

**Sample Exam 1 - STAT 303 Session 201**  
**Summer 2009**

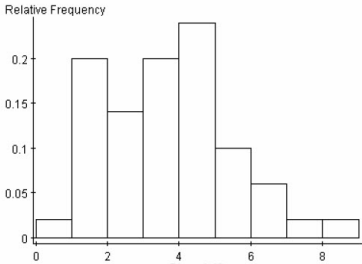
**Name:**

**UIN:**

**Signature:**

1. Do not open this test until told to do so.
2. Turn in your exam with your answers circled when you are done with the exam. You should not take the exam with you.
3. This is a closed book examination. You may use one one-sided sheet of formulas that you have brought with you. You should have no other printed or written material with you on the exam.
4. You have 60 minutes to work on this exam. There are 20 multiple choice questions, each worth 5 points.
5. You may use a calculator but not a phone during the exam.
6. If you are unsure of what a question is asking for, do not hesitate to ask the instructor or course assistant for clarification.
7. Do not sit directly next to another student.
8. Good Luck!!!

1. Boxplots are the best numeric graphs that we have discussed for
  - (a) determining if there are outliers
  - (b) comparing distributions
  - (c) finding the mean
  - (d) All of the above are true.
  - (e) *Only two of the above are true.*
  
2. Suppose you are trying to see if there is a linear relationship between the two numerical variables,  $X$  and  $Y$ . You find that the correlation  $r = 0.006$ . What can be concluded from this?
  - (a) *There is no linear relationship between  $X$  and  $Y$ .*
  - (b) There is a strong linear relationship between  $X$  and  $Y$ .
  - (c)  $X$  and  $Y$  are the same variable.
  - (d) There may or may not be a linear relationship between  $X$  and  $Y$ . You would need to see a scatterplot before coming to a conclusion about the relationship between  $X$  and  $Y$ .
  - (e) None of the above.
  
3. Suppose you know that the data you have is normal in shape and the mean,  $\mu = 30$ , and the standard deviation,  $\sigma = 5$ . Which of the following would you also know?
  - (a) *The minimum would be about 15.*
  - (b) The IQR would be less than 5.
  - (c) The 95th percentile would be more than 40.
  - (d) All of the above are true.
  - (e) Only two of the above are true.
  
4. This is a discrete variable histogram and the numbers at the bottom indicate the values within the bins to the right. Which of the following is true?
 



Bin Range	Relative Frequency
[0, 1)	0.025
[1, 2)	0.200
[2, 3)	0.140
[3, 4)	0.200
[4, 5)	0.250
[5, 6)	0.100
[6, 7)	0.060
[7, 8)	0.025
[8, 9)	0.025

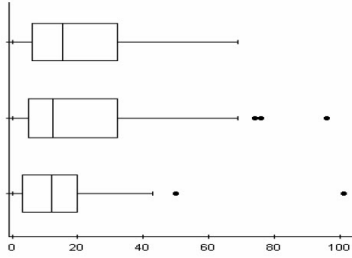
- (a) The median is 4 and since the data is skewed to the right, the mean is more than 4.

- (b) *The median is 3 and since the data is skewed to the right, the mean is more than 3.*
- (c) The median is 4 and since the data is skewed to the left, the mean is more than 4.
- (d) The mean is 4 and since the data is skewed to the right, the median is more than 4.
- (e) The median is 3 and since the data is skewed to the left, the mean is more than 3.
5. The data below came from a SRS taken from all Americans. It was done General Social Survey of Americans in 1991 to determine whether an association existed between gender and perception of level of excitement in life. How likely is a Female to think her life is Exciting?

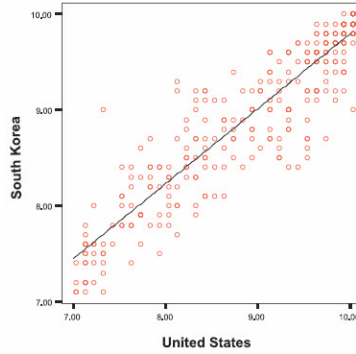
<i>Original</i> <b>Counts</b>	<b>Exciting</b>	<b>Routine</b>	<b>Dull</b>	<b>Total</b>
<b>Male</b>	260	140	12	412
<b>Female</b>	221	305	29	555
<b>Total</b>	481	445	41	967

- (a) 221
- (b) 221 / 967
- (c) 221 / 555
- (d) 221 / 481
- (e) 481 / 555
6. Suppose you know that your data is approximately bell-shaped and the mean is 50 with a standard deviation of 5. Approximately how much of the data falls between (45, 60)?
- (a) 47.5%
- (b) 68%
- (c) 81.5%
- (d) 95%
- (e) 61%
7. Suppose we weigh 50 people. 49 of them are between 180 and 280 lbs. One weighs only 150 lbs. Which of the following is true for this sample?
- (a) *The sample mean,  $\bar{X}$ , is NOT the best measure of center for this data.*
- (b) The sample mean,  $\bar{X}$ , will be larger than the sample median.
- (c) The data cannot be normally distributed since there is an outlier.
- (d) All of the above are true for this data.

- (e) Only two of the above are true for this data (excluding D).
8. The only statistics that we have discussed that are NOT affected by outliers are
- (a) the correlation,  $r$ , and z-scores
  - (b) the mean and the standard deviation
  - (c) the median and the standard deviation
  - (d) Two of the above are not affected by outliers.
  - (e) *None of the above are correct.*
9. Select the statements which are TRUE.
- (a) We can use of a regression line for prediction far outside the range of values of the explanatory variable  $x$  that you used to obtain the line.
  - (b) The further an  $(x, y)$  data point is from the mean, the less it affects the sample statistics.
  - (c) Influential points should be deleted since they change the linear relationship between  $x$  and  $y$ .
  - (d) *The size of a sample,  $n$ , affects the spread of the sampling distribution of the statistic.*
  - (e) None of the above are TRUE.
10. When studying the association between two numeric variables X and Y, which of the following is always true?
- (a)  $r$  being equal to zero implies that X and Y are not related.
  - (b) If all of the observed points fall exactly in a straight line, then  $r = 1$ .
  - (c) *Lurking variables can lead to correlation when there is not causation.*
  - (d) All of the above
  - (e) None of the above
11. What happens to the spread of a statistic's sampling distribution as the sample size increases?
- (a) The spread stays the same since it depends on the population size not the sample size.
  - (b) *The spread gets smaller.*
  - (c) The spread gets bigger.
  - (d) Any of A, B, and C are possible.
  - (e) This question makes no sense since a statistic is a fixed number corresponding to a population, and therefore has no variability.
12. Which of the following is/are true for the data represented above? There are 50 points in each.



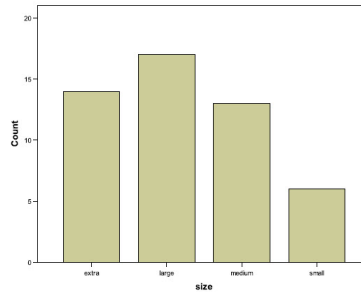
- (a) Although the range of the bottom boxplot is larger than the middle one, the middle one has a larger standard deviation. (Remember the definition of standard deviation).
- (b) The middle boxplot has the largest mean.
- (c) Only the top boxplot has no outliers.
- (d) *All of the above are true.*
- (e) We need summary numbers to determine any of the above.
13. If the correlation between X and Y is 0.98, which of the following is/are true?
- (a) 98% of the variation in Y is explained by X
- (b) *There is a strong, positive linear relationship.*
- (c) There is a cause and effect relationship between the two variables.
- (d) Two of the above.
- (e) All of the above.
14. Which of the following statements is FALSE?
- (a) The distribution of a categorical variable lists the categories and gives the counts or the percent of individuals in each category.
- (b) A bar chart is a useful graphical tool for describing the distribution of a categorical variable.
- (c) *A stemplot is particularly valuable for displaying the shape of the distribution of a categorical variable when there are few observations.*
- (d) A histogram shows the distribution of counts or percents among the values of a single quantitative variable.
- (e) Bar graphs, unlike histograms, have blank spaces between the bars to separate the items being compared.
15. The scatterplot shows the relationship between the U.S. and South Korean judges for some Olympics event. Is the American judges score a good predictor for the South Korean's?
- (a) Yes, since the slope is positive.



- (b) *Yes, since the correlation is strongly positive.*
- (c) Yes, but the correlation is only moderately positive.
- (d) No, the judging is secret, so the South Korean couldnt know how the American would vote.
- (e) No, just because the correlation is strong, we can't say the Americans score could predict the South Korean's.
16. Which of the following is TRUE for bell-shaped data?
- (a) Approximately 95% of the observations in the dataset fall within 2 standard deviations of the mean.
- (b) Approximately 50% of the observations in the dataset fall within the IQR.
- (c) The IQR is less than 2 standard deviations (in length).
- (d) *All of the above are correct statements for bell-shaped data.*
- (e) Only two of A, B and C are correct statements for bell-shaped data.
17. Suppose the mean score for the first exam is 70 with a standard deviation of 10. Assuming the data is bell-shaped, which of the following is true for an exam score of 90?
- (a) The 90 means that this person did better than 90% of the students who took the same exam.
- (b) *The 90 did better than almost 97.5% of the other students since it has a z-score of 2.*
- (c) The 90 did better than almost 85% of the other students since it has a z-score of 1.
- (d) Since the z-score of the 90 is 2, this is the second highest score on the exam.
- (e) All we can say is that the 90 made an A on the test.
18. Which of the following is true?
- (a) You can calculate both the mean and median from a stemplot.
- (b) You can calculate both the mean and median from a histogram.

- (c) You can estimate both the mean and median from a boxplot.
- (d) All of the above are true.
- (e) *Only two of the above are true.*

19. Which of the following best describes the distribution above?



- (a) The data is skewed to the right.
  - (b) The data is skewed to the left.
  - (c) The data is almost symmetric.
  - (d) The shape cant be determined.
  - (e) *Shape is not a valid description for this data.*
20. Suppose data was collected from statistics students regarding the length of time they studied for an exam,  $x$ , and their grade on the exam,  $y$ . The data contained scores for students that studied anywhere from 0 to 8 hours. We find that the prediction (regression) equation is  $\hat{y} = 22 + 10x$  and the correlation coefficient,  $r = 0.74$ . What would be the expected increase in their exam grade if a student studied one additional hour?
- (a) 22 points
  - (b) 10 points
  - (c) 74 points
  - (d) 0.74 points
  - (e) 32 points