

STAT303 Sec 508-510  
Spring 2006  
Exam #1

Instructor: Julie Hagen Carroll

Name: \_\_\_\_\_

1. **Don't even open this until you are told to do so.**
2. Please PRINT your name in the blanks provided.
3. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers **clearly**. Multiple marks will be counted wrong.
4. You will have 60 minutes to finish this exam.
5. If you have questions, please write out what you are thinking on the back of the page so that we can discuss it after I return it to you.
6. If you are caught cheating or helping someone to cheat on this exam, you both will receive a grade of **zero** on the exam. You must work alone.
7. This exam is worth the same as a regular exam (this may differ from section to section).
8. Good luck!

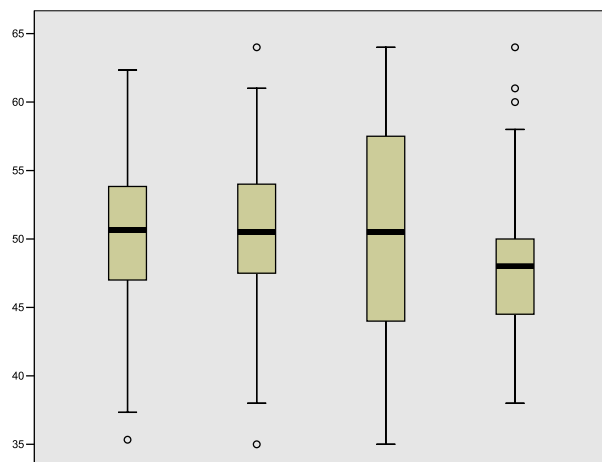
- Why do we use  $s$ , the standard deviation, rather than  $s^2$ , the variance, when describing a sample and doing calculations?
  - It's easier to calculate the standard deviation.
  - It's in the same units as the mean,  $\mu$ .
  - You can't determine the variance without knowing the standard deviation.
  - It doesn't matter which one you use since they are the same quantity.
  - Exactly two of the above are correct.
- It is known that the mean price for a carton of milk is \$1.00 with a standard deviation,  $s$ , of \$0.20. Also, the mean price per pound for beef is \$2.50 with a standard deviation of \$0.60. If a local store, X, sells milk for \$0.85 per carton and beef for \$2.00 per pound, which of the following is true for the store?
  - Beef is relatively cheaper since it's \$0.50 below the average vs. \$0.15 below the average for milk.
  - Milk is relatively cheaper since it's only \$0.85 where beef is \$2.00 per pound.
  - You can't compare the two since the average prices are different.
  - Beef is relatively cheaper since it's 0.83s below the average vs. 0.75s below the average for milk.
  - Milk is relatively cheaper since it's 0.75s below the average vs. 0.83s below the average for beef.

	Freq	Pct	Cum Pct
65.00	1	1.0	1.0
72.00	2	2.0	3.0
73.00	3	3.0	6.0
74.00	6	6.0	12.0
75.00	3	3.0	15.0
76.00	5	5.0	20.0
77.00	7	7.0	27.0
78.00	15	15.0	42.0
79.00	11	11.0	53.0
80.00	11	11.0	64.0
81.00	6	6.0	70.0
82.00	11	11.0	81.0
83.00	3	3.0	84.0
84.00	7	7.0	91.0
85.00	2	2.0	93.0
86.00	3	3.0	96.0
87.00	2	2.0	98.0
89.00	2	2.0	100.0
Total	100	100.0	

- Which is the correct list of the 5 Number Summary for this table?
  - 1, 25, 50, 75, 100
  - 65, 75, 79, 84, 89
  - 65, 77, 79, 81, 89
  - 65, 77, 79, 82, 89
  - 65, 71, 77, 83, 89

- Which of the following indicate that the data is skewed left?
  - The mode (tallest bin) of the histogram is on the right, and the other bins get continually shorter as you go left.
  - The boxplot has the median,  $\tilde{x}$ , closer to  $Q_3$  and the maximum than to  $Q_1$  and the minimum.
  - The mean,  $\bar{x}$ , is greater than the median,  $\tilde{x}$ .
  - All of the above indicate left skewness.
  - Exactly two of the above (excluding D.)

- or 4. You can tell if a dataset is skewed by looking at
- the mean vs. the median
  - the standard deviation vs. the interquartile range
  - the boxplot of the dataset
  - All of the above would be useful.
  - Two of the above would be useful.



- Which of the following is/are true about the boxplots(5)?
  - All of the boxplots are normal.
  - Only number 2 is normal since it has outliers on both ends (therefore symmetric).
  - Only number 3 is normal.
  - Numbers 1 and 2 are normal.
  - You can't determine whether they are normal or not with boxplots. We need normal quantile plots instead.

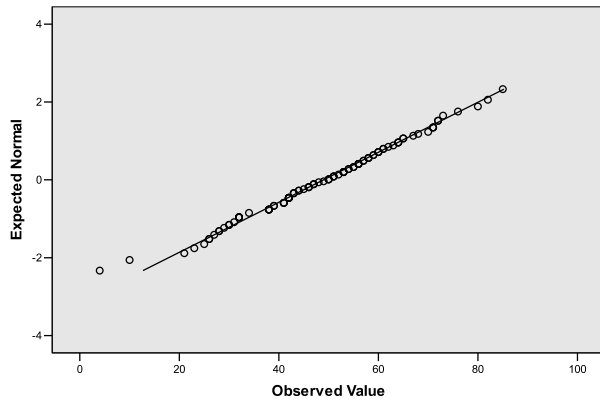
MomHair	Hair				Total
	Black	Blond	Brown	Red	
Black	1	0	12	0	13
Blonde	0	7	10	0	17
Brown	1	13	40	1	55
Gray	0	0	3	0	3
Red	0	0	3	0	3
White		1	0	0	1
Total	2	21	68	1	92

- The Two-way table above is showing the relationship between a child's hair color and their mom's. Which is the following is/are true?
  - All of the boxplots are normal.
  - Only number 2 is normal since it has outliers on both ends (therefore symmetric).
  - Only number 3 is normal.
  - Numbers 1 and 2 are normal.
  - You can't determine whether they are normal or not with boxplots. We need normal quantile plots instead.

- A. If your mom has gone gray, you must have brown hair.
  - B. In this survey, the child most likely has brown hair no matter what color their mom has.
  - C. If the child has brown hair, the likelihood of the mom having brown hair is 40/55.
  - D. All of the above are true.
  - E. Only two of the above are true.
7. Again using the table of hair color, which of the following best describes the proportion 7/21? Out of the 92 people in the survey, it is
- A. how likely both mom and child have blond hair.
  - B. how likely a blond child will have a blonde mom.
  - C. how likely a blonde mom will have a blond child.
  - D. how likely the mom is blonde.
  - E. how likely the child is blond.
- or 7. Again using the table of hair color, which of the following best describes the proportion 7/92? Out of the 92 people in the survey, it is
- A. how likely both mom and child have blond hair.
  - B. how likely a blond child will have a blonde mom.
  - C. how likely a blonde mom will have a blond child.
  - D. how likely the mom is blonde.
  - E. how likely the child is blond.
8. Which of the following statements is TRUE?
- A. Increasing the sample size,  $n$ , increases the strength of the linear relationship between  $x$  and  $y$ , *i.e.*, the correlation coefficient,  $r$ .
  - B. Increasing the sample size,  $n$ , improves our estimates (statistics) since their variability is reduced, so our estimated line is closer to the true, population, line.
  - C. Increasing the sample size,  $n$ , makes the points fall closer to the regression line, so  $R^2$  is increased.
  - D. All of the above are false.
  - E. All of the above are true.
9. Which of the following is/are true?
- A. A correlation coefficient = 0 means the least squares line is horizontal (flat).
  - B. A correlation coefficient = 0 means that the points don't deviate from the least squares line.
  - C. A slope,  $b_1 = 1$  means the correlation coefficient,  $r = 1$ .
  - D. All of the above are true.
  - E. None of the above are true.
- or 9. Consider a set a bivariate data,  $(x_i, y_i)$ , of size  $n$ . If we *add* 5 to each of the  $x$ 's, then which of the following will change?
- A. the correlation coefficient for  $x$  and  $y$
  - B. the slope of the linear regression line  $\hat{y} = b_0 + b_1x$
  - C. the intercept of the linear regression line  $\hat{y} = b_0 + b_1x$
  - D. All of the above will change.
  - E. None of the above will change.
10. If a dataset has outliers
- A. the mean is a better measure of center than the median.
  - B. the IQR is a better measure of spread than the standard deviation.
  - C. the correlation coefficient is closer to zero than it would be without the outliers.
  - D. All of the above are correct statements.
  - E. Only two of the above are correct statements.
- or 10. Suppose you have a dataset of catagorical data. Which of the following would be appropriate statistics to describe your dataset?
- A. the mean and standard deviation
  - B. the median and IQR (since you don't know if the data is skewed)
  - C. the correlation coefficient
  - D. the proportion
  - E. Exactly two of the above would be appropriate statistics to report.
11. Suppose your data is approximately normally distributed with a mean,  $\bar{x} = 25$  and standard deviation,  $s_x = 15$ . What is the approximate 95th percentile of your data? (Note: the 95th percentile has 95% of the data is below it.)
- A. 35
  - B. 44
  - C. 45
  - D. 54
  - E. 55
- or 11. Assuming that the distribution of grades on this exam is approximately normal, what is the minimum  $z$ -score would you have to have to be in the top 10% of the class?
- A. 0.90
  - B. between 0 and 1
  - C. between 1 and 2
  - D. between  $-1$  and  $-2$
  - E. It's impossible to tell without knowing the mean and standard deviation.

12. Which of the following is TRUE?

- A. Randomization is ALWAYS necessary if we want 'good' statistics.
- B. Randomization prevents any bias in our statistics.
- C. Confounding variables affect the response variable, but their effects cannot be separated from other explanatory variables effects.
- D. All of the above are true.
- E. Only two of the above are true.



13. If the mean is 50 and the standard deviation is 15, then we could say

- A. About 95% of the observations will fall between 20 and 80.
- B. The approximate minimum and maximum are 5 and 95.
- C. Both A and B are true.
- D. Nothing since we don't know if the data is normal or not.
- E. Nothing since the data isn't normal.

14. For a certain dataset, you are told that the standard deviation,  $s = 0$ . What else can you say about the dataset?

- A. the mean,  $\bar{x} = 0$
- B. the median,  $\tilde{x} = 0$
- C. the IQR = 0
- D. all of the above
- E. none of the above

or 14. You're trying to figure out the relationship between Major and GPR at A&M. Which of the following would be helpful?

- A. the correlation coefficient between Major and GPR
- B. a comparison of boxplots of GPR by Major
- C. comparing the sample means of GPR for each Major
- D. all of the above
- E. only two of the above

15. Suppose we stand on the corner of Texas Avenue and University Drive and count the number of cars that pass through the intersection during a green light on Texas. What type of variable would this be?

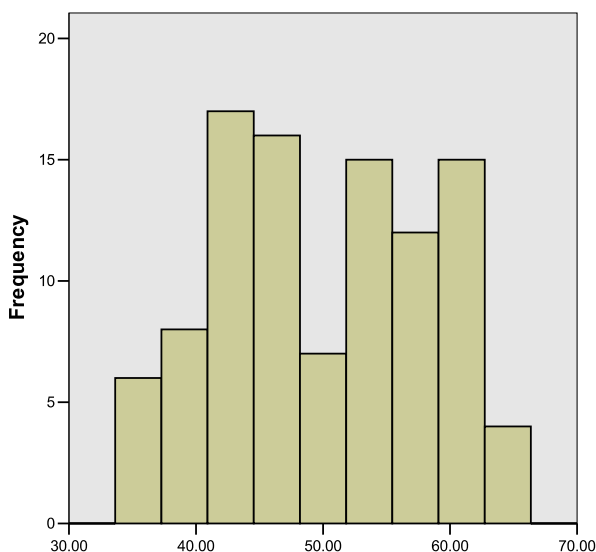
- A. a numerical continuous since the length of a green light could vary
- B. a numerical discrete since the number of cars would vary
- C. a categorical since the cars are different types/makes
- D. You can't tell without the data.
- E. Inaccurate since so many people run reds lights!

16. Given a least squares line of  $\hat{y} = 14 - 0.3x$ , what is the residual for the point (2, 9)?

- A. -4.4
- B. -2.3
- C. 11.3
- D. 13.4
- E. 2.3

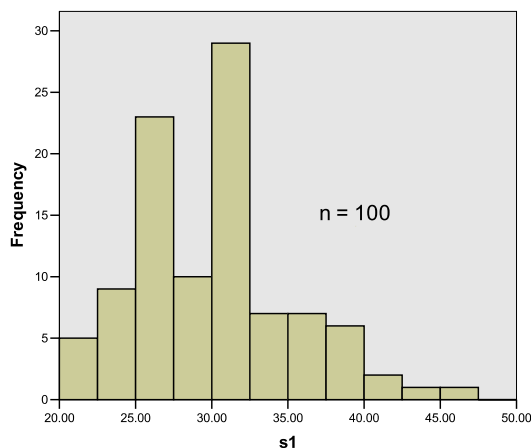
17. Suppose we want to study the quality of education in the Texas public schools. We want to include every school district, but it's too costly to visit every school within each district, so we only take a random sample of 3 schools in every district and gather data from every teacher/class within that school. What type of sample did we get?

- A. a multi-stage sample since we had two levels of sampling
- B. a cluster sample since we took everyone within a school
- C. a stratified sample since we gather data from every school district
- D. a simple random sample since we didn't know ahead of time which classes would be in the sample
- E. a biased sample since we didn't look at all of the schools in Texas



18. What can you say about this histogram? There are 100 observations.

- A. The median,  $\tilde{x}$ , and mean,  $\bar{x}$ , are about 50.
- B. The IQR is about 15 (less than 20).
- C. It is mostly likely a normal distribution.
- D. All of the above are true.
- E. Only two of the above are true.



19. What percent of the observations in the histogram above are 25 or more?

- A. 75%
- B. almost 85%
- C. about 23%
- D. not quite 15%
- E. Percents can't be determined here since it's a frequency histogram.

20. Suppose you get a  $z$ -score = 1.2 on this exam. Nearly everyone missed one particular question, so I decide to give everyone credit, *i.e.*, I give 5 points to everyone who missed it. What SHOULD you think about this if you got the problem right originally?

- A. It doesn't matter since your  $z$ -score would stay the same.
- B. You get gyped since your  $z$ -score would decrease.
- C. You benefit since your  $z$ -score would increase.
- D. Your  $z$ -score would increase, but so would everyone else's.
- E. Your  $z$ -score would decrease, but so would everyone else's.

or 20. Suppose I give you your  $z$ -score on this exam instead of your points out of 100. If you get a  $z$ -score of 0.25, this means

- A. you only got 25% of the questions correct.
- B. you only missed 25% of the questions.
- C. you were 25% above the median.
- D. you did worse than 25% of the class.
- E. you did better than the average.

1B,2D,3D,4E,5D,6B,7B(or7A),8B,9A(or C),  
10E(or D),11D(or C),12E,13C,14C(or E),15B,  
16A,17A,18E,19B,20B(or20E)