42.5	31.5	48
CITY B	51	58	64.5	13.5	66

d. Construct side-by-side box plots and compare the two groups.

City A more spread out
City B longer commutes
Both sets skewed right.

e. For each group, determine the z-score for a commute of 75 minutes. For which group would a 75 minute commute be more unusual.

$$Z\text{-score} = \frac{75 - \bar{X}}{S}$$

|	$\bar{X}$	S	Z
City A	29.06	25.15	1.81
City B	57.00	12.12	1.48

75 is more unusual for City A

2. The following data represents the heights (in feet) of 20 almond trees in an orchard.

14 14 14 14 15 18 18 20 21 21
22 24 25 25 25 27 27 29 31 45

a. Construct a box plot of the data.

b. Do you think the tree with height of 45 feet is an outlier? Use both methods we covered in class to justify your answer.

$$IQR = 9.5$$
$$Fence \ 1 = 26 + (9.5)(1.5) = 40.25$$
$$45 \ is \ a \ possible \ outlier$$
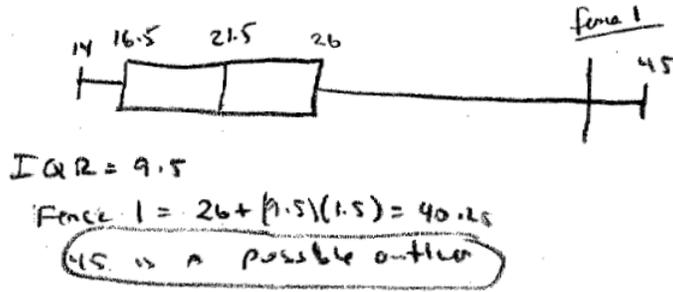
3. Rank the following correlation coefficients from weakest to strongest.

   .343,   -.318,   .214,   -.765,   0,   .998,   -.932,   .445

$$0, \ .214, \ -.318, \ .343, \ .445, \ -.765, \ -.932, \ .998$$

4. A poll was taken of 150 students at De Anza College. Students were asked how many hours they work outside of college. The students were interviewed in the morning between 8 AM and 11 AM on a Thursday. The sample mean for these 150 students was 9.2 hours.

   a. What is the Population?
      All students at De Anza College

   b. What is the Sample?
      150 students at De Anza College

   c. Does the 9.2 hours represent a statistic or parameter? Explain.
      Statistic – it is from sample data, parameters are from populations.

   d. Is the sample mean of 9.2 a reasonable estimate of the mean number of hours worked for all students at De Anza? Explain any possible bias.
      No, only morning students were asked, night students probably work longer hours.

5. If you were trying to think of factors that affect health care costs:

   a.   Choose a variable you believe would be positively correlated with health care costs.
   b.   Choose a variable you believe would be negatively correlated with health care costs.
   c.   Choose a variable you believe would be uncorrelated with health care costs.

(a) year, income level, Dr. salarys, age, etc

(b) poverty rate, hours exercising each week, etc

(c) number of pets, street address, etc

6. A researcher wanted to know if students who use the library at a college have higher GPAs than students who do not use the library. The researcher decided used a random number generator to choose 20 random classes at the college. Students in each of these classes were given surveys that could be filled out anonymously. Students that completed the surveys were given a $5 gift card for the bookstore. 82% of students in the sampled classes returned the surveys.

Here are the two questions of interest:

How often do you use the library?
    a. Never
    b. Less than once a week
    c. More than once a week, but not every day
    d. Every day

What is your current GPA? _____

    a. What method of sampling was used by the researcher?
    Cluster sampling

    b. Discuss the wording of the questions for possible bias.
    Seems ok, choices for library may be too limited.

    c. Is this an observational study or an experiment? Explain.
    Observational, no manipulated variable

    d. The researcher concluded that students who use the library more frequently have higher GPAs. Is this a valid conclusion for this type of study? Explain.
    Yes, sampling method is representative and no cause and effect is mentioned.

7. Identify the Steps of a Statistical Process for the library use/GPA example in problem 6. The steps are listed below:

    a. Ask a question that can be answered with sample data.
    Do students who use the library at a college have higher GPAs than students who do not use the library?

    b. Determine the information needed
    Library Use and GPA

    c. Collect sample data that is representative of the population.
    Cluster sampling done correctly is usually representative

    d. Summarize, interpret and analyze the sample data.
    no data given in study

    e. State the results and conclusion of the study.
   **Students who use the library more frequently have higher GPAs.**

8. A researcher for an electric car company was testing a new battery system. The goal of the battery system was to extend the life of the battery before recharging is necessary.

48 identical model electric cars were selected. 24 cars were given the new battery system (treatment group), while the remaining 24 cars kept the old system (control group). All cars were then fully charged. 24 drivers were then assigned a car. They were not told whether they were driving a car with the new batteries or a car with the regular batteries. The drivers were all given the same route to drive. The drivers drove the cars until the battery ran dead. The mileage driven was then recorded.

The 24 drivers then returned the next day to repeat the experiment with the remaining cars.

Each driver was assigned a new battery car and a regular battery car, but neither the driver nor the person assigning the car knew the order in which they drove the cars.

The results are shown in the box plot. The researchers concluded that new battery system did extend the life of the battery by about 7%.
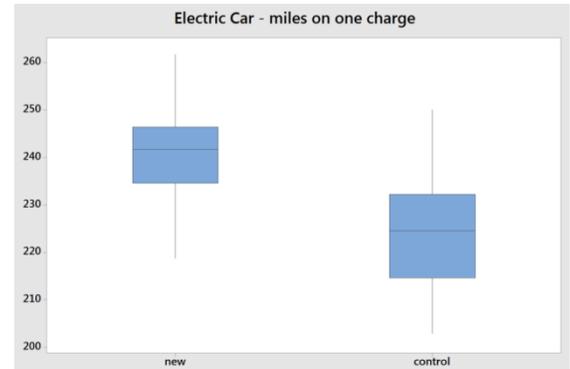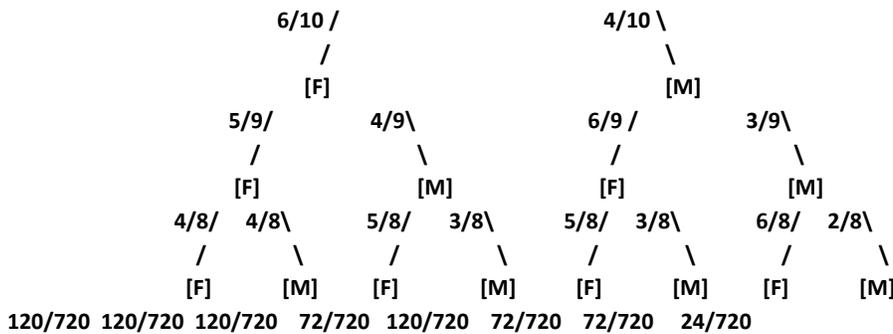
a. In this experiment, what is the explanatory variable and what is the response variable?
   Explanatory: battery system   Response: miles on one charge

b. Was there blinding done in this experiment? Explain.
   Yes, they were not told whether they were driving a car with the new batteries or a car with the regular batteries.

c. Suppose the researcher instead chose 48 drivers and each driver drove a single car. Would this create any lurking variables for the experiment?
   Yes, individual driver habits.

9. Identify the Steps of a Statistical Process for the multiple measures example in problem 8. The steps are listed below:

a. Ask a question that can be answered with sample data.
   Does the new battery system extend the life of the battery before recharging is necessary?

b. Determine the information needed.
   Battery system (new and control), miles on one charge.

c. Collect sample data that is representative of the population.
   Random assignment was use so sampling is representative. Results shown in box plots

d. Summarize, interpret and analyze the sample data.
   Box plots show that cars with new battery systems travel farther.

e. State the results and conclusion of the study.
   The new battery system extends the life of the battery by about 7%.

10. A student has a 90% chance of getting to class on time on Monday and a 70% chance of getting to class on time on Tuesday. Assuming these are independent events, determine the following probabilities:

   a. The student is on time both Monday and Tuesday.   **(.9)(.7)=.63**

   b. The student is on time at least once (Monday or Tuesday).   **.9+.7-.63 = .97**

   c. The student is late both days. **1-.97=.03   or (.1)(.3)=.03**

11. A class has 10 students, 6 females and 4 males. 3 students will be sampled without replacement for a group presentation.
   a. Construct a tree diagram of all possibilities (there will be 8 total branches at the end)

```
                6/10 /                          4/10 \
                    /                                \
                  [F]                                [M]
              5/9/        4/9\              6/9 /         3/9\
                /            \                 /             \
              [F]           [M]              [F]            [M]
          4/8/   4/8\    5/8/   3/8\     5/8/   3/8\    6/8/   2/8\
            /       \      /       \       /       \      /       \
          [F]      [M]   [F]      [M]    [F]      [M]   [F]      [M]
    120/720 120/720 120/720  72/720  120/720  72/720  72/720  24/720
```

   b. Find the following probabilities:
        i. All male students in the group presentation.
           24/720 = 1/30=.0333

        ii. Exactly 2 female students in the group presentation.
           (120+120+120)/720 = 360/720 = 1/2 = .5000

        iii. At least 2 female students in the group presentation.
           (120+120+120+120)/720 = 480/720 = 2/3 = .6667

12. 20% of professional cyclists are using a performance enhancing drug. A test for the drug has been developed that has a 60% chance of correctly detecting the drug(true positive). However, the test will come out positive in 2% of cyclists who do not use the drug (false positive).
   a. Construct a tree diagram where the first set of branches are cyclist with and without the drug, and the 2nd set is whether or not they test positive.
   b. From the tree diagram create a contingency table.

```
              /          \
       .20 /              \ .80
        (D+)              (D-)
       /  \              /  \
  .60 /    \ .40   .02 /    \ .98
   (T+)   (T-)    (T+)    (T-)
   .120   .080    .016    .784
```

	D+	D-	total
T+	.120	.016	.136
T-	.080	.784	.864
total	.200	.800	1.000

P(D+|T+) = .120/.136=88.2%

c. What percentage of cyclists will test positive for the drug?
    13.6%

d. If a cyclist tests positive, what is the probability that the cyclist really used the drug?
    .120/.136=88.2%

13. We wish to determine the morale for a certain company. We give each of the workers a questionnaire and from their answers we can determine the level of their morale, whether it is 'Low', 'Medium ' or 'High'; also noted is the 'worker type' for each of the workers. For each worker type, the frequencies corresponding to the different levels of morale are given below.

WORKER MORALE

Worker Type	Low	Medium	High	TOTAL
Executive	1	14	35	**50**
Upper Management	5	30	65	**100**
Lower Management	5	40	55	**100**
Non-Management	354	196	450	**1000**
TOTAL	**365**	**280**	**605**	**1250**

a. We randomly select 1 worker from this population. What is the probability that the worker selected

- is an executive? **50/1250 = 4%**

- is an executive with medium morale? **14/1250 = 1.12%**

- is an executive or has medium morale? **(50+280-14)/1250 = 316/1250 = 25.3%**

- is an executive, given the information that the worker has medium morale.
  14/280 = 5%

b. Given the information that the selected worker is an executive, what is the probability that the worker

- has medium morale? **14/50 = 28%**

- has high morale? **35/50 = 70%**

c. Are the following events independent or dependent? Explain your answer:

- is an executive', 'has medium morale', are these independent?
  not independent   P(E|M)=5% ≠P(E) = 4%

- is an executive', 'has high morale', are these independent?
  not independent P(H|E)=70% ≠P(H) = 605/1250 = 48.4%