PRACTICE FINAL EXAM (COMPREHENSIVE)

PRACTICE FINAL EXAM MUST INCLUDE QUESTIONS FROM PRACTICE EXAM 1 (1/3) + PRACTICE EXAM 2 (1/3) + FOLLOWING QUESTIONS (1/3).

- 159. Which graph is used to display the relationship between two quantitative variables?
 - A. Scatterplot
 - B. Bar Graph
 - C. Stemplot
 - D. Histogram
- 160. Which of the following pairs of variables CANNOT be studied using correlation?
 - A. Adult height and yearly income
 - B. Smokestack height and air pollution particulate count
 - C. Worker's age and number of sick days taken per year
 - D. Family size and religion
- 161. Which of the following gives the strongest linear association?
 - A. r = 0.
 - B. r = 0:4.
 - C. r = 0:9.
 - D. r = 0.6.
- 162. A coach in a large high school thinks that ballet training will improve the batting performance of his baseball team. He decides to have a randomly selected half of the team take six weeks of ballet training before the baseball season begins while the other half does not take such training. He will then compare the season batting averages of group A (those with ballet training) and group B (those without ballet training) by comparing the mean of group A with the mean of group B. This is an example of a

A. stratified random sample.

- B. simple random sample.
- C. completely randomized design.
- D. randomized block design.
- 163. The regression line relating average August temperatures in degrees Fahrenheit (Y) and geographic latitude (Y) for a sample of 20 North American cities is $\hat{y} = 114 x$ where $r^2 = 0.64$. What is the value of the correlation (r)?
 - A. r = -0.8
 - B. r = 0.8
 - C. r = 0.64
 - D. r = -0.64
- 164. A researcher examines data from cities in the United States. The scatterplot of the number of churches (X) in the city versus the average number of annual homicides (Y) in the city shows a strong positive linear relationship. Can we conclude that more churches cause more homicides to occur? If so, explain. If not, give a potential lurking variable?
- 165. Suppose that x = height in inches and y = weight in pounds of men aged 18 to 29 years old. The correlation is r = 0.71 and the equation of the least-squares line is $\hat{Y} = -250 + 6x$:
 - (a) Interpret the slope value in the regression equation in terms of the problem.
 - (b) Predict the weight (in pounds) of a 23-year-old male who is 72 inches tall.
 - (c) A 23-year-old male who is 72 inches tall is found to weigh 178 pounds. What is the residual?
 - (d) Give the value of r^2 . Give 4 decimal places.

(e) Interpret the value of r^2 [the answer to (d)] in terms of the problem.

Use the following description to answer the next two questions.

Waiters and waitresses at a local restaurant are being audited because the IRS thinks they are underreporting the tips they receive. An IRS agent predicts the tip left at lunch to be $\hat{y} = -0.40 + 0.20x$, where the explanatory variable is the price of the lunch for one person and the response variable is the amount of the tip.

- 166. If a person's lunch costs \$5, what does the IRS agent's model predict the person will leave the waiter/waitress as a tip?
 - A. \$0.20
 - B. \$0.40
 - C. \$0.60
 - D. \$1.39

167. For a \$5 lunch with a 55 cents (\$0.55) tip, what is the value of the residual?

- A. \$0.55
- B. \$0.60
- C. -\$0.05
- D. \$0.05
- 168. Which one of the following is true?

A. Correlations are always between 0 and 1.

B. When the correlation between two variables is 0, this means that the two variables are not associated in any way.

C. High correlations are always good evidence of cause-and-effect relationships.

D. None of the above statements are true.

- 169. TRUE or FALSE: If the correlation coefficient is greater than zero, it indicates that higher values of one variable are associated with higher values of the other variable.
- 170. TRUE or FALSE: The correlation coefficient, r, measures the proportion of the variation in the response variable y that is explained by the least-squares regression of y on x.
- 171. Which of the following pairs of variables cannot be studied using correlation?
 - A. Adult height and yearly income.
 - B. Smokestack height and air pollution particulate count.
 - C. Worker's age and number of sick days taken per year.
 - D. Family size and religion.

172. Which graph(s) is(are) useful for displaying the relationship between two quantitative variables?

- A. Scatterplot
- B. Bar Graph
- C. Stemplot
- D. Both A. and B.
- 173. The point in the lower righthand corner of the plot is

A. an influential point with a negative residual.

B. an influential point with a positive residual.

C. an outlier, but not an influential point.

D. neither an outlier nor an influential point.

0 5 10 15 20 25 30 35 22 24 26 28 30 32 x y

174. A scatterplot of fasting plasma glucose (in mg/dl) versus long term blood glucose (in percent) is displayed on page 155 (Figure 2.22) of your textbook. The least-squares regression line is drawn on the plot. Subject 18

A. has a large residual and is influential.

- B. has a small residual and is influential.
- C. has a large residual and is not influential.
- D. has a small residual and is not influential.
- 175. If the correlation between age of an automobile (x) and money spent for repairs (y) is 0.90, then A. 81% of the variation in the money spent for repairs is explained by the linear relationship with automobile age.

B. 81% of money spent for repairs is explained by the linear relationship with automobile age.

C. 90% of the variation in the money spent for repairs is explained by the linear relationship with automobile age.

D. 90% of money spent for repairs is explained by the linear relationship with automobile age.

- 176. A study is conducted to determine if one can predict the yield of a crop based on the amount of yearly rainfall. The response variable in this study is
 - A. yield of the crop
 - B. inches of water
 - C. the experimenter
 - D. amount of yearly rainfall
- 177. Calories and diet: A study is conducted to determine the relationship between the percentage of calories in the diet of juvenile mice that comes from protein and their average daily weight gain (in grams). The study measured the growth rates of mice with diets ranging from 0-60% protein. The estimated regression relationship was $\hat{y} = 0: 12 + 0: 007x$. Which of the following statements about the relationship is NOT reasonable?

A. Mice with no protein in their diet should gain an average of 0.12 grams per day.

B. An increase from a 20% to a 30% protein diet should result in an average increase in weight gain of 0.07 grams per day.

C. A mouse with a 40% protein diet should gain an average of 0.4 grams per day.

D. A mouse with a 90% protein diet should gain an average of $0.75~{\rm grams}$ per day.

- 178. TRUE or FALSE: Correlation is appropriate for investigating the relationship between two categorical variables.
- 179. **Bears:** Biologists are interested in the weight of the bear population. A bear's weight is a very difficult measurement to take. For this reason, it would be helpful to find a regression equation to predict a bear's weight (in pounds) based on other measurements that can be obtained more easily, such as length of the bear, head width, and chest girth. A random sample of 97 bears was used to obtain the following regression analyses.

Regression Analysis:

Weight versus Length

Weight = -405.03 + 9.8023 Length

Predictor	Coef	SE Coef	Т	Р
Constant	-405.03	34.28	-11.82	0.000
Length	9.8023	0.5599	17.51	0.000
S = 55.00	R-Sq = 7	6.3%	R-Sq(adj)	= 76.1%

Regression Analysis:

Weight versus Head Width

Weight = - 198.82 + 60.652 Head Width

Predictor	Coef	SE Coef	Т	Р
Constant	-198.82	32.86	-6.05	0.000
Head Width	60.652	5.038	12.04	0.000

S = 71.14 R-Sq = 60.4% R-Sq(adj) = 60.0%

Regression Analysis:

Weight versus Chest Girth

Weight = - 266.64 + 12.6317 Chest Girth

Predictor	Coef	SE Coef	Т	Р
Constant	-266.64	13.61	-19.60	0.000
Chest Girth	12.6317	0.3685	34.28	0.000
S = 30.93	R-Sq = 92	2.5%	R-Sq(adj)	= 92.4%

Which is the best predictor of a bear's weight?

A. Length

B. Head Width

C. Chest Girth

```
180. Suppose the correlation coefficient for X and Y is known to be zero. We can conclude that
```

- A. X and Y have standard distributions.
- B. the variances of X and Y are equal.
- C. there exists no relationship between X and Y.
- D. there exists no linear relationship between X and Y.
- 181. **Predictors for Success:** Colleges are interested in predicting future success of incoming freshmen. Explanatory variables used to predict success include high school GPA (HS GPA), rank in high school graduating class (HS Rank), score on ACT exam (ACT score), and number of credits the student registers for the first semester (Term Credits). In one study, a simple random sample of 50 freshmen was taken and their first-semester GPA was recorded. A simple linear regression of the response variable versus each explanatory variable was used. Some results are provided in the table below.

Simple Linear Regression	r^2	\mathbf{S}	p-value
First-semester GPA vs. HS GPA	0.3622	0.4290	$3:79 \times 10^{-6}$
First-semester GPA vs. HS Rank	0.5066	0.3795	$9:18 \times 10^{-9}$
First-semester GPA vs. ACT Score	0.1311	0.5006	0.0098
First-semester GPA vs. Term Credits	0.0028	0.5363	0.7145

Among the explanatory variables examined here, what is the single best predictor of first-semester college GPA?

A. HS GPA

B. HS Rank

C. ACT Score

D. Term Credits

182. Sharks: Physical characteristics of sharks are of interest to marine researchers. Data on shark length (in feet) and jaw width (in inches) for 44 sharks were collected. The range of x values used to calculate the least-squares regression line was 9.4 to 22.8 feet. Because it is difficult to measure jaw width in living sharks, researchers would like to determine whether it is possible to accurately estimate jaw width from shark length, which is more easily measured.

Regression Analysis: JawWidth versus Length

The regression equation is

JawWidth = 0.688 + 0.963 Length

Predictor	Coef	SE Coef	Т	Р
Constant	0.688	1.299	0.53	0.599
Length	0.963	0.08228	11.71	0.000

S = 1.37574 R-Sq = 76.6% R-Sq(adj) = 76.0%

(a) Give the number for the proportion of variability in jaw width that is explained by the linear relationship with shark length? Round your answer to 4 decimal places.

(b) Predict the jaw width of a randomly-selected 17-foot shark. Round your answer to 4 decimal places.

(c) Calculate the 95% confidence interval for β_1 , the slope of the population regression line. Round your answer to 4 decimal places.

(d) Provide a written interpretation of the 95% confidence interval in terms of the problem.

(e) State the appropriate hypotheses to test if there is a linear relationship between shark length and jaw width.

(f) Based on the 95% confidence interval for β_1 , what is the decision about the null hypothesis at the 0.05 significance level?

(g) Briefly explain how you used the 95% confidence interval to make the decision.

183. Old Faithful geyser: Can we predict the time of the next eruption of the Old Faithful geyser? The Old Faithful geyser is the most popular attraction in Yellowstone National Park and tourists want to be sure to see at least one eruption. Park rangers help tourists by posting the predicted time of the next eruption. Below you will see regression output where the response variable is the time after the eruption (in minutes) versus duration of the eruption (in seconds). All the necessary assumptions are met.

Regression Analysis:

Time After Eruption versus Duration

Predictor	Coef	SE Coef	Т	Р
Constant	47.396	7.637	6.21	0.000
Duration	0.17969	0.03084	5.83	0.000

S = 7.002 R-Sq = 47.2% R-Sq(adj) = 45.8%

(a) Give the equation of the least-squares regression line for this problem.

(b) Predict the number of minutes until the next eruption (Time After) when the most recent eruption (Duration) lasted 250 seconds. Give 4 decimal places. You must draw a box around your answer.

(c) Continuation of (b). If the actual time after the 250-second eruption was 92 minutes. Calculate the residual for this data point. Give 4 decimal places. You must draw a box around your answer.

(d) Interpret the value of 0.17969 in the context of this problem.

(e) State the appropriate hypotheses to test for a linear relationship between Time After and Duration.

- [i.] What is the decision regarding H_o ? Use $\alpha = 0.05$.
- [ii.] State the conclusion in the context of this problem.

184. TRUE or FALSE: The test of slope for a linear relationship is a t-test with n-2 degrees of freedom.