

Name _____ Period _____

Statistics Honors
Midterm Exam Review Packet 2019

<u>Exam:</u>	Thursday, January 24 th	Periods 1 & 12
	Friday, January 25 th	Periods 2 & 9/10/11
	Monday, January 28 th	Periods 3 & 7/8/9
	Tuesday, January 29 th	Periods 4 & 5/6/7

Please bring the following the day of your exam:

- 1.) Graphing calculator – I DO NOT HAVE ANY TO LEND OUT!!!
Make sure your batteries work!
- 2.) #2 pencils for the scantron section of your exam
- 3.) This completed packet – 25 points
(1 pt per page – 20 points - checked daily, and 5 additional pts when you turn it in on the day of your exam)

Exam: 30 multiple choice @ 2 pts each (60pts) & 4 open ended (40 pts)

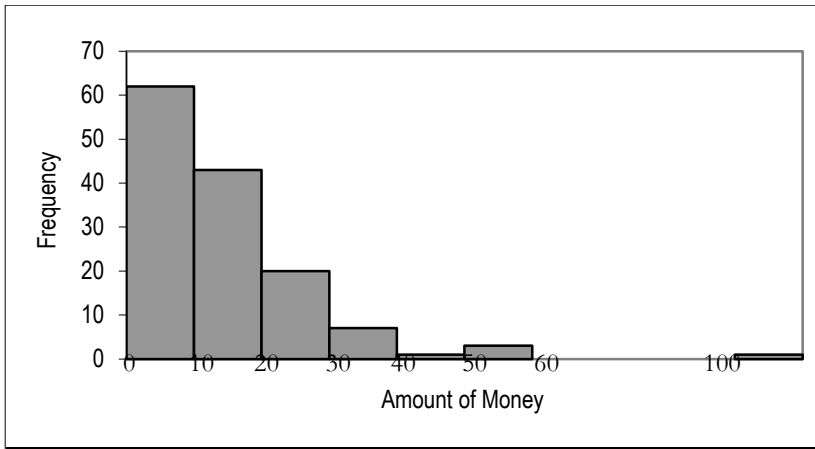
TO COMPLETE THIS REVIEW YOU NEED YOUR NOTES AND PAST ASSESSMENTS. YOU ALSO NEED YOUR FORMULA SHEET AND Z-SCORE TABLE.

The following are the review assignments that are DUE on the listed days below. Each assignment will be checked (1 point per page). An answer key will be provided for you to check and make corrections, but NOT TO KEEP! If you are absent, all work must still be completed.

Classwork will be assigned the day before & is DUE at the beginning of the period on:

Friday, January 18 th	Chapter 1 (p. 2 – 4), OE #1a-d (p. 16) #8a-j (p. 18-20)
Tuesday, January 22 nd	Chapters 2 and 3 (p. 5 – 9), OE #2 - 7 (p. 17)
Wednesday, January 23 rd	Mixed Review (p. 10 – 15), OE #9a-k (p. 20-21)

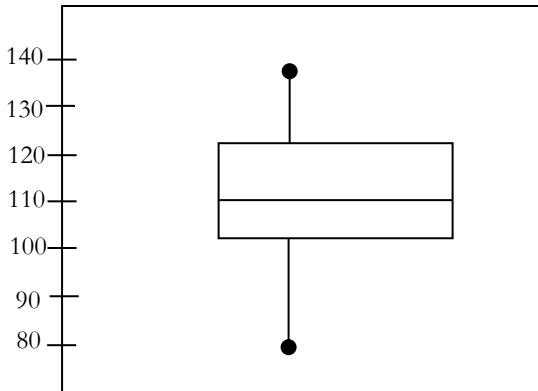
- 1.) In a statistics class with 136 students, the professor records how much money each student has in his or her possession during the first class of the semester. The histogram below is of the data collected.



The histogram

- | | |
|----------------------|--------------------------------|
| A.) is skewed right. | B.) appears to have an outlier |
| C.) is asymmetric | D.) all of the these are true |

- 2.) The boxplot below is of the birthweights (in ounces) of a sample of 160 infants born in a local hospital.



The median birthweight is approximately

- | |
|----------------|
| A.) 80 ounces |
| B.) 100 ounces |
| C.) 110 ounces |
| D.) 120 ounces |
| E.) 140 ounces |
- 3.) What happens to the standard deviation when you add a constant to every data point in the set?
- | |
|---|
| A.) standard deviation stays the same |
| B.) standard deviation changes by the same constant |
| C.) standard deviation shifts to the right |
| D.) standard deviation shifts towards the outliers |
| E.) cannot tell from the given information |

- 4.) What happens to the mean when you add a constant to every data point in the set?
- A.) mean stays the same
 - B.) mean changes by the same constant
 - C.) mean shifts to the right
 - D.) mean shifts towards the outliers
 - E.) cannot tell from the given information
- 5.) Which of the following are examples of categorical variables?
- I. method of payment for purchases (cash, check, credit card)
 - II. zip code of customer's mailing address
 - III. number of items purchased
 - IV. amount of tax on a purchase
 - V. type of items purchased
- A.) I
 - B.) I & II
 - C.) I, II, IV & V
 - D.) I, II & V
 - E.) all of them
- 6.) If a data set has one outlier, what is the best measure of center?
- A.) mean
 - B.) median
 - C.) standard deviation
 - D.) interquartile range
 - E.) mode
- 7.) If a data set has one outlier, what is the best measure of spread?
- A.) mean
 - B.) median
 - C.) standard deviation
 - D.) IQR & range
 - E.) mode
- 8.) Which graph is best to use if a data set has one outlier?
- A.) histogram
 - B.) stem-and-leaf plot
 - C.) dotplot
 - D.) ogive
 - E.) box-and-whisker plot
- 9.) Which of the following is a resistant measure?
- A.) mean
 - B.) range
 - C.) standard deviation
 - D.) variance
 - E.) median

10.) The grades on the last test in Statistics class are given below:

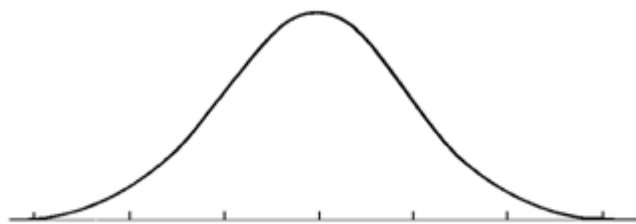
94	91	90	88
87	85	84	84
82	81	81	79
77	76	76	74
73	71	70	70
68	68	67	65
60	59	57	35

- a.) Find the five – number summary for this set of data.
- b.) Use the graphing calculator to *check* your results.
- c.) Determine if there are any outliers in this data. Justify your answer.
- d.) Draw a boxplot (modified) for this set of data.
- e.) Find the mean, mode and standard deviation for this set of data.

11.) State an example of a categorical variable. _____

12.) State an example of a quantitative variable. _____

- 1.) **Directions:** Fill in each blank to describe the area under any density curve including the area under a normal distribution:
- a.) In a normal distribution the _____, _____, and _____ are all equal.
 - b.) The normal curve is _____ shaped and _____ about the mean.
 - c.) The total area under the normal curve is equal to _____.
 - d.) The normal curve approaches, but _____ the x-axis as it extends further and further from the mean.
- 2.) **Directions:** Fill in each blank to describe Standard Normal Calculations:
- a.) A standardized value is called a ____ - _____.
 - b.) A z-score tells us how many _____ the original observation falls away from the mean and in which direction.
 - c.) The **standard** normal distribution is the normal distribution $N(\text{____}, \text{____})$ with mean _____ and standard deviation _____.
 - d.) The Empirical Rule states that approximately _____ of all observations fall within 1 standard deviation from the mean, _____ of all observations fall within 2 standard deviations from the mean and _____ of all observations fall within 3 standard deviations from the mean.
 - e.) Label the diagram of a Standard Normal Distribution below: (Be sure to include appropriate labels for the x-axis and information from the Empirical Rule.)



- 3.) Scores on the American College Testing (ACT) college entrance exam are approximately normal with $N(18,6)$. Lisa's standardized score on the ACT was $z = 0.7$. What was her actual ACT score?

- 4.) IQ scores have a mean of 100 and a standard deviation of 15. Suppose that Diane's IQ score is 115 and Fred's IQ score is 95. Duran's IQ z -score is 1.13 and Azuree's IQ z -score is -0.95 . List the students in order of ascending IQ scores.
- (a) Azuree, Duran, Fred, Diane (b) Azuree, Fred, Duran, Diane
(c) Fred, Azuree, Diane, Duran (d) Fred, Azuree, Duran, Diane
(e) Azuree, Fred, Diane, Duran
- 5.) A college admissions office needs to compare scores of students who take the SAT (scholastic Aptitude Test) with those who take the ACT (American College Test). The mean score on the SAT is 896 with a standard deviation of 174. The mean score on the ACT is 20.6 with a standard deviation of 5.2. Suppose student A scores 1080 on the SAT and student B score 28 on the ACT. Which of the following statements is true?
- (a) Student A did better than Student B because his score is higher.
(b) Student A did better than Student B because the SAT is a more difficult test than the ACT.
(c) Student A scored 1.5 standard deviations above the norm.
(d) Student B has a z -score of 1.42.
(e) no decision can be made because the tests are different.

Directions for # 6 - 11: **Draw and label a diagram where appropriate** and show all work including calculator steps used to answer each question.

- 6.) The lifetime of a 2-volt non-rechargeable battery in constant use has a normal distribution with a mean of 516 hours and a standard deviation of 20 hours. The proportion of batteries with lifetimes exceeding 520 hours is approximately:
- (a) 0.2000 (b) 0.5793 (c) 0.4207
- 7.) The lifetime of a 2-volt non-rechargeable battery in constant use has a normal distribution with a mean of 516 hours and a standard deviation of 20 hours. 90% of all batteries have a lifetime shorter than:
- (a) 517.28 hours (b) 536.00 hours (c) 541.60 hours
- 8.) The scores on a Statistics exam are approximately normally distributed with a mean of 75 and a standard deviation of 5. Which of the following would you expect, if any, to be an outlier.
- (a) 65 (b) 75 (c) 85 (d) 100

- 9.) The heights of American men aged 18 to 24 are approximately normally distributed with mean 68 inches and standard deviation 2.5 inches.
- a.) What percent of the young men are taller than 6 feet?
- b.) Only about 5% of young men have heights outside what range of height values.
- 10.) If the weights of males are distributed normally with mean of 170 and standard deviation of 20, what weight (**to the nearest pound**) would correspond with the 40th percentile?
- 11.) The heights of American women ages 18 to 24 are approximately normally distributed with $N(64.5, 2.5)$
What percent of young women are taller than 66 inches?

Directions: Circle the best answer for each multiple choice question.

- 1.) A group of college professors and a group of college students were asked to rate a collection of essays. The correlation coefficient was 1.03. A reasonable conclusion of this report is that
 - a.) professors and students tend to agree on what is good in an essay
 - b.) there is little relationship between professor and student ratings of the essays
 - c.) there is a strong relationship between professor and student ratings of the essays
 - d.) there is a strong causation relationship between professor and student ratings of the essays
 - e.) an arithmetic error was made.

- 2.) What does the square of the correlation (r^2) measure:
 - a.) The slope of the least – squares regression line.
 - b.) The intercept of the least – squares regression line.
 - c.) The extent to which cause and effect is present in the data.
 - d.) The percent of the variation in the values of y that is explained by the least – squares regression of y on x .
 - e.) None of the above. The correct answer is _____.

- 3.) What is the least squares regression line?
 - a.) a line that minimizes the sum of the squares of the vertical distances of the observed response variable
 - b.) a line that minimizes the correlation coefficient on the response variable
 - c.) a line that minimizes the coefficient of determination on the response variable
 - d.) a line that minimizes the sum of the squares of the vertical distances of the observed explanatory variable
 - e.) a line that minimizes the sum of the squares of the horizontal distances of the observed explanatory variable

- 4.) A researcher quantified the personality traits "greed" and "superciliousness" and found the correlation between them to be -0.40 . What percentage of the variation in greed can be explained by the relationship with superciliousness:
 - a.) 0%
 - b.) 16%
 - c.) 20%
 - d.) 40%
 - e.) 60%

- 5.) In regression, the residuals are which of the following:
 - a.) Those factors unexplained by the data.
 - b.) The difference between the observed responses(y) and the values predicted by the regression line(\hat{y}).
 - c.) Those data points which were recorded after the formal investigation was completed.
 - d.) Possible models unexplored by the investigator.
 - e.) None of the above. The correct answer is _____.

Directions: Fill in the blanks or answer the following the questions.

- 6.) The correlation coefficient (r) is always greater than or equal to _____ and less than or equal to _____.
- 7.) For each of the following pairs of variables, would you expect a substantial **negative** correlation, a substantial **positive** correlation, or **no** correlation?
- a.) The advertising expenses of a company and that company's sales _____
- b.) The age of a child and the number of words in their vocabulary _____
- c.) The heights and the IQ scores of adult women _____
- 8.) The weights of children in the Egyptian village of Nahya were recorded. Here are the mean weights of the 170 children in that village.

Age(months) (x)	1	2	3	4	5	6	7	8	9	10	11
Weight (kg) (y)	4.3	4.1	5.7	6.3	6.8	7.1	7.2	7.2	7.2	7.2	7.5

- a.) What is the correlation coefficient?
- b.) What percent of the variation in weight is explained by the linear regression of age?
- c.) What is the residual of an age of 5?
Use your formula sheet & check answer by making a table of residuals on calc

- 1.) Suppose we fit the least squares regression line to a set of data. What do you call any individual point with a unusually large residual?
- (a) response variables (b) the slope (c) outliers
(d) correlations (e) none of the above
- 2.) If removing an observation from a data set would have a marked change on the position of the LSRL fit to the data, what is the point called?
- (a) robust (b) a residual (c) a response
(d) influential (e) none of the above
- 3.) Suppose the correlation between two variables x and y is due to the fact that both are responding to changes in some unobserved third variable. What is this due to?
- (a) cause and effect between x and y (b) the effect of the lurking variable
(c) extrapolation (d) common sense
(e) none of the above
- 4.) Suppose a straight line is fit to data having response variable y and explanatory variable x . Predicting values of y for values x outside the range of observed data is called:
- (a) correlation (b) causation (c) extrapolation
(d) sampling (e) none of the above
- 5.) State whether the variable is categorical or quantitative
- (a) Person's age _____ (b) Gender _____
(c) Salary _____ (d) Monthly water bill _____
(e) State _____ (f) Zip Code _____

- 6.) A 1969 study among Pima Indians of Arizona investigated the relationship between a mother's diabetic status and the appearance of birth defects in children. The results appear in the two – way table below.

<i>Birth Defects</i>	<i>Diabetic Status</i>			Total
	<i>Non – diabetic</i>	<i>Pre – Diabetic</i>	<i>Diabetic</i>	
None	754	362	38	1154
One or more	31	13	9	53
Total	785	375	47	1207

- a.) Compute all of the **marginal distributions**. (Write as percents rounded to the nearest **tenth**.)

Probability No Birth defect =

Probability Non Diabetic =

Probability 1+BD =

Probability Pre Diabetic =

Probability Diabetic =

- b.) Compute the **conditional distributions** of **diabetic status among babies with no birth defects**.
HINT - Only look at the row with "No Birth defects".

Diabetic Status: *Percent:*

Non Diabetic

Pre Diabetic

Diabetic

- c.) Compute the **conditional distributions** of **birth defects among Diabetics**. *HINT - Only look at the column with "Diabetics".*

Birth Defects: *Percent:*

None

One or More

- 7.) A researcher suspected a relationship between people's preferences in beverage and preference in snacks. A random sample of 96 people produced the following two – way table:

		Beverage Preference			
		<i>Milk</i>	<i>Water</i>	<i>Soda</i>	TOTAL
Favorite Snack	<i>Cookies</i>	22	5	7	34
	<i>Potato Chips</i>	2	8	20	30
	<i>Fruit</i>	11	15	6	32
TOTAL		35	28	33	96

- (a) Find each of the probabilities, in percents:
- i.)* Probability preferred Milk
 - ii.)* Probability preferred Fruit
 - iii.)* Probability preferred Soda
 - iv.)* Probability of preferring Milk among those that preferred Cookies
 - v.)* Probability of preferring Potato Chips among those that preferred Soda
- 8.) Research shows there is a strong positive correlation in developing nations between number of television sets per capita and the life span of citizens. It is unlikely that buying more television sets will cause a person to live longer. A more reasonable explanation could be that wealthier people have more money to buy television sets and have more money to take care of themselves, and therefore live longer.
- (a) Identify the variables x , y & z .
- (b) Draw a diagram of the relationship and identify all variables.
- (c) This problem represents *causation*, *common response*, or *confounding*?

- 9.) A researcher observes that, on average, the number of divorces in cities with major league baseball teams is larger than in cities without major league baseball teams. Another variable could be the size (population) of the city. Cities with larger populations tend to have major baseball teams and overall more divorces because there are more people.
- (a) Identify the variables x , y & z .
- (b) Draw a diagram of the relationship and identify all variables.
- (c) This problem represents *causation*, *common response*, or *confounding*?
- 10.) A soccer coach wanted to improve the team's playing ability, so he had them run two miles a day. At the same time the players decided to take vitamins. In two weeks the team was playing noticeably better, but the coach and players did not know whether it was from the running or the vitamins.
- (a) Identify the variables x , y & z .
- (b) Draw a diagram of the relationship and identify all variables.
- (c) This problem represents *causation*, *common response*, or *confounding*?

Determine the population and sample for each of the following situations.

- 11.) In a recent survey, 3002 adults in the United States were asked if they read news on the Internet at least once a week. Six hundred of the adults said yes.

The population is: _____

The sample is: _____

- 12.) A study of 254 patients with sleep disorders was conducted to find a link between obesity and sleep disorders.

The population is: _____

The sample is: _____

- 13.) The U.S. Department of Energy conducts weekly surveys of approximately 900 gasoline stations to determine the average price per gallon of regular gasoline. On December 29, 2003, the average price was \$1.478 per gallon.

The population is: _____

The sample is: _____

- 14.) A study of 33,043 infants in Italy was conducted to find a link between a heart rhythm abnormality and sudden infant death syndrome.

The population is: _____

The sample is: _____

- 15.) A survey of 2104 households in the United States found that 65% subscribe to cable television.

The population is: _____

The sample is: _____

- 16.) A survey of 546 women found that more than 56% are the primary investor in their household.

The population is: _____

The sample is: _____

- 17.) A survey of 791 vacationers from the United States found that they planned on spending at least \$2000 for their next vacation.

The population is: _____

The sample is: _____

- 18.) A member of Congress wants to know what his constituents think of proposed legislation on health insurance. His staff reports that 228 letters have been received on the subject, of which 193 oppose the legislation.

What is the population in this situation?

- (a) The constituents
- (b) The 228 letters received
- (c) The 193 opposing the legislation
- (d) Congress
- (e) None of the above. The correct answer is _____

1.) The grades on the last test in Statistics class are given below:

95	91	90	88
87	85	84	84
82	81	81	81
77	76	76	74
74	74	70	70
68	68	67	65
60	59	57	25

- a.) Find the five – number summary for this set of data.
- b.) Determine if there are any outliers in this data. Justify your answer.
- c.) Draw a boxplot (modified) for this set of data.
- d.) Find the mean, mode and standard deviation for this set of data.

- 2.) The heights of American men aged 18 to 24 are approximately normally distributed with mean 68 inches and standard deviation 2.5 inches.
- a.) What percent of the young men are taller than 5.5 feet?
- b.) Only about .3% of young men have heights outside what range of height values?
- 3.) If the weights of males are distributed normally with mean of 175 and standard deviation of 25, what weight (**to the nearest pound**) would correspond with the 30th percentile?
- 4.) The heights of American women ages 18 to 24 are approximately normally distributed with $N(64.5, 2.5)$. What percent of young women are taller than 64 inches?

The following table contains the IQ and SAT scores for a group of high school seniors. Use the table to answer #'s 5-7.

IQ	100	110	120	125	130	135	140
SAT	875	900	910	925	1010	1135	1210

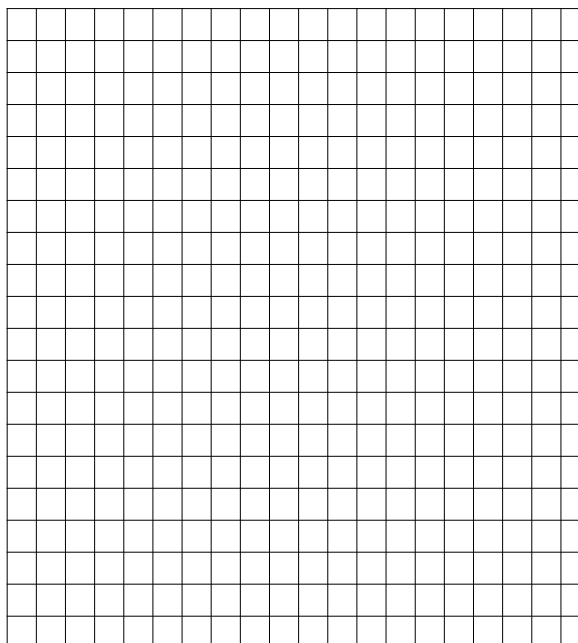
- 5.) Using the table above, what is the correlation coefficient for this data?
- 6.) Using the table above, what percent of the variation or error in the SAT score is explained by the linear regression on the IQ?
- 7.) Using the table above, what is the residual for an IQ of 110?

Statistics Exam Scores				
72	75	70	74	82
47	64	58	64	78
98	39	77	77	72
55	72	94	50	63
83	68	72	97	66

8.) Use the data above to answer the following questions.

(a) Construct a frequency distribution for the data. (Table with “CLASS” and “COUNT”)

(b) Create a histogram for this data.



(c) Describe the distribution (SOCS).

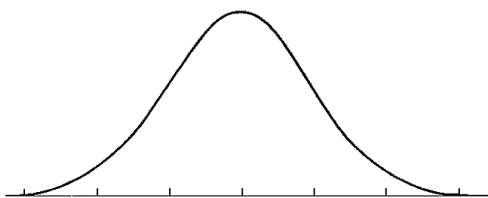
(d) Determine if there are any outliers. Use the formulas & graphing calculator to check work.

(e) Find the five-number summary for the data.

(f) Construct a modified boxplot for the data.

(g) Find the mean and standard deviation for this data.

(h) Assume the data is normally distributed. Use the Empirical Rule to find the range within which the scores for 68% of the grades fall.



...
distribution?

n, IQR and range to describe the center and spread of a

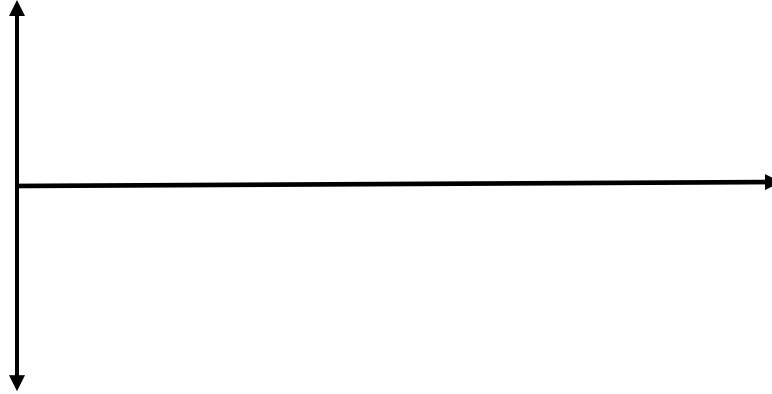
- (j) In general, when is it best to use the mean and standard deviation to describe the center and spread of a distribution?

- 9.) **TV and Test Scores** An instructor wants to show students that there is a linear relationship between the number of hours they watch television during a certain weekend and their scores on a test taken the following Monday. The table below shows the # of hours spent watching and that student's score on the test.

Hours spent watching TV	0	1	2	3	5	6
Test Score	96	85	82	74	68	58

- (a) View a scatterplot of the data on your calculator.
- (b) Find the correlation coefficient.
- (c) Determine the equation of the LSRL **using formulas**.
- (d) Determine the equation of the LSRL **from calculator**.
- (e) Plot the LSRL on your scatterplot on the calculator.
- (f) Use your calculator to find the table of residuals.

(g) Draw the residual plot.



(h) Is it appropriate to call the data linear? EXPLAIN.

(i) Predict the score a student would likely receive if he watched 5 hours of TV.

(j) Compute the residual for watching 5 hours of TV.

(k) What percent of the variability in test grade can be explained by the hours spent watching TV.