

Please circle one. Section: 10:30 or 11:30

Please enter all of your answers on these exam pages. There are 25 questions. Notice the Defective Question Report page at the end of the exam.

1. (2 pts) In a designed experiment, treatments are randomly assigned to the study units (people or things). The reason for the randomization is to prevent bias due to other factors which are not controlled in the experiment.  
A) True B) False
2. (2 pts) If an experimental treatment is deemed to have harmful side effects, it is called a placebo.  
A) True B) False
3. (2 pts) A lurking variable is a variable that plausibly explains a relationship between two other variables.  
A) True B) False
4. (2 pts) In multiple regression the most basic residual plot is the graph of residuals versus fitted values.  
A) True B) False
5. (2 pts) For a quadratic model the *adjusted* R-squared will always be greater than the *adjusted* R-squared for a straight-line model with the same  $x$  and  $y$  variables.  
A) True B) False

6. (3 pts) A straight-line model was fit to several data points using least squares. *Some* of the results are shown in the table at the right. What is the slope of the least-squares line (to the nearest tenth)?

$x$	$y$	fitted	residual
7	510	504.877	5.12308
8	503	505.185	?
...	...	...	...
12	504	?	-2.41538
13	512	506.723	5.27692

- A) -7.0
  - B) 0.3
  - C) 0.4
  - D) 3.2
  - E) Cannot be determined from the information given.
7. (2 pts) Use the information contained in question 6. What is the value of the residual for the second case where  $x=8$  and  $y=503$ ? Show your work.
  8. (2 pts) Use the information contained in question 6. What is the fitted value for the next-to-last case where  $x=12$ ? Show your work.

9. (4 pts) A economist obtains a regression model that relates a person's annual Savings (in \$1000's), the person's annual Income (in \$1000's), and an indicator variable House where House=1 if the person owns a house and House=0 if not a homeowner. The fitted equation is  $\text{Savings} = -0.32 + 0.0675 \text{ Income} + 0.827 \text{ House}$ . Using this model, about how much *more* does a homeowner with annual income of \$50,000 save compared to a non-homeowner with annual income of \$20,000? Show your work.

10. (4 pts) Two regression models are developed to relate monthly Sales (in \$1000's) to monthly advertising expenditures (Ads in \$1000's). The first model is a straight-line model while the second uses Ads squared in a quadratic model. Here are some of the results:

Straight-Line Model				Quadratic Model		
Source	SS	DF		Source	SS	DF
Regression	858	1		Regression	889	2
Error	1230	10		Error	1199	9
Total	2088	11		Total	2088	11

Which model fits the data better? Explain your choice.

11. (3 pts) In a taste test of two pizza recipes, a coin was flipped to decide which tasters got which recipe. This is an example of

- A) replication
- B) blinding
- C) placebo
- D) randomization
- E) control

12. (2 pts) For a quadratic model R-squared will always be at least as large as the R-squared for a straight-line model with the same  $x$  and  $y$  variables.

- A) True
- B) False

13. (3 pts) A professor asked his class to guess his age in years. Half the class were given the hint that he received his B.A. degree in 1961. The other half received no hint. The half that received no hint is called the
- A) blind group
  - B) control group
  - C) factor group
  - D) treatment group
  - E) None of the above.
14. (3 pts) Consider the experiment reported in question 13. How many *factors* were considered in this experiment?
- A) 1
  - B) 2
  - C) 3
  - D) 4
  - E) None of the above.
15. (2 pts) A probability survey supports rigorous, objective inferences about the characteristics of a population.
- A) True
  - B) False
16. (2 pts) The least squares regression line is that line which makes the sum of squared vertical distances between the observations and the fitted line as small as possible.
- A) True
  - B) False
17. (3 pts) There is a strong positive correlation between the number of highly specialized doctors at a hospital and the proportion of patients that die at the hospital. This is an example of:
- A) a lurking variable
  - B) ecological correlation
  - C) cause and effect
  - D) consistency
  - E) mechanism
18. (2 pts) Ecological correlations are useful for measuring the strength and direction of straight-line relationships in groups and then applying that measure to individuals.
- A) True
  - B) False
19. (5 pts) A set of data pairs has the following summary statistics:  $\bar{x} = 40$ ,  $\bar{y} = 100$ ,  $s_x = 3$ ,  $s_y = 12$ , and  $r = 0.5$ . What is the equation for the least squares regression line in *original terms* of  $x$  and  $y$ ? Show your work.

20. (3 pts) In the multiple regression example given in the British Steel video, they used two continuous predictor variables (amount of carbon in the steel plate and amount of carbon in the welding wires) to predict the amount of carbon in the final weld. Their regression model is best described as

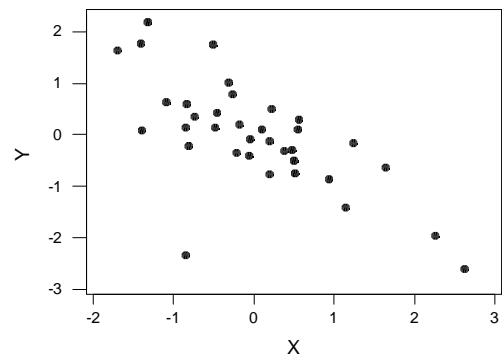
- A) one line
- B) two parallel lines
- C) one quadratic curve
- D) a quadratic surface
- E) None of the above.

21. (3 pts) Suppose  $y$  is a response variable of annual salary,  $x$  is a continuous predictor variable of years of experience, and  $z$  is a binary indicator variable indicating gender ( $z=1$  means female,  $z=0$  means male). Consider the regression model with fitted equation:  $\hat{y} = b_0 + b_1x + b_2z$ . What coefficient (or combination of coefficients) represents the predicted *starting* annual salary (experience=0) for a female?

- A)  $b_0$
- B)  $b_1$
- C)  $b_2$
- D)  $b_0+b_1$
- E)  $b_0+b_2$

22. (3 pts) What is the correlation coefficient for the data shown in the scatterplot at the right?

- A)  $-0.39$
- B)  $-0.51$
- C)  $-0.73$
- D)  $-0.95$
- E) The correlation cannot be calculated because of the outlier.



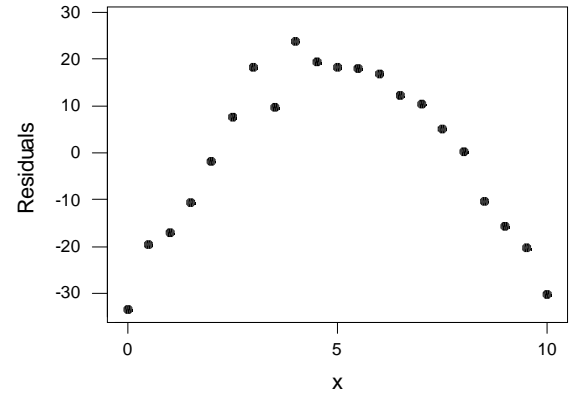
23. (2 pts) In the scatterplot for the previous question one point would be considered an outlier. Circle that point.

24. (2 pts) **Replication** is important in experimentation since process results vary even under carefully controlled experimental conditions.

- A) True
- B) False

25. (3 pts) The graph at the right shows a plot of residuals versus  $x$  from a fitted straight-line model. Which of the following best describes what we learn from this graph?

- A) Since the graph is mound-shaped the plot indicates that the straight-line model explains the relationship well.
- B) Since the residuals sum to zero the plot indicates that the straight-line model is a good fit for the relationship.
- C) Since about half of the residuals are positive and half negative the plot indicates that the straight-line model is a good fit for the relationship.
- D) Since the graph shows zero correlation between residuals and  $x$  the plot indicates that the straight-line model could be improved.
- E) Since the graph shows a systematic pattern in the residuals—negative then positive then negative—the plot indicates that the straight-line model could be improved—possibly with a quadratic model.



# Defective Question Report

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

Section: \_\_\_\_\_

If you believe that a test question is defective in some way, please list your complaint here. All complaints will be considered in our interpretation of the test results.

Question number: \_\_\_\_\_

Your complaint:

Question number: \_\_\_\_\_

Your complaint:

Question number: \_\_\_\_\_

Your complaint: