Exam 2 March. 20, 1997, Quant II Form A

The 40 exam questions and answer sheet are both to be turned in to your Discussion Section instructor at the end of the exam. Code your name and ID number on the answer sheet. **Code your Section number under OPTIONAL CODES in positions L M N**.

1. **Data production quality** refers to assessment of whether the data collected are relevant to the problem we wish to solve.

   A) True B) False

2. A factorial marketing experiment has two factors. The first factor has three levels and the second factor has four levels. If the experiment has three replications, how many observations will be recorded?

   A) 3
   B) 4
   C) 12
   D) 36
   E) None of the above.

3. A **placebo** is a nontreatment that is disguised so that the subjects think it is real.

   A) True B) False

4. The least squares regression line is that line which makes the sum of squared perpendicular distances between the observations and the fitted line as small as possible.

   A) True B) False

5. **Blinding** is the name for the process that randomly assigns treatments to subjects.

   A) True B) False

6. Regression models may always be thought of in the form

   Observed Response = Fitted Value + Residual

   Which of these three terms depends upon the model being used?

   A) Observed Response and Fitted Value
   B) Observed Response and Residual
   C) Fitted Value and Residual
   D) Observed Response, Fitted Value, and Residual
   E) None of the above.

7. In multiple regression modeling, the most basic residual plot is the plot of residuals versus the corresponding fitted values.

   A) True B) False

8. A portion of an ANOVA (analysis-of-variance) table from a multiple regression calculation is shown to the right. What is the value of *s* (the residual standard deviation) for this regression?

Source	SS	df
Regression	0.427	3
Error		
Total	0.949	93

A) 0.0058
B) 0.0762
C) 0.0102
D) 0.1423
E) 0.5220

9. The table below shows four data pairs together with two possible models—one linear and one quadratic. ***These models were not necessarily found using least squares.***

Data			**Curve I** $\hat{y} = 0.7x$				**Curve II** $\hat{y} = 1 - 0.5x + 0.25x^2$	
y	***x***		**FITTED**	**RESIDUAL**			**FITTED**	**RESIDUAL**
1	1						0.75	
1	2						1.00	
2	3		2.1					
3	4		2.8				3.00	

Which curve fits the data better in the sense of least squares?

A) Curve I fits better since its residuals add to zero.
B) Curve I fits better since its sum of squared residuals is smaller than for Curve II.
C) Curve I fits better since it is the least squares regression line for these data.
D) Curve II fits better since its sum of squared residuals is smaller than for Curve I.
E) Curve II fits better since two of its residuals are zero.

10. If a sequence of longitudinal data is random we expect the lag two autocorrelation coefficient to be near zero.

A) True B) False

11. The **major** difference between simple straight-line regression models and multiple regression models is:

A) Multiple regression models are curved.
B) Regression coefficients in multiple regression models are not found by least squares.
C) Multiple regression models have several response variables.
D) Multiple regression models have an associated adjusted R^2 value.
E) None of the above.

12. Here is the fitted regression equation for modeling points scored per minute for the NBA guards in the 92-93 season.

Pts/Min = –0.78 + 0.0056Min/Game + 0.0041Height + 0.0034%FT
with s = 0.076, R^2 = 45.0%, and R^2(adj) = 43.2%

Reggie Miller averaged 36.0 minutes per game, is 201 centimeters tall, and had an 88% free throw percentage. What is Reggie Miller's residual for this model?

A) 0.076
B) 0.558
C) –0.78
D) 43.2%
E) Cannot be determined from the information given.

13. Referring to the previous question: If the additional predictor variable, rebounds per minute, is added to the model what will happen to s and R^2?

A) R^2 will increase and s will decrease.
B) R^2 will increase and s will increase.
C) R^2 will increase but s may decrease or increase.
D) R^2 and s will both decrease.
E) Cannot say what either R^2 or s will do.

14. In a taste test of two recipes for pizza, a coin was flipped to decide which tasters got which recipe. This an example of:

A) randomization
B) replication
C) blinding
D) stratification
E) placebo

15. In the Physicians' Health Study described in the video, 22,000 male doctors were used as subjects to measure the effect of aspirin in the prevention of heart attacks. By saying that the experiment was **double blinded** we mean that neither the doctors nor the evaluators knew whether the subjects received aspirin or placebo.

A) True B) False

16. There is a high positive correlation between the number of fire trucks on the scene of a fire and the amount of damage suffered. This is an example of

A) a lurking variable
B) ecological correlation
C) cause and effect
D) consistency
E) mechanism

17. A company's research analyst developed a regression equation that used salary as a response and years of experience as a single predictor. She then displayed the residual plot shown below with plotting symbol o for males and + for females. Which of the following best describes the conclusions we get from this plot? The plot shows that

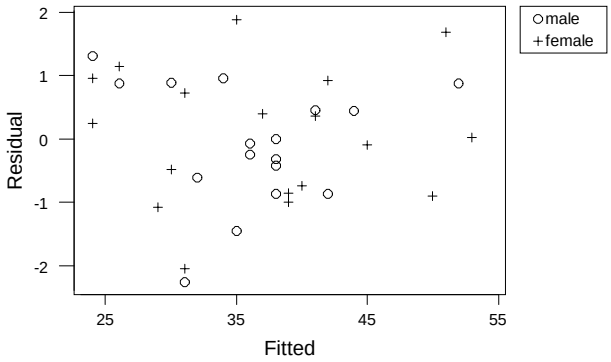
A) she should use a parallel-lines model instead with an indicator variable for gender.

B) the straight-line model captures the relationship between salary and years of experience *without* using an indicator for gender.

C) she should use a quadratic curve instead.

D) she should use a regression plane instead with years of experience and gender as continuous predictor variables.

E) None of the above.



18. For the sequence 5, 6, 6, 8, 10, find the lag one autocorrelation. (Hint: The mean for the sequence is 7 and the standard deviation is 2.)

A) 0.1873

B) 0.3125

C) 0.3750

D) 0.4428

E) 0.5096

19. A set of data pairs has the following summary statistics: $\bar{x} = 30$, $\bar{y} = 50$, $s_x = 3$, $s_y = 2$, and $r = 1/2$. What is the equation for the least squares regression line in ***original terms*** of x and y?

A) $\hat{y} = 1/2x$

B) $\hat{y} = 1/3x$

C) $\hat{y} = 27.5 + 3/4x$

D) $\hat{y} = 40 + 1/3x$

E) $\hat{y} = -7.5 + 3/4x$

20. The model GP100M = –1.3 + 0.0025Weight + 0.7Trans is obtained from fitting Gallons per 100 miles to Weight and Transmission type for a number of cars of various typical weights and a mix of transmission types. Here Weight is in pounds, Transmission type is an indicator variable with Trans = 1 for automatic transmissions, = 0 for manuals, and GP100M is gallons of gasoline used per 100 miles. A new car, the Ford Aspire, comes on the market. It weighs 2110 pounds, has a 4-cylinder engine and a 4 speed manual transmission. How many gallons per 100 miles does the regression model predict for the Ford Aspire? (Round to the nearest whole gallon.)

A) 3

B) 4

C) 5

D) 6

E) None of the above.

21. Refer to the preceding question: You are about to take a 100 mile trip in a typical rental car of unknown weight. It has an automatic transmission. How many *additional* gallons of gas does the model predict you will use if you take your 200 pound friend with you?

A) 0.07

B) 0.14

C) 0.25

D) 0.50

E) Cannot be determined from the information given.

22. Describe the lag one autocorrelation in the sequence plot shown at the right.

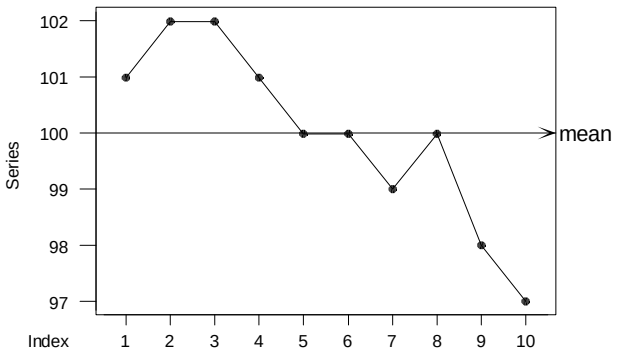
A) negative

B) near zero

C) positive

D) Not enough information given.

E) None of the above.



23. In the Physicians' Health Study described in the video, 22,000 male doctors were used as subjects to measure the effect of aspirin in the prevention of heart attacks. In this study the physicians receiving the placebo are said to be in the **control group**.

A) True B) False

24. The following table shows the responses of ten students to questions about their gender and their music preference.

Student ID	Gender†	Music Preference‡		Student ID	Gender	Music Preference
1	0	2		6	0	2
2	1	3		7	0	1
3	1	1		8	1	3
4	1	2		9	0	1
5	0	2		10	1	2

† 1 = female, 0 = male
‡ 1 = Country, 2 = Rock, 3 = Other

What percentage of students preferred Country?

A) 20%
B) 30%
C) 40%
D) 50%
E) None of the above.

25. Referring to the table in question 24: What percentage of females preferred Country?

A) 20%
B) 30%
C) 40%
D) 50%
E) None of the above.

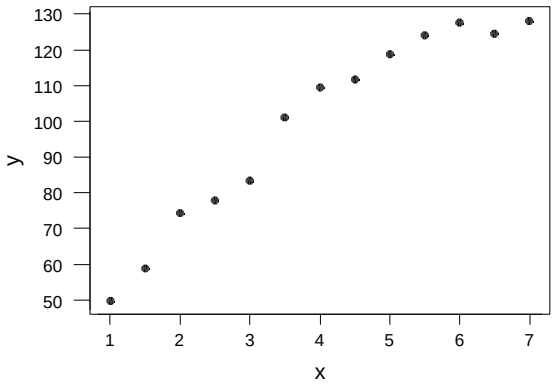
26. Referring to the table in question 24: What percentage of those who preferred Rock were males?

A) 20%
B) 30%
C) 40%
D) 50%
E) None of the above.

27. Ecological correlations are very useful for assessing the strength of relationships.

A) True B) False

28. Which of the following is closest to the correlation coefficient for the scatterplot shown at the right?



A) −0.9
B) 0.0
C) 0.4
D) 0.9
E) Cannot be calculated since the relationship is curved.

29. In class we developed the equation for compound interest: $P_t = P_0(1 + i)^t$ where P_t is the value of the investment at time t, P_0 is the initial investment, and i is the interest rate. Then using time as a predictor variable, we used this equation and straight-line regression to estimate the effective interest rate for S&P500 time series. The slope of the regression line is an estimate of

A) P_0
B) $\log(P_0)$
C) $1+i$
D) $\log(1+i)$
E) i

30. A simple experiment asked students to guess the instructors age. Randomly, half the class was given the hint that the instructor's oldest son graduated from Iowa in 1985. The other half received no hint. How many factors does this experiment have?

A) 1
B) 2
C) 3
D) 4
E) None of the above.

31. The Gallup Poll Organization reports a recent worldwide study of perceptions of quality. People were asked "In general, how would you rate the quality of manufactured goods produced in the United States; excellent, very good, good, only fair, or poor?" Then they reported a "Gallup Quality Score" as the percent who rated manufactured goods as either "Excellent" or "Very Good." This is an example of

A) a lurking variable
B) ecological correlation
C) cause and effect
D) consistency
E) collapsing tables

32. We modeled a car's gasoline efficiency using car weight x, and an indicator variable z for transmission type (1=automatic, 0=manual). Consider the regression model:

    $\hat{y} = b_0 + b_1x + b_2z + b_3zx$.

    What is the slope coefficient for the weight predictor for all automatic transmission cars?

    A) b_0

    B) b_1

    C) $b_0 + b_2$

    D) $b_1 + b_3$

    E) None of the above.

33. Refer to the previous question: If y is measured in gallons per 100 miles and weight is in pounds, what are the units for the coefficient b_0?

    A) pounds

    B) gallons per hundred miles

    C) gallons per hundred miles per pound

    D) It is unitless

    E) None of the above.

34. In a probability survey the persons to be interviewed are selected by chance.

    A) True B) False

35. A student team conducts an experiment that measures the height that a ball bounces on different surfaces. They use concrete, vinyl, hard carpet, and wood. How many levels are there in this experiment?

    A) 1

    B) 2

    C) 3

    D) 4

    E) None of the above.

36. Since correlation coefficients always lie between −1 and +1, they are not affected by outliers.

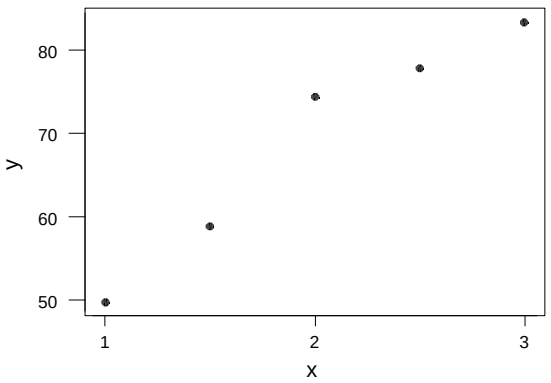
    A) True B) False

37. What is the correlation coefficient for the data in the table?

x	y							
201	1000							
203	1002							
202	1004							

A) 0.2
B) 0.3
C) 0.4
D) 0.5
E) None of the above.

38. The scatterplot at the right has a correlation coefficient of 0.97. If the case with x=3 is altered so that it's y-value is about 50, what will be the effect on the correlation?



A) It will be negative and close to −1.
B) It will be negative and close to zero.
C) It will be positive but close to zero.
D) It will be positive and close to 0.5.
E) It will not change.

39. The table below displays counts of people by gender and marital status.

Gender	Marital Status		
	Single	Married	Other
female	100	40	60
male	200	60	40

What percentage of the males are married? (Round to the nearest whole percent.)

A) 20%
B) 30%
C) 40%
D) 50%
E) None of the above.

40. Referring to the table in the previous question, what percentage of the people are male? (Round to the nearest whole percent.)

A) 20%
B) 30%
C) 40%
D) 50%
E) None of the above.

Formulas

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)\right] \qquad r_k = \frac{1}{n-1}\sum_{i=1}^{n-k}\left[\left(\frac{y_{k+i}-\bar{y}}{s}\right)\left(\frac{y_i-\bar{y}}{s}\right)\right]$$

$$b_1 = r\frac{s_y}{s_x} \qquad b_0 = \bar{y} - b_1\bar{x} \qquad s = \sqrt{\frac{1}{n-k-1}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$$

$$R^2 = 100\left[1-\frac{\text{Error SS}}{\text{Total SS}}\right]\% \qquad \text{adjusted } R^2 = 100\left[1-\frac{s^2}{s_y^2}\right]\%$$

Defective Question Report

Name:_____
Section:_____
ID: _____
Circle one: Form A B C D

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.

To correctly identify the test question we must know which **form** of the test you have taken. ***We also must know how you answered the question.***

Remove this last page from the exam questions and turn it in to one of the instructors in the course.

Question number: