AMERICAN UNIVERSITY STAT 614 EXAM 1
Instructions: Submit all code comments and output in an Rmarkdown file and a Word File.

library(tidyverse)

1) Use and show R code that will produce a boxplot for the variable **price** from the diamonds data frame

2) Using your boxplot, estimate Q_1 , Q_2 , and Q_3 for the variable **price**

3) Does the boxplot indicate that the data for the variable **price** has outliers?

4) Use and show R code to indicate the maximum value of the variable **price**

5) Use and show R code to generate a histogram for the variable **price.**

6) Is the histogram that you generated for the variable **price** skewed left or right ?, Also, indicate which is larger for the histogram, *mean or median*.

7) Use and show Tidyverse R coding to create a scatter plot for the variables **carat** and **price** from the diamonds data frame. Let carat be explanatory **x** and let price be response **y.**

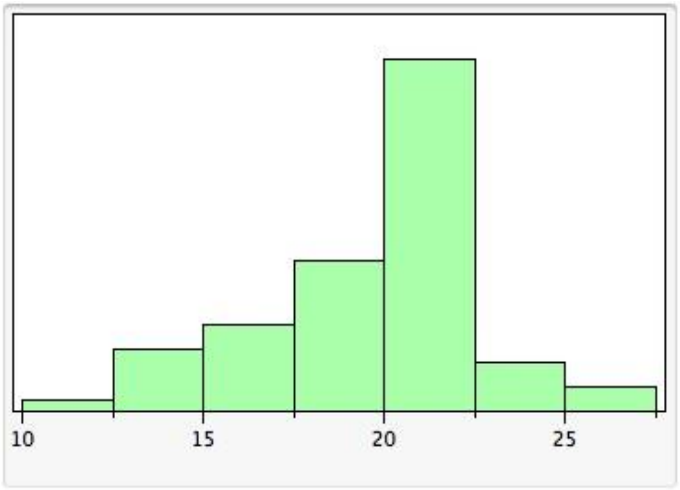
8) Comment on the relationship that the scatter plot shows for the two variables.

Use and show R coding to obtain correct answers/output for problems 9 and 10.

9) Test scores for a Calculus class at American University are normally distributed with a mean of 72 and a standard deviation of 5.5. Find the percentage of scores that are greater than 78.

10) Test scores for a Calculus class at American University are normally distributed with a mean of 72 and a standard deviation of 5.5. Find the 85^{th} percentile for the scores.

11. In the fuel efficiency study of 2007 compact model automobiles the following histogram of the distribution of the miles per gallon fuel efficiency rating in city driving (MPG-City) for automobiles manufactured in Europe was obtained:



From the histogram above, showing the distribution of MPG-City, we can see that

A) the shape of the distribution is roughly symmetric with one peak.
B) the distribution is skewed to the left.
C) the distribution is skewed to the right.
D) the distribution is roughly symmetric with outlier values to the left.
E) the shape of the distribution would be easier to see if a stemplot had been constructed instead of the histogram.

Answer:

12.
The Environmental Protection Agency records data on the fuel economy of many different makes of cars. Some of the variables they collect are listed below. Identify each variable as categorical or quantitative. Also classify a variable that is categorical as **nominal** or **ordinal.**

.

A) The manufacturer of the car (GM, Ford, Mazda, etc.).
B) Mileage of the car (miles per gallon).
C) Weight of the car (in lbs).
D) Size of the car (small, medium, full-size, pick-up truck, etc.).

Answer:

13. Data on the mileage of 20 randomly selected cars are listed below. The values are ordered for convenience.

12	13	15	16	16	17	18	18	19	19
20	20	22	23	24	26	26	27	27	29

What is the median mileage for these 20 cars?

A) 17.5 miles per gallon
B) 19 miles per gallon
C) 19.5 miles per gallon
D) 20 miles per gallon

Answer:

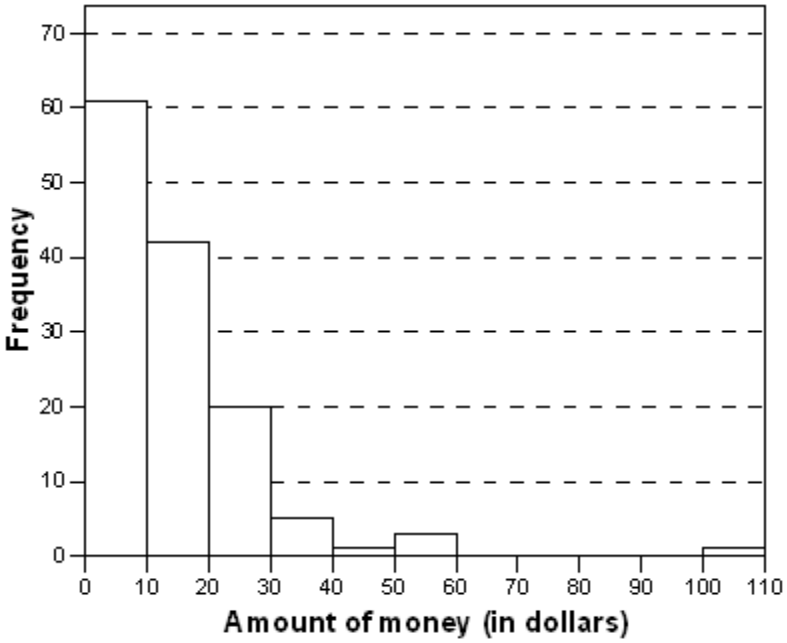
14. If the value 29 were misrecorded and should really be 21, what would the median mileage be for these 20 cars? (Use the same data set from #13)

A) It would change to 20 miles per gallon.
B) It would stay the same.
C) It would change to 21 miles per gallon.
D) We cannot determine this from the given information.

Answer:

Use the following to answer questions 15 and 16:

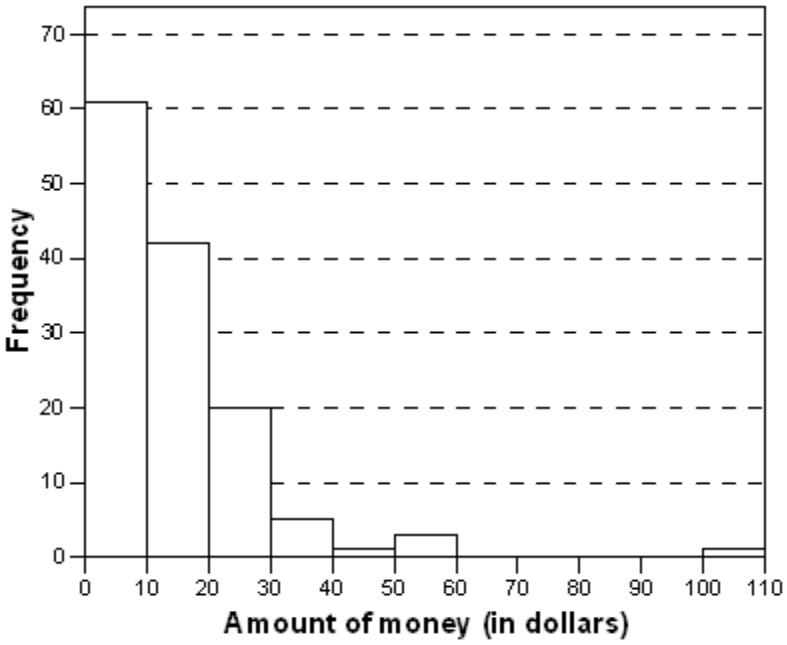
In a statistics class with 136 students, the professor records how much money each student has in their possession during the first class of the semester. The histogram shown below represents the data he collected:



The questions are on the next page.

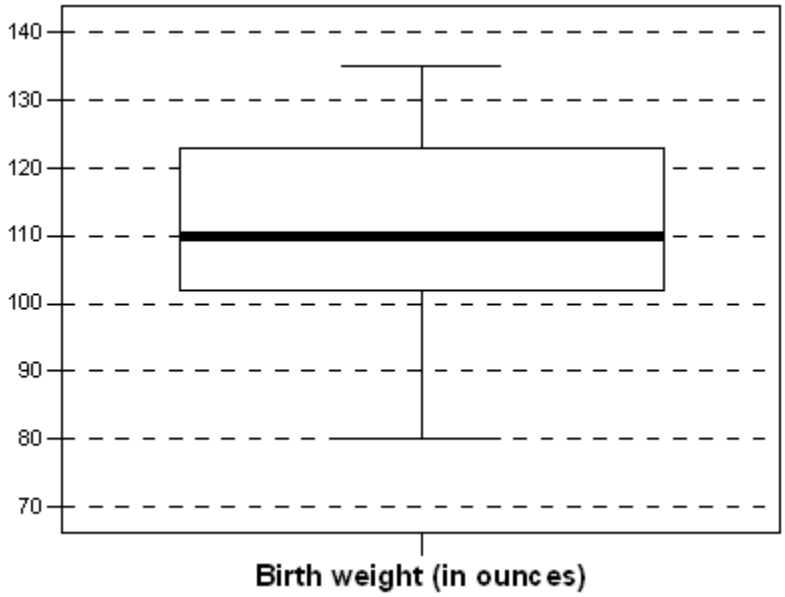
15. From the histogram, which of the following is true?
A) The mean is larger than the median.
B) The mean is smaller than the median.
C) The mean and median are approximately equal.
D) It is impossible to compare the mean and median for these data.
Answer:

16. What is the range of the data set?
A) 90
B) 100
C) 110
D) This cannot be determined from just the histogram.
Answer:



Use the following to answer questions 17–19:

The following boxplot is of the birth weights (in ounces) of 160 infants born in a local hospital:



17. Fill in the blank. The median birth weight is approximately _____.
A) 90 ounces
B) 100 ounces
C) 110 ounces
D) 120 ounces
Answer:

18. Fill in the blank. About 40 of the birth weights were below _____.
A) 92 ounces
B) 102 ounces
C) 112 ounces
D) 122 ounces
Answer:

19. Fill in the blank. Approximately _____ children had birth weights between 102 and 122 ounces.
A) 40
B) 50
C) 80
D) 100
Answer:

Use the following to answer questions 20 and 21:

Researchers wish to determine if a new experimental medication will reduce the symptoms of allergy sufferers without the side effect of drowsiness. To investigate this question, the researchers gave the new medication to 50 adult volunteers who suffer from allergies. Forty-four of these volunteers reported a significant reduction in their allergy symptoms without any drowsiness.

20. How could this study be improved?

A) Include people who do not suffer from allergies in the study in order to represent a more diverse population.

B) Repeat the study with only the 44 volunteers who reported a significant reduction in their allergy symptoms without any drowsiness, giving them a higher dosage this time.

C) Give the spouses of the volunteers a placebo.

D) Use randomization to divide the volunteers into 2 groups—one to receive the experimental medication and one to receive the placebo.

Answer:

21. What are the experimental units in this study?

A) The researchers.

B) The 50 adult volunteers.

C) The 44 volunteers who reported a significant reduction in their allergy symptoms without any drowsiness.

D) The six volunteers who did not report a significant reduction in their allergy symptoms without any drowsiness.

Answer:

22. A study was conducted with the object of determining the effect on blood pressure and cholesterol levels in women over the age of 60 of two different strength-training programs, one involving the use of a balance ball and the other using stationary exercise equipment. Eighty women over the age of 60 were identified and randomly divided into two equal-sized experimental groups. Each member of the two groups was trained and allowed to follow the prescribed routine for eight weeks. The subjects were aware of the purpose of the study but not that there were two different treatment groups involved. The cholesterol levels and blood pressure for subjects were measured at the start and at the end of the eight weeks. The experimenter and the medical technician were not told to which group any of the subjects were assigned. This is an example of an attempt to conduct

A) a matched-pairs design.

B) a design involving blocking.

C) a designed experiment that failed because it had systematic bias.

D) a double-blind randomized experiment.

E) a randomized experiment with two factors: type of exercise and type of medical measurement.

Answer:

23. In order to select a sample of undergraduate students in the United States, a simple random sample of four states is selected. From each of these states, a simple random sample of two colleges or universities is then selected. Finally, from each of these eight colleges or universities, a simple random sample of 20 undergraduates is selected. The final sample consists of 160 undergraduates. What sampling technique is being used?

A) Simple random sampling
B) Stratified random sampling
C) Multistage sampling
D) Convenience sampling

Answer:

Use the following to answer questions 24-26:

A call-in poll conducted by *USA Today* concluded that Americans love Donald Trump. This conclusion was based on data collected from 7800 calls made by *USA Today* readers.

24. What sampling technique is being used?

A) Simple random sampling
B) Stratified random sampling
C) Volunteer sampling
D) Convenience sampling

Answer:

25. *USA Today* reported that of the 7800 calls, 6435 calls were supportive of Donald Trump. This results in a percentage of 82.5%. Of the 6435 supportive calls, about half came from female callers. Which value(s) can be labeled as statistics?

A) 7800 and 6435
B) 6435 and 82.5%
C) 6435, 82.5%, and 50%
D) 82.5% and 50%

Answer:

26. *USA Today* later reported that 5640 of the 7800 calls for the poll came from the offices owned by one man, Cincinnati financier Carl Lindner, who is a friend of Donald Trump. What can we conclude about the results of this poll?

A) They are surprising, but reliable, because it was conducted by a nationally recognized organization.
B) They are biased, but only slightly because the sample size was quite large.
C) They are biased understating the popularity of Donald Trump.
D) They are biased overstating the popularity of Donald Trump.

Answer:

:

27. A market research firm has been asked to survey the population of people in a particular city who are 16 years of age or over with respect to their preferences for television programming. To do this the firm divides the list of the target population into five age groups: 16 to 25, 26 to 35, 36 to 45, 46 to 55, and 56 or older. From each of these age groups a simple random sample of 225 people is selected for a total sample of size 1125 individuals. Assume that all of these selected individuals respond honestly to the survey questions. The resulting sample is an example of what kind of sample design?

A) Multistage sample design.
B) Volunteer sample design.
C) Stratified random sample design.
D) Simple random sample design.
E) Systematic random sample design.

Answer:

28. A large university wishes to determine the percentage of its students that have committed some form of academic dishonesty, such as cheating on an examination or plagiarism on assignments during their academic career. To determine the desired percentage, a random sample of their current students is selected. Each selected student is then interviewed by a staff member and asked if they had cheated. The results of this survey likely will be unreliable because

A) the students likely will refuse to answer the questions.
B) those students who answer the questions may not do so honestly.
C) the interviewer being a staff member may be intimidating and hence there may be response bias.
D) All of the above are reasons for concern.
E) For this purpose, the sample results will be reliable since the students were selected purely by chance.

Answer:

29. Test scores for a Chemistry class were recorded below. Use the 1.5IQR rule to determine if you have outliers. **Show all of your work and steps and also use and show R code to find Q_1 , Q_3 , and the IQR.**

Chemistry Test Scores = {82,80,75,77,60,16,56,73,19,88,12,77,73,70,9,98,96}