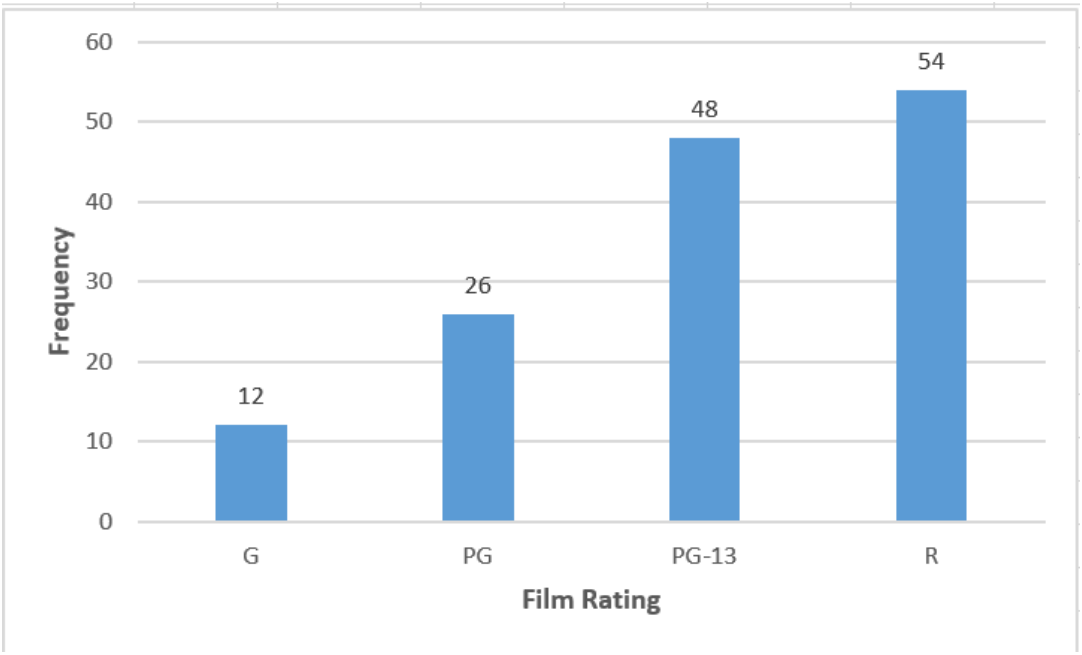
STAT 3090 – UNIT ONE REVIEW

Show relevant work for the following.

1. The following bar chart shows the film rating for 140 randomly selected films at local theaters.



What is the relative frequency of the films rated "PG"?

2. A campus survey at Clemson University found that 72% of those surveyed bought a national championship t-shirt. Is 72% is a parameter or statistic?

3. In a survey of 1300 American high school students, 32% of the respondents reported that someone had bullied them in school. What was the target population of this survey?

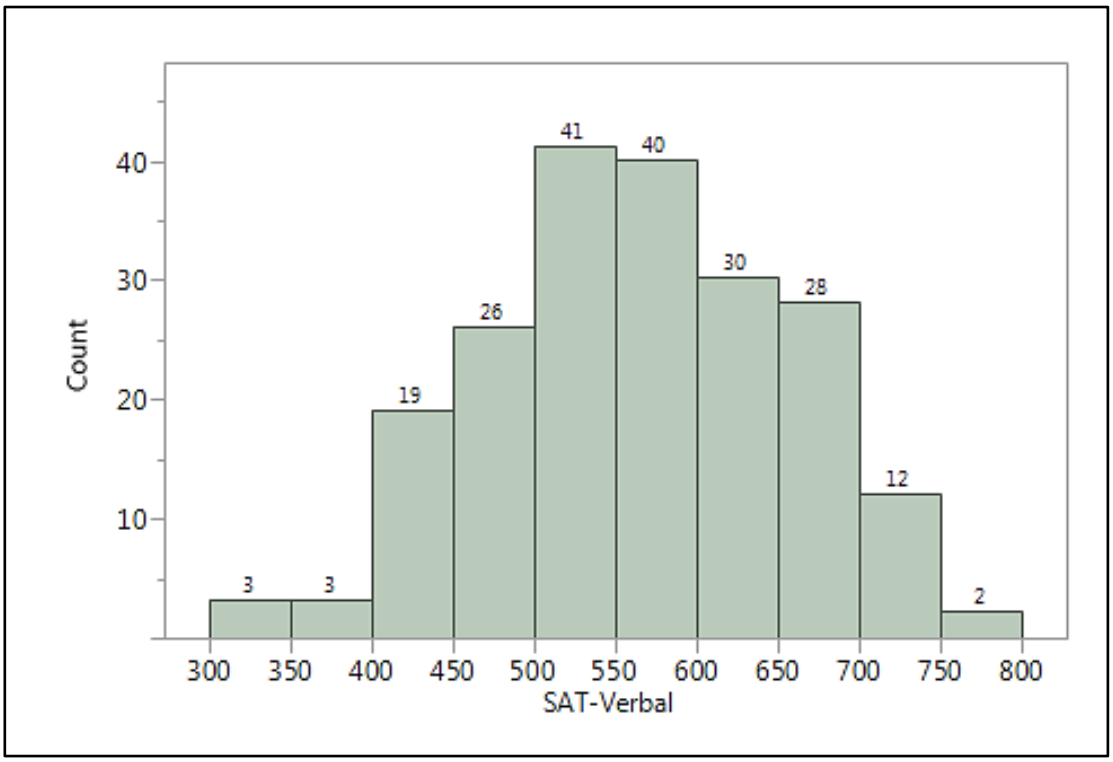
4. On Test 1, Devin computes his z-score using his test score, the mean test score and standard deviation. Devin's z-score on the test is 2. Write an interpretation of the z-score.

5. Classify the following as continuous or discrete and state the level of measurement.

the number of courses in which a Clemson student is currently enrolled

6. A. Approximately what percentage of data points in a data set are greater than the 38^{th} percentile value?
B. Approximately what percentage are less than the 38^{th} percentile?

7. The following histogram shows the Verbal SAT scores for 205 students entering a local college in the fall of 2016.



A. In what class is the median score?

B. In what class is the 30th percentile value?

8. A survey collects many variables of interest about the competitors at the 2014 Winter Olympics. Classify the following variables as qualitative or quantitative.

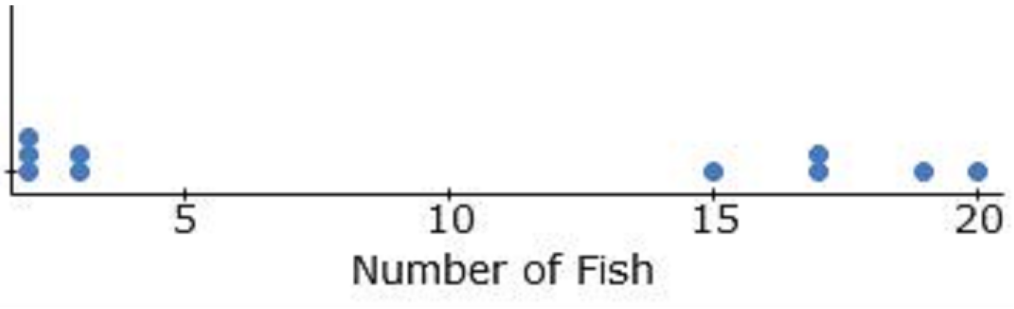
A. country of origin

B. age of the athlete

C. time spent traveling to the venue

D. resting pulse rate of the athlete

9. Andrew decides to go on a fishing adventure. He records the number of catches per day for 10 days and creates the following dot plot to illustrate his results.



A. What is Q1?

B. What is the median value?

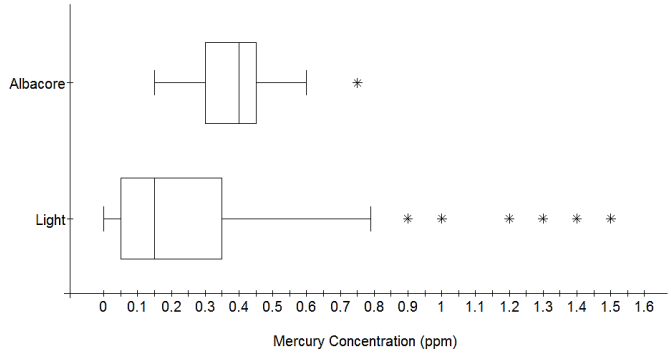
C. On what proportion of days did Andrew catch more than 15 fish?

D. Find the standard deviation of the number of fish caught. Interpret this value.

10. A data set comprised of 1000 measurements is bell-shaped with mean 142 kg. 95% of the data lies between 99 kg and 185 kg.
A. What is the standard deviation of these measurements?
B. What is the variance of these measurements?

11. The monthly average number of phone calls to a local police station is 1022 with standard deviation 130.
A. State the range in which 75% of the number of phone calls per month will reside.
B. State the range in which 88.89% of the number of phones calls per month will reside.

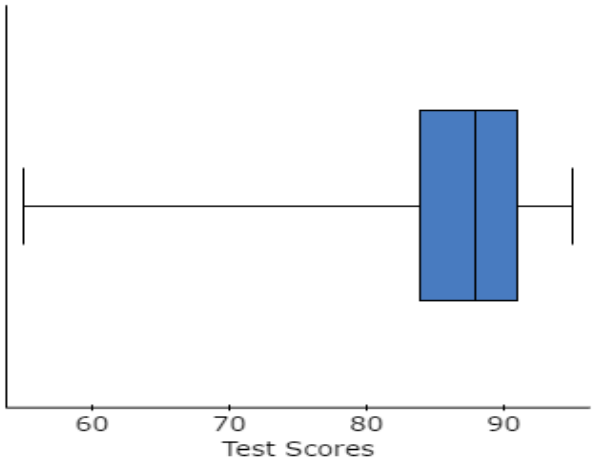
12. A study conducted by the Defenders of Wildlife examined the typical mercury concentration in cans of tuna sold in stores. The Defenders collected a sample consisting of 20 cans of albacore tuna and 144 cans of light tuna from stores across the United States. The boxplots below summarize the mercury concentrations found in the sampled cans (in parts per million, ppm).



A. What is the approximate value of the interquartile range of the distribution of mercury concentrations found in the sampled cans of light tuna?
B. Which distribution has greater variability?

C. Which type of tuna has a greater mercury concentration overall?

13. The following boxplot shows the distribution of scores on the first exam in a calculus class. Which of the following is true?



A. Describe the shape of this distribution.

B. Compare the mean and median score.

14. It was reported by the South Carolina Cattle Association that the weights yearling steers are symmetric and bell-shaped with a mean of 1152 pounds and standard deviation 84 pounds.
A. Approximately what percent of steers weigh ***more than*** 1,068 pounds?
B. What percent of steers weigh between 984 pounds and 1236 pounds?

15. The following data are the heights (in feet) of the 37 waterfalls in Greenville County described in the book *Waterfall Hikes of Upstate South Carolina* (King, 2008), listed in ascending order:

8	10	10	10	12	12	12	15	15	15	15	15	15	15	18	18	20	20	20
25	25	28	30	40	40	50	50	50	70	75	80	80	100	120	125	140	150	

A. Create a stem-and-leaf plot of this distribution.
B. Find the height that corresponds to the 60^{th} percentile value for these data.
C. Find the IQR value. State the meaning of this value.
D. Write a few sentences describing center, shape and spread of the distribution of these waterfall heights. Be sure to include appropriate summary measures for center and spread.

16. Your body's immune system protects you from disease and infection. If you have an autoimmune disease, your immune system attacks healthy cells in your body.A local newspaper published an article with the following headline.

Study Finds Strong Association Between Smoking and Autoimmune Disease

The article reported that a study tracked the medical histories of 15,016 men and women for 17 years. The article stated that for those who smoked at least two packs of cigarettes a day, the risk of autoimmune disease was 2.4 times the risk for those who did not smoke or smoked less than two packs per day.

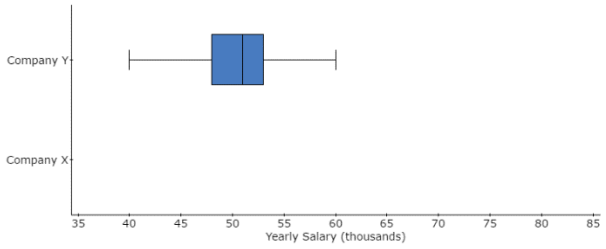
A. Identify the explanatory and response variables in the study.
B. Is the study described in the article an observational study or an experiment? Explain.
C. Exercise status (regular weekly exercise versus no regular weekly exercise) was mentioned as a possible confounding variable. Explain how exercise could be a confounding variable in the study. (what impact might it have on the study?

17. Two companies, X and Y, hire many new college graduates as managers at entry-level positions. In 2010, the starting salary for an entry-level manager was $36,000 per year at both companies. At each company, data were collected from 20 employees who were hired in 2010 as entry-level managers who were still employed 5 years later.

The lower quartile, median, and upper quartile for the salaries of the sample of 20 employees at Company X are shown in the table below, along with all values below the lower quartile and all values above the upper quartile. The values are given in 1,000s.

	Values Below Q_1	Q_1	**Median**	Q_3	**Values Above Q_3**
Company X	$36, $38, $39, 39	$47	$50	$55	$59, $59 61, $70, 80

A. Determine if there are outliers in the company X salary distribution. Justify your response clearly.
B. The boxplot of the 20 salaries for Company Y is shown below. Construct the boxplot for the Company X salaries on this same scale.



C. Compare the two distributions.