QUANTITATIVE METHODS 830:200:H1

SUMMER 2020

Midterm

Name

_____

Show your work – all questions are one point unless otherwise noted

---

You participate in a fishing tournament. Here is a summary of the weights (expressed in pounds) of all the fish caught:

Min	Q1	Median	Q3	Max
2.3	2.8	3.0	3.3	4.5

1. Would you expect the mean to be higher or lower than the median?

   A. Lower

   B. Higher

2. You caught 3 fish whose weight was (a) 2.3 pounds, (b) 3.9 pounds, and (c) 4.2 pounds. Were any of these outliers (according to summary above)?

   A. (b) and (c) are outliers
   B. There are no outliers
   C. (c) is an outlier
   D. (a) is an outlier

The boxplots show mileage per gallon (mpg) for three different types of cars (Large, Truck, SUV).

3. Which type offers the best value of mpg?

   A)Large B)Truck C)SUV

4. Which type has the highest median mpg?

   A)Large B)Truck C)SUV

5. Which type has the smallest range of mpg?

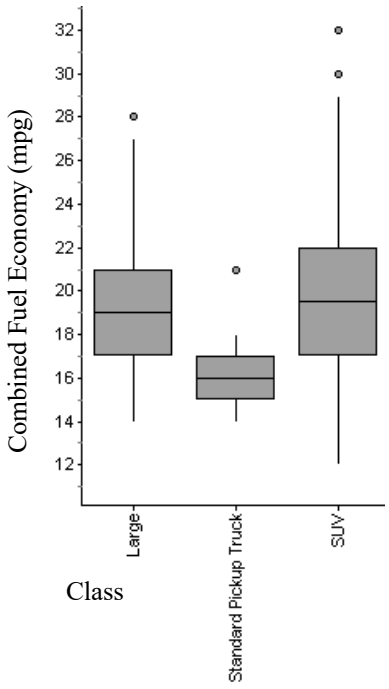
   A)Large B)Truck C)SUV

6. Which type has the smallest IQR?

   A)Large B)Truck C)SUV

7. Which type generally gets the best mileage?

   A)Large B)Truck C)SUV



A survey asked students about their preferred means of transporation to school.

Some came with the local bus, some got a ride with parents or friends, and others used their own means of transportation – they biked, skateboarded, or walked. See the summary table.

	Male	Female	Total
Bus	30	34	**64**
Ride	37	45	**82**
Personal	19	23	**42**
Total	**86**	**102**	**188**

Find each percent.

8.What percent are females who took the bus?
A)18.1% B)53.1% C)17% D)33%

9.What percent of the females take the bus?

A)18.1% B)53.1% C)17% D)33%

10.What percent of the bus riders are females?
A)18.1% B)53.1% C)17% D)33%

11.What is the marginal distribution of gender?

A) 30 males and 34 females

B) 30/37/19 and 34/45/23 respectively

C) 86 males and 102 females

D) 64 bus, 82 ride, and 42 personal transportation

12. Is their choice of means of transportation independent of gender?
A) Yes
B) No

The lifespans of a certain type of appliances are approximately normally distributed, in days, as N(620, 82). The company is deciding which warranty to use.

13. If the warranty is set at 540 days, what percentage of appliances will need to be replaced? (2 POINTS)
A) 16.45%
B) 32.90%
C) 48.50%
D) 65.80%

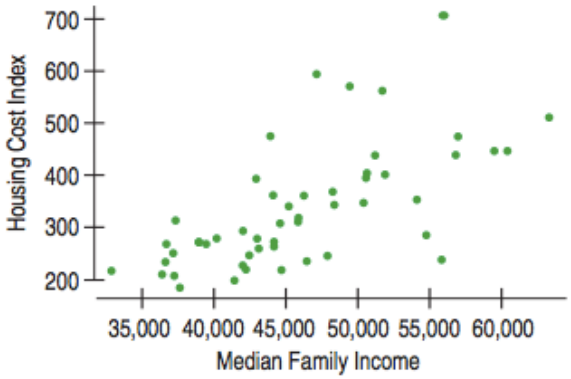
14. They refuse to replace more than 1% of the appliances sold. How many days should they set for the warranty?
(2 POINTS)
A) 365 days
B) 90 days
C) 429 days
D) 54 days

15. They still want a warranty of 540 days, and still refuse to replace more than 1% of the appliances sold. They think they could reduce the standard deviation of the appliances' lifespans. What standard deviation would be needed?
(2 POINTS)
A) s = 54 days
B) s= 34.4 days
C) s= 15.2 days
D) s= 90 days

Here is a graph of an index of *Housing Cost* (HCI, y-axis) and of *Median Family Income* (MFI, x-axis) for each of the contiguous American states. Consider the following:

r = 0.65.

HCI: mean = 338.2, s= 116.55
MFI: mean = $46,234, s = $7072.47



16. Can we use a regression analysis?
A) Yes B) No

17. Write the equation predicting HCI from MFI (3 POINTS)
A) HCI = -0.0107 + 156.50MFI
B) HCI = -156.50 + 0.0107MFI
C) HCI = -116.55 + 0.0107MFI
D) HCI = -338.20 – 116.55MFI

18. What would be the y-axis index's *predicted value* if a state has a median family income of $44,993?
A) 247.56
B) 324.93
C) 112.20
D) 289.67

19. The District of Columbia has a median income of $44,993 and an *actual* index of 548.2. How far off is the prediction you got from the previous question from the actual index?
A) it's 223.09 too low
B) it's 345.00 too high
C) it's 324.93 too low
D) it's 267.8 too high

You are taking a statistics exam. The exam is composed of five problems. They are multiple choice, so there are four possible answers for each problem. *If you just answer at random*:

20. what is the probability that none of the problems has a correct answer?
A) 0.5668
B) 1.3456
C) 0.1567
D) 0.2373

21. what is the probability that all of the problems have a correct answer?
A) 0.00098
B) 0.0098
C) 0.0039
C) 0.00024

22. what is the probability to get at least one of the problems wrong?
A) 0.870
B) 0.952
C) 0.999
D) 0.237

23. what is the probability to get your first incorrect answer on the fourth problem?
A)0.1406
B) 0.0351
C) 0.0263
D) 0.0117

---

24. A dataset has been collected from 50 male factory workers. Which variable is categorical?

A) eye color B) head circumference C) hours of work last week
D) number of steps walked daily E) number of devices at home

25. Which one is quantitative?

26. Which one is most likely to follow a Normal distribution?

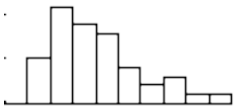
---

27. We are now working on a dataset collected from 30 male and 20 female factory workers. The mean number of hours worked for the men was 6, and for the women was 9.
The overall mean number of hours worked is …

A) is 6.5 B) is 7.2 C) is 7.5 D) is none of these E) cannot be determined.

---

28. I collected some data. I draw a histogram to check out the distribution. Which of the following is correct?



I. The histogram is skewed to the right.
II. The median is probably bigger than the mean.
III. It's a good idea to use mean and standard deviation.

A) I only B) II only C) I and II D) II and III E) I, II, and III

---

29. A Normal model describes the acidity level (pH) of household water. You tested the water in your own house, and its z–score was 1.8. This means that your household water's acidity …

A) had a pH of 1.8.
B) had a standard deviation of 1.8
C) its pH was 1.8 above than average household water.
D) its pH was 1.8 times that of average household water.
E) its pH was 1.8 standard deviations above than that of average household water.

---

30. If two variables have a zero correlation, that means that

A) our calculation is wrong
B) there is no association between them
C) there is no linear association between them
D) using a logarithmic transformation will generate a linear association under any condition
E) None of the above.

Match each statement with the value of r most likely to describe the association

31. Association between how many hours you study and score on the exam
32. Association between how many siblings you have and your GPA score.
33. Association between hours spent practicing a certain task and the minutes needed to complete it.
34. Association between the alcohol level in your blood and your reaction time in seconds

A. - 0.78
B. 0.73
C. 0.46
D. 0.13

.

---

35. Why do we look at a residual plot?

I. To check the appropriateness of the regression model
II. It could reveal a pattern in the data that was difficult to detect in the original graph.

A) I only B) II only C) I and II D) Neither I nor II

---

36. I am analyzing a scatterplot. I think that a point might be influential. This may be true if…

A) It has a small residual.
B) It is extreme in the vertical direction.
C) It is extreme in the horizontal direction.
D) It will modify the regression equation's slope
E) none of these

---

37. A phone survey asked a random sample of 1,028 U.S. boys and asked these teens to name their favorite movie from 2004. *Napoleon Dynamite* had the highest percentage with 8% of teens ranking it as their favorite movie. Which is true?

I. The population of interest is all U.S. teens.
II. 8% is a statistic and not the actual percentage of all U.S. teens who would rank this movie as their favorite.
III. This sampling design should provide a reasonably accurate estimate of the actual percentage of all U.S. teens who would rank this movie as their favorite.

A) I only B) II only C) III only D) I and II E) I, II, and III

---

38. More dogs are being diagnosed with thyroid problems than have been diagnosed in the past. A researcher identified 50 puppies without thyroid problems and kept records of their diets for several years to see if any developed thyroid problems. This is a(n)

A) randomized experiment B) survey C) prospective study
D) retrospective study E) blocked experiment

---

39. Which is true about bias?

I. it results from random variation and it's unavoidable.
II. it appears when we use a sampling method producing samples that are not representative
III. it is usually reduced as the sample size grows larger.

A) I only B) II only C) III only D) I and III only E) II and III only

---

40. Which statement is correct?

I. An event is composed of one or more outcomes from a specific trial
II. A trial is composed of many events
III. An outcome is composed of many trials
IV. An outcome is the value of a trial
V. The sample space is a collection of some events

A) I only B) II only C) III only D) I and IV only E) I and V only