

Statistics 201 Exam 1 – Practice Exam  
Fall 2021  
Chapters 01 - 06 including Pivot tables

Disclaimer:

This practice exam is provided solely for the purpose of familiarizing you with the format and style of the Stat 201 exams. There is no explicit or implicit guarantee that the upcoming exam will ask similar questions. If you use the practice exam as your only tool to help you prepare for the upcoming exam, you most likely will not do well on the exam. You should still do the things you would have done if you did not have access to this practice exam, such as re-read the text, go over your class notes, re-work the online homework problems, and look at the list of exam topics provided and make sure that you understand all the concepts listed within it.

NOTE: These questions come from actual older exams. As such, the points on this practice exam may total more than 100 points.



\_\_\_\_\_Score

1. A study of potential age discrimination considered promotions among middle managers in a large company.

	Age		Total
	Up to 39	40 and Over	
Promoted	38	43	81
Not Promoted	82	88	170
Total	120	131	251

- i) (2 points) What percent of employees were promoted?
- ii) (2 points) What percent of employees up to age 39 were promoted?
- iii) (2 points) What percent of employees were 40 and over?
- iv) (2 points) What percent of promoted employees were 40 and over?

2. The demand for bottled water increases during hurricane season in Florida. The number of 1-gallon bottles of water sold for a random sample of  $n=9$  hours in one store during hurricane season is:

64, 80, 74, 85, 82, 63, 67, 65, 75

- i) (3 points) What is the average hourly number of 1-gallon bottles of water sold?
- ii) (3 points) What is the median number of 1-gallon bottles of water sold?
- iii) (2 points) Partial JMP output is provided below. From this output, calculate the IQR of these data.

Quantiles		
100.0%	maximum	85
99.5%		85
97.5%		85
90.0%		85
75.0%	quartile	81
50.0%		
25.0%	quartile	64.5
10.0%		63
2.5%		63
0.5%		63
0.0%	minimum	63

3. Suppose the number of minutes a customer is put on hold at a particular software company's technical support call center is approximately normally distributed, with a mean of 4.9 minutes and a standard deviation of 1.4 minutes. Betty called this call center and was on hold for 7.5 minutes.

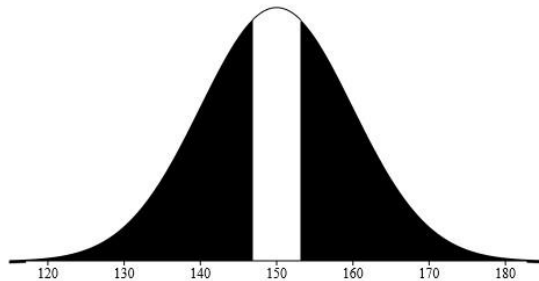
i) (2 points) Calculate the z-score for Betty's hold time.

ii) (3 points) Interpret the number you calculated in part (i) above. Don't comment on the magnitude of this number, rather explain what this number means. [Note: If you have no answer for part (i), use 2.34, which is NOT the correct answer to part (i)]

iii) (2 points) Another customer (David) called the same technical support call center and had a hold time that had a z-score of -0.55. The percent of customers that wait longer than David is approximately (circle the best answer):

- a) 97.5%
- b) 71%
- c) 50%
- d) 29%
- e) 2.5%
- f) 0.3%

4. (3 points) The reaction time of a professional hockey goalie was measured many times, and the distribution of these reaction times was approximately normally distributed, with an average of 150 milliseconds and a standard deviation of 10 milliseconds. What is the 75<sup>th</sup> percentile of this hockey goalie's reaction times? Use the correct screen shot below from our Normal Curve Calculator to answer this question (**hint**: only one of the images below can help you answer this question).



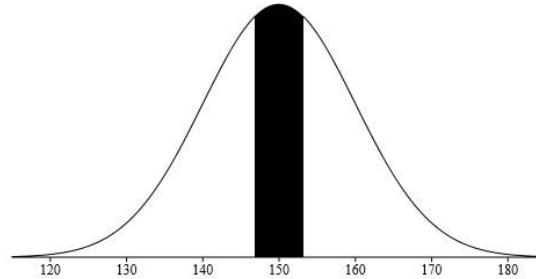
- ☐ Area from a value (Use to compute p from Z)  
☒ Value from an area (Use to compute Z for confidence intervals)

Specify Parameters:

Area 0.75  
Mean 150  
SD 10

Results:

Recalculate  
☐ Above  
☐ Below  
☐ Between  
☒ Outside 146.818 and 153.182



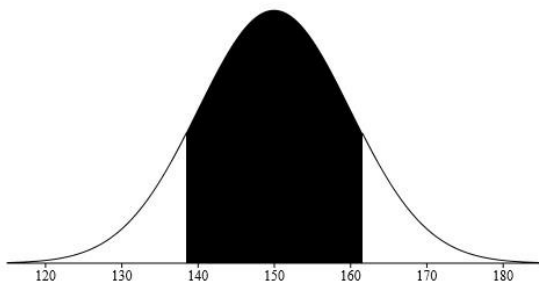
- ☐ Area from a value (Use to compute p from Z)  
☒ Value from an area (Use to compute Z for confidence intervals)

Specify Parameters:

Area 0.25  
Mean 150  
SD 10

Results:

Recalculate  
☐ Above  
☐ Below  
☒ Between 146.818 and 153.182  
☐ Outside



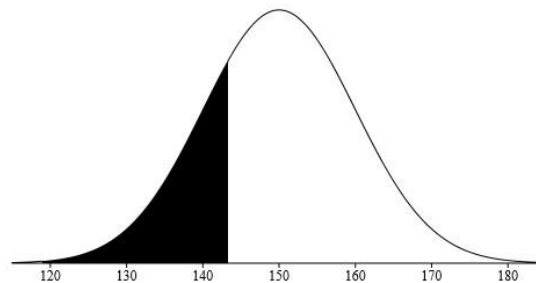
- ☐ Area from a value (Use to compute p from Z)  
☒ Value from an area (Use to compute Z for confidence intervals)

Specify Parameters:

Area 0.75  
Mean 150  
SD 10

Results:

Recalculate  
☐ Above  
☐ Below  
☒ Between 138.496 and 161.504  
☐ Outside



- ☐ Area from a value (Use to compute p from Z)  
☒ Value from an area (Use to compute Z for confidence intervals)

Specify Parameters:

Area 0.25  
Mean 150  
SD 10

Results:

Recalculate  
☐ Above  
☒ Below 143.258  
☐ Between  
☐ Outside

5. A particular brand of cell phone has a battery that will last, under normal conditions, 2.5 years on average, with a standard deviation of 0.32 years. The distribution of these lifetimes is approximately normally distributed.

i) (2 points) Calculate the z-score for a battery that lasts 2 years under normal conditions.

ii) (5 points) What percent of this brand of batteries can be expected to last 2 years or more under normal conditions?

Without the normal curve calculator, you can't give an exact answer here. Base your answer on the 68-95-99.7 rule, and your answer to part [i] above. Give as narrow of an interval as you can that contains the exact answer. Fill in the blanks below. Also, show your work (and/or reasoning) below.

- (Note 1: if you have a calculator that could give you the exact answer, DO NOT use that capability! Use only the 68-95-99.7 rule.)
- (Note 2: if you have NO answer for part [i] above, use -1.2 as your answer there, which is NOT the correct answer to part [i]).

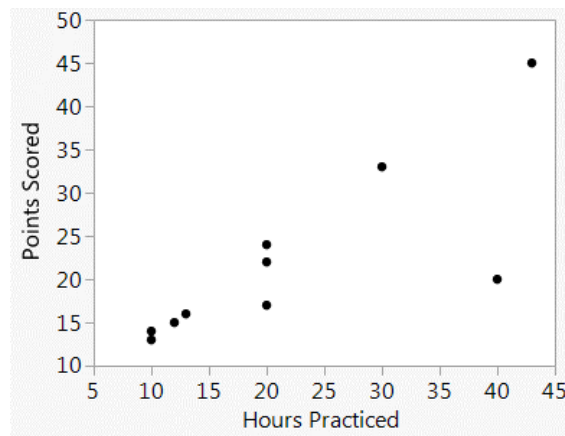
“The exact answer must be between \_\_\_\_\_% and \_\_\_\_\_%.”

6. (4 points) A gas station is trying to increase food sales inside the store. Many customers pay at the pump and leave. The gas station begins collecting data hoping to discover something that will bring customers inside after they pump their gas. A random sample of the data set is shown below. Above each of the eight columns, write a **C** if that variable is categorical, **Q** if that variable is quantitative, or **I** if that variable is an identifier variable.

Transaction number	Type of gas	Number of gallons	Pay at pump?	Inside food sale?	Type of payment	Gas purchase in dollars	Day of week
9853	premium	22	y	n	Visa	85.58	Mon
9211	diesel	26	y	n	Am Exp	110.5	Tues
8875	regular	19	y	y	Visa	70.11	Tues
8824	regular	21	y	y	Visa	77.49	Fri
8313	regular	14	y	y	MasterCard	51.66	Wed
7699	premium	22	n	n	cash	85.8	Wed
7645	diesel	45	y	y	Am Exp	191.25	Sat
3145	diesel	38	y	n	Am Exp	161.5	Sat
2588	regular	17	n	y	Visa	62.73	Sun
2499	regular	22	n	n	cash	81.18	Sat
2325	premium	15	y	n	MasterCard	58.35	Fri
2291	diesel	22	y	n	MasterCard	92.4	Mon
2078	regular	14	y	y	Visa	51.66	Thur
1843	regular	35	y	n	Visa	129.15	Thur
2103	regular	25	n	y	cash	92.25	Sat



7. A high school football coach believes that the number of points scored in each game is correlated with how many hours his players spend practicing the week before. He records the hours practiced before 10 games and the points his team scored in those games.



- i) (6 points) There are 3 conditions that must be checked before calculating a correlation coefficient ( $r$ ). List each condition, and briefly comment on whether or not each condition is met in this case.

Condition 1 is:

Is condition 1 met? (Circle one):      Yes    No  
Briefly explain your answer:

Condition 2 is:

Is condition 2 met? (Circle one):      Yes    No  
Briefly explain your answer:

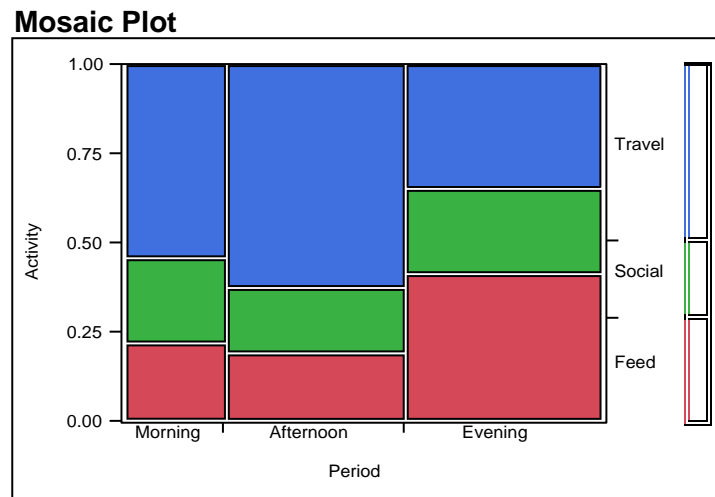
Condition 3 is:

Is condition 3 met? (Circle one):      Yes    No  
Briefly explain your answer:

Question 7 (continued)

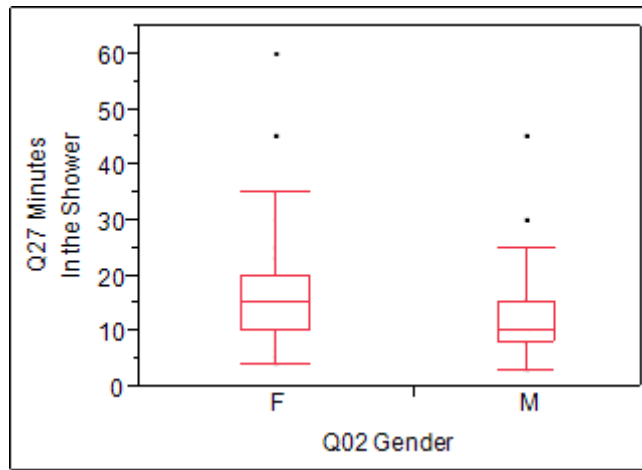
- ii) (2 points) Suppose that the correlation coefficient,  $r$  is calculated using all of the data in the scatter plot. In one week, the players practiced 40 hours and scored 20 points. How would the removal of this point change the values of  $r$ ? Circle the best answer:
- a)  $r$  would be unchanged
  - b)  $r$  would get closer to  $-1$
  - c)  $r$  would get closer to  $0$
  - d)  $r$  would get closer to  $+1$
  - e) The impact on  $r$  cannot be determined

8. Are Icelandic dolphins' behavioral activities related to a specific period of the day? Icelandic marine biologists observed the behavior (categorized into 3 behaviors: Travel, Social, Feeding) and time of day (categorized into 3 distinct periods: Morning, Afternoon, and Evening) of 1200 dolphins. They hoped to observe a relationship between these two variables over the course of 2 years. Below is a mosaic plot of the collected data:



- i) (2 points) Based on the mosaic plot, approximately what percent of these dolphins' mornings were spent traveling?
- ii) (2 points) Based on the mosaic plot, at what time of the day were the Icelandic dolphins observed the least? Briefly explain how you came to this conclusion.
- iii) (1 point) Is there a relationship between the variables Activity and Period? CIRCLE your answer.
- Relationship
No Relationship
- iv) (3 points) Make reference to the mosaic plot to justify your choice above.

9. Who takes longer showers? The following boxplots show the shower times in minutes for Stat 201 students, as reported in a previous survey.



- i) (2 points) The distribution of the male shower times is (Circle the BEST answer)
- a) Right-Skewed
  - b) Left-Skewed
  - c) Symmetric
  - d) Unimodal
  - e) Bimodal
- ii) (2 points) What is the approximate range of shower times for these females?
- iii) (2 points) Which group has a higher median shower time? CIRCLE the best answer.
- males      females      cannot be determined      the median times are the same
- iv) (2 points) 75% of males spend less than \_\_\_\_\_ minutes in the shower.
- v) (2 points) What is the IQR (approximately) for the female shower times?

10. (2 points) Adding an outlier in a scatterplot will do which of the following to the correlation coefficient? (Circle the best answer)

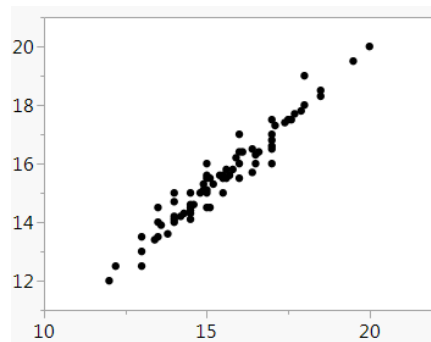
- a) Increase the correlation coefficient
- b) Decrease the correlation coefficient
- c) Change the correlation coefficient from positive to negative
- d) Change the correlation coefficient from negative to positive
- e) All options above are possible

11. (2 points) When the correlation coefficient is close to +1, it indicates (Circle the best answer)

- a) changes in one variable cause changes in the other, but we don't know which one causes the other to change.
- b) changes in the variable on the x-axis cause changes in the variable on the y-axis.
- c) changes in the variable on the y-axis cause changes in the variable on the x-axis.
- d) nothing regarding the possible cause and effect relationship between the two variables.

12. (2 points) What is the most likely correlation coefficient for the scatterplot below?

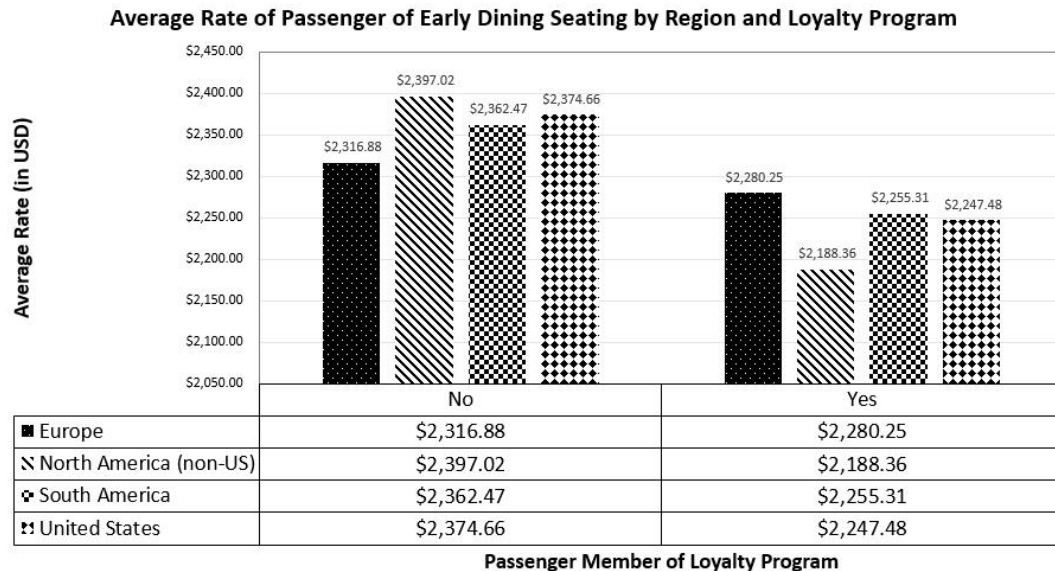
- a) 1.00
- b) 0.98
- c) 0.50
- d) 0.00
- e) -0.50
- f) -0.98
- g) -1.00



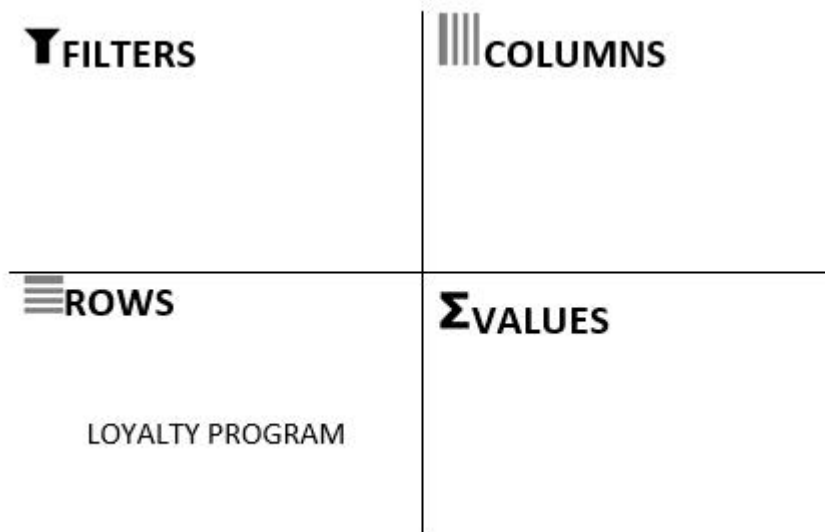
13. (3 points) Which of these statements is inaccurate (Circle the best answer)?

- a) Large outliers can inflate the mean
- b) The IQR is a measure of distance between 75<sup>th</sup> and 25<sup>th</sup> percentile
- c) When measuring variation, the standard deviation should be used for normal data and IQR should be used for skewed data
- d) A data set skewed to the right will generally have a median below the mean
- e) The median is the average of all the data values

14. Below is a pivot table of the average rate (**Rate**) paid by a passenger by the region (**Region**) and the passenger's participation in the ACL Loyalty Program (**Loyalty Program**). In addition, the passengers in the late seating for dining (**Dinner Seating**) have been filtered out.



- i) In the grid below, write in the name of the variable in the appropriate place which would create the pivot table above. The variables names are in **bold**, in the paragraph above. For example, the variable **Loyalty Program** has been added.



- ii) Briefly explain, one inference from the pivot chart. Be sure your inference has numerical support from the graph.

True/False Questions  
Circle the best answer  
(2 points each)

- T   F   The normal model can be used to describe all distributions.
- T   F   An observation with a z-score of 0.5 would be considered an outlier.
- T   F   If all of the data in a skewed right data set are converted into z-scores, the distribution of these z-scores will be normally distributed.
- T   F   The middle 50% of a data set will be within two standard deviations of the mean for normally distributed data.
- T   F   When the correlation is above 0.95 or below -0.95, it is safe to conclude that changes in X cause changes in Y.
- T   F   Bar charts are used to graphically display both quantitative and categorical variables.
- T   F   When doing a Pivot Table in Excel, the default value displayed is the SUM of the selected variable.