Name:_____ Date:_____

SOLUTIONS AP Statistics Assignment #1 Exploring Data

1. Identify each study as being either observational or experimental:
   a. Subjects were randomly assigned to two groups, and one group was given an herb and the other group a placebo. After six months, the numbers of respiratory tract infections each group had were compared
   This is an experiment because 1) a treatment was given (Herb or no herb) and 2) Participants were randomly assigned into groups

   b. A researcher stood at a busy intersection to see if the colour of the automobile that a person drives is related to running red lights
   Observation: No treatment was given and no group assignment

   c. A researcher finds that people who are more hostile have higher total cholesterol levels than those who are less hostile

   Observation: No treatment was given and no group assignment

   d. Subjects are randomly assigned to four groups. Each group is placed on one of four special diets: low fat diet, high fish diet, a combination of low-fat diet and high fish diet, and a regular diet. After six months, the blood pressures of the groups are compared to see if diet has any effect on blood pressure

   This is an experiment because 1) a treatment was given (the four types of diets) and 2) Participants were randomly assigned into groups

2. Researchers analyzed standardized test results and television viewing habits of 1700 children. They found that children who averaged more than two hours of television viewing per day when they were younger than 3 tended to score lower on measures of reading ability and short term memory. *(Seattle Times, July 6, 2005)*
   a. Is this an "Observation study" or "Experiment"?
   This is an Observational study because no treatment was given and groups were not assigned

   b. Is it reasonable to conclude that watching two or more hours of television is the cause of lower reading scores? Explain?
   It is not reasonable to derive any cause and effect conclusions from an observational study.
   TO establish any cause and effect conclusions, we need to show that, with every other variable consistent, that:
   1) Having "event A" → "event B" occurs and also 2) Not having Event A → Event B does not occur. (This is when a control group is present).

3. A study on children allergies led researchers to conclude that babies raised with two or more animals were about half as likely to have allergies by the time they turned six. (*San Luis Obispo Tribune*, August 28, 2002) . What are the potential confounding variables that illustrates why it is unreasonable to conclude that being raised with two or more animals is the cause of the observed lower allergy rate?

   Genetics could come into play. If parents are not allergic to pets, they may be more likely to have pets. Also, if parents are not allergic to pets, their genes could be passed onto their kids so that they are also less likely to not be allergic to pets.

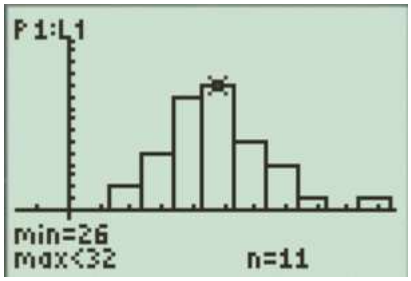
4. Observational studies in recent years have concluded that moderate drinking of alcohol (espescially red wine) is the cause of reduced risk of heart disease. What are some of the confounding variables that can prevent us from making this conclusion?

   One confounding variable is "wealth". Someone who is wealthy is able to afford drinking red wine on a regular basis. At the same time, being wealthy also allows them to afford better health care, leading to better health.

5. The amount of protein (in grams) for a variety of fast food sandwiches is reported here. Construct:
   i) a frequency distribution using 6 classes ii) A histogram with a relative frequency distribution
   Iii) an ogive

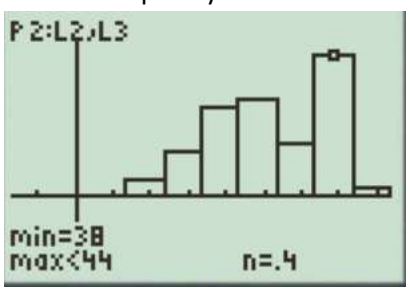
23	30	20	27	44	26	35	20	29	29
25	15	18	27	19	22	12	26	34	15
27	35	26	43	35	14	24	12	23	31
40	35	38	57	22	42	24	21	27	33

Frequency distribution:



X axis: amount of proteins in grams
Y -axis: Count

Relative Frequency Distribution



X axis: amount of proteins in grams
Y axis: Percentage

OGIVE

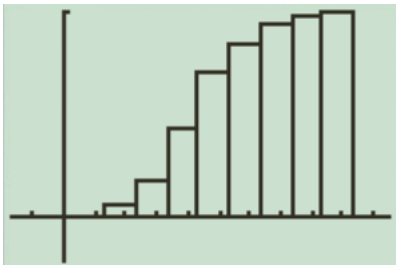


Table of frequency:

Range	Frequency	Relative Frequency	Cumulative Frequency
8 → 14	2	0.05	2
14 → 20	5	0.125	7
20 → 26	10	0.25	17
26 → 32	11	0.275	28
32 → 38	6	0.15	34

38 → 44	4	0.10	38
44 → 50	1	0.025	39
56 → 62	1	0.025	40

6. State which graph: (Bar, Pie, or Line graph would be the best for each situation)

   a. The number of students enrolled at a local college for each year during the last five years

      A line graph can be used to see the changes in enrollment over the past five years. Line graphs are good for recognizing trends/changes over time.
      A bar graph can also be used, however each of the past five years will need to be treated categorically.

   b. The budget for the student activities department at a certain college for each year during the last five years
      We can use a separate pie chart to see how the budget was allocated in each year.
      Another option is a segmented bar graph, where the bars are split into separate segments to illustrate the allocation of funds.

   c. The means of transportation the students use to get to school

      A Pie Chart can be used to illustrate the comparison of each type of transportation relative to all the students in the school

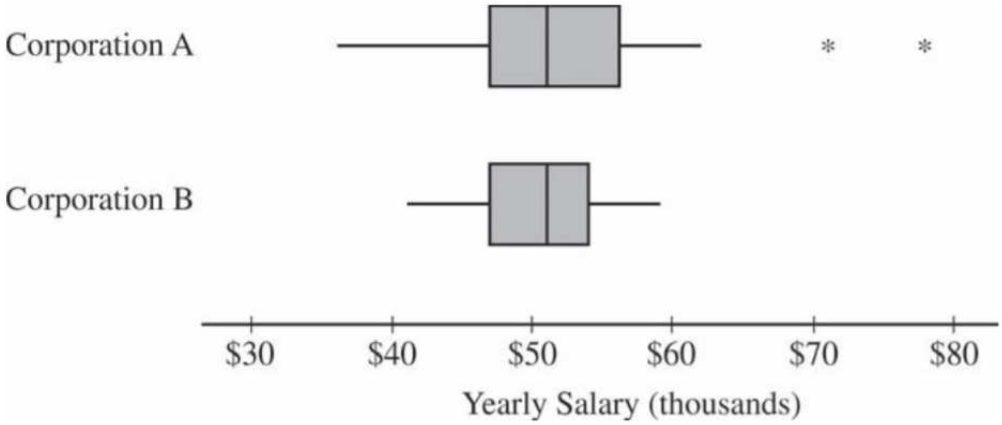
   d. The percentage of votes each of the four candidates received in the last election
      Pie Chart – better for making comparisons in the number of votes each candidate received relative to the total number of voters.

   e. The record temperatures of a city for the last 30 years
      If we are looking at progression over time, then a line graph would be ideal

   f. The frequency of each type of crime committed in a city during the year
      Pie Chart – a part will allow us to compare the number of times each crime was committed relative to all the crimes of that year.

Question7

Two large corporations, A and B, hire many new college graduates as accountants at entry-level positions. In 2009 the starting salary for an entry-level accountant position was $36,000 a year at both corporations. At each corporation, data were collected from 30 employees who were hired in 2009 as entry-level accountants and were still employed at the corporation five years later. The yearly salaries of the 60 employees in 2014 are summarized in the boxplots below.



(a) Write a few sentences comparing the distributions of the yearly salaries at the two corporations.

(b) Suppose both corporations offered you a job for $36,000 a year as an entry-level accountant.

(i) Based on the boxplots, give one reason why you might choose to accept the job at corporation A.

(ii) Based on the boxplots, give one reason why you might choose to accept the job at corporation B.

a) Centers are approx. equal ... medians ≈ $51K/year.
More spread in Corp. A ... IQR = 56-47 ≈ 9K vs. 53-47 ≈ 6K for Corp. B
Also, outliers for Corp. A, and overall lower values. Corp. B is less varied & closer to the median overall.
Corp B is slightly skewed left and corp. A is approx symmetric.
Corp A also had gaps b/w 62-70K & 70-75K.

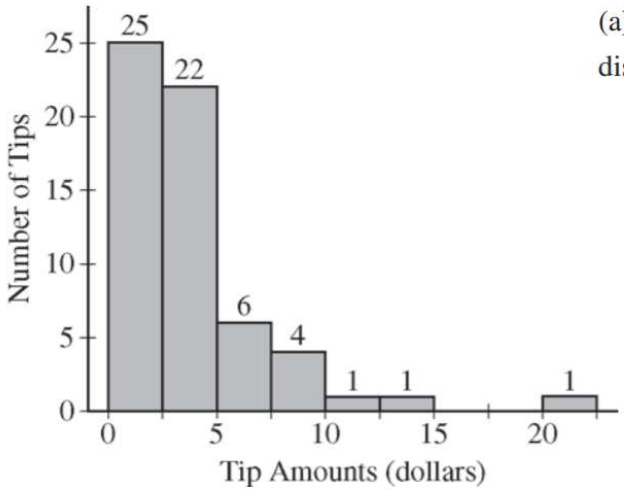
b) i) The upper quartile (75th %tile) is larger for Corp A., so I would have overall higher earning potential.

ii) Corp B has less spread and more consistent salaries. I would be more guaranteed to know what my salary would be b/c there are no extreme highs & lows like Corp A

GO ON TO THE NEXT PAGE.

Question 8

Robin works as a server in a small restaurant, where she can earn a tip (extra money) from each customer she serves. The histogram below shows the distribution of her 60 tip amounts for one day of work.



(a) Write a few sentences to describe the distribution of tip amounts for the day shown.

(b) One of the tip amounts was \$8. If the \$8 tip had been \$18, what effect would the increase have had on the following statistics? Justify your answers.

The mean:

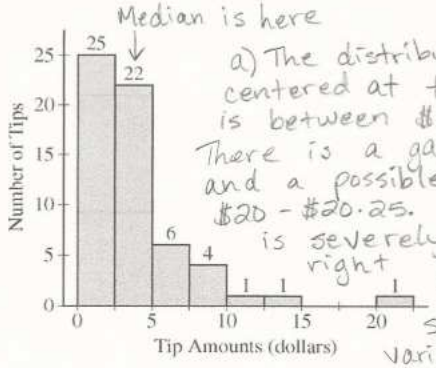
The median:

1. Robin works as a server in a small restaurant, where she can earn a tip (extra money) from each customer she serves. The histogram below shows the distribution of her 60 tip amounts for one day of work.

$\frac{60}{2} = 30$

So Median is between 30^{th} + 31^{st} tips

Median is here



a) The distribution of tips is centered at the median which is between \$2.50 and \$5.00. There is a gap from \$15 - \$20 and a possible outlier from \$20 - \$20.25. The tips distribution is severely skewed to the right and unimodal. There is a large spread and high variability with a range of \$22.50.

(a) Write a few sentences to describe the distribution of tip amounts for the day shown.

(b) One of the tip amounts was \$8. If the \$8 tip had been \$18, what effect would the increase have had on the following statistics? Justify your answers.

The mean: The mean would increase because it is highly affected by outlying data points. Since \$18 is greater than almost all the other tips it would increase the mean by $\frac{10}{60} \approx \$.17$.

The median:

The median would remain unaffected because it is resistant to outliers. Since the \$8 tip is above the median, changing it to \$18 would not change the position of the median tip.