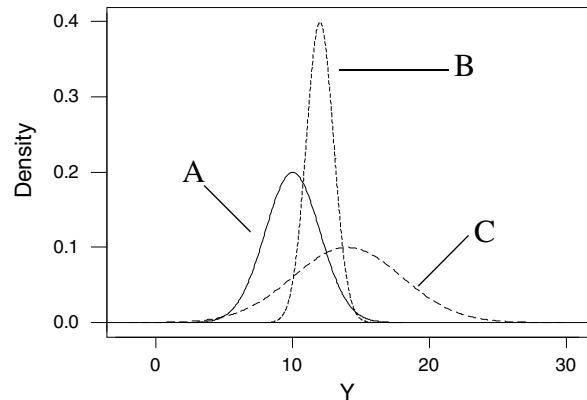


Be sure to code your Discussion Section number into the OPTIONAL CODES in positions JKL.

There are 40 questions. A table for the normal distribution is attached and a Defective Question Report is available from your instructor.

1. The graph at the right displays three normal curves with different means and different standard deviations. What is the correct order for the curves from *smallest* to *largest standard deviation*?

- A) A, B, C
- B) A, C, B
- C) B, A, C
- D) B, C, A
- E) C, A, B



2. Refer to question 1. What is the correct order for the curves from *smallest* to *largest mean*?

- A) A, B, C
- B) A, C, B
- C) B, C, A
- D) B, A, C
- E) C, B, A

3. An analysis of 17 pairs of data produced a least squares regression line with equation  $\hat{y} = 17 + 2x$ . What is the fitted value when  $x = 3$ ?

- A) 21
- B) 22
- C) 23
- D) 24
- E) 25

4. Refer to question 3. What is the residual when  $x = 3$  and  $y = 22$ ?

- A) -2
- B) -1
- C) 0
- D) 1
- E) 2

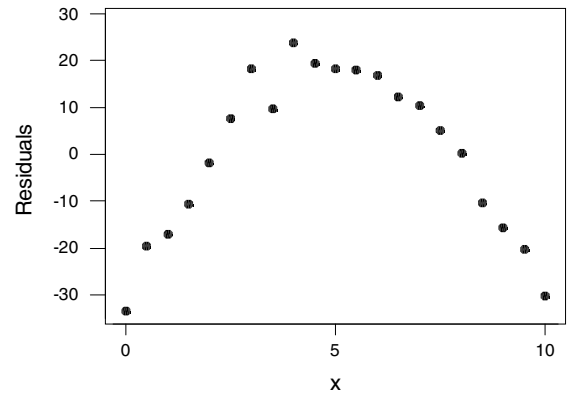
5. We have 28 pairs of  $x$ - $y$  data and are fitting a quadratic regression curve to these data by least squares. How many degrees of freedom do the residuals have?
- A) 24  
 B) 25  
 C) 26  
 D) 27  
 E) 28
6. A model of the form  $\hat{y} = b_0 + b_1x + b_2x^2$  has been used to model  $y$ , the height in inches of young girls, where  $x$  is the age of the girl in years. The least squares estimates are  $b_0 = 25$ ,  $b_1 = 4$ , and  $b_2 = -0.07$ . What are the units for the coefficient,  $b_1$ ?
- A) inches  
 B) inches per year  
 C) years  
 D) years per inch  
 E) inches per year squared
7. Refer to question 6. What does the model predict for the height of a 10 year old girl?
- A) 25  
 B) 58  
 C) 65  
 D) 72  
 E) None of the above.
8. A small table of data is given below. The means of the  $x$  and of the  $y$  are both 2. The standard deviations of both  $x$  and  $y$  are 1. What is the correlation coefficient between  $x$  and  $y$ ?

$x$	$y$			
2	3			
3	2			
1	1			

- A) 0.3333  
 B) 0.5000  
 C) 0.5773  
 D) 0.7071  
 E) 1.5000

9. An investigator is studying the relationship between Market value of houses and their size as measured by their square footage (Sq.ft). Her data consist of information on 60 houses and her Total SS = 2092.4. She first fits the model  $\text{Market} = b_0 + b_1 \text{ Sq.ft}$  and finds a Error SS of 919.3. She then fits a second model  $\text{Market} = b_0 + b_1 \text{ Sq.ft} + b_2 (\text{Sq.ft})^2$  and gets a new Error SS of 687.33. Which one of the following is true?
- A) The second model is better since  $919.3 > 687.33$
- B) The second model is better since the adjusted  $R^2$  for the second model is larger than for the first model.
- C) The second model is better since the  $R^2$  for the second model is larger than for the first model.
- D) The first model is better since  $919.3 > 687.33$
- E) The first model is better since the  $R^2$  for the first model is larger than for the second model.
10. An investigator is studying the relationship between Market value of houses and their size as measured by their square footage (Sq.ft) and a categorical variable that measures construction grade. Construction grade has three values that have been coded into indicator variables Low, Medium, and High. She fits a model  $\text{Market} = b_0 + b_1 \text{ Sq.ft} + b_2 (\text{Sq.ft})^2 + b_3 \text{ Low} + b_4 \text{ Medium}$ . This relationship when plotting Market versus Sq.ft would be best described as:
- A) a quadratic curve
- B) two parallel lines
- C) two parallel quadratic curves
- D) three parallel quadratic curves
- E) a plane
11. 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 male,  $z=0$  means female). Consider the regression model with fitted equation:  $\hat{y} = b_0 + b_1 x + b_2 z$ . What coefficient or combination of coefficients represents the predicted annual salary for a male with one year of experience?
- A)  $b_0$
- B)  $b_1$
- C)  $b_0 + b_1$
- D)  $b_0 + b_2$
- E)  $b_0 + b_1 + b_2$
12. 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  $y$  data.
- A) True    B) False
13. In multiple regression the most basic residual plot is the graph of residuals versus the corresponding fitted values.
- A) True    B) False

14. 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.

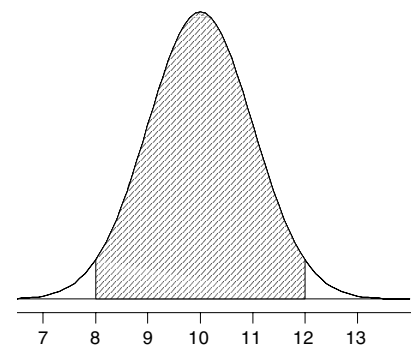
15. An analyst is using regression to model the Market value of houses using the size of the house, Sq. ft, and a categorical variable, Grade, that has values Grade = 1 for Low grade houses, Grade = 2 for Medium grade houses, and Grade = 3 for High grade houses. The model is  $\text{Market} = b_0 + b_1 \text{Sq. ft} + b_2 \text{Grade}$ . Which of the following statements best describes this model?

- A) This model is a parallel-planes model.
- B) This model is one plane.
- C) This model is a parallel-lines model with arbitrary vertical distances between the lines.
- D) This model is a parallel-lines model with equal vertical distances between the lines.
- E) None of the above.

16. In the general U.S. adult population, Stanford-Binet IQ scores are approximately normally distributed with a mean of 100. Just 2.5% of U.S. adults score above 132 on this IQ test. What is the standard deviation of the IQ scores?

- A) 1
- B) 2.5
- C) 16
- D) 32
- E) 116

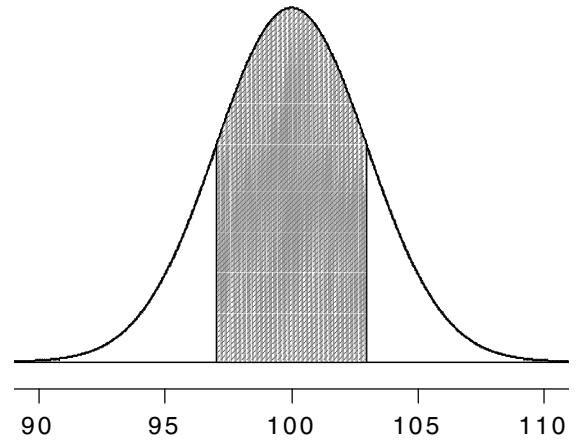
17. The graph at the right shows a normal curve with standard deviation of 1. What is the value of the shaded area?



- A) 0.0228
- B) 0.8413
- C) 0.9544
- D) 0.9772
- E) 0.9970

18. The 381 students in *Statistics for Business* took a 35 question exam. The average score was 28.4 and the standard deviation was 2.6. Class policy is that you must get 88% or more correct to get an A. Here that means you must get 31 questions or more correct. The distribution of scores is approximately a normal curve. About how many students got A's on this test?
- A) 20
  - B) 29
  - C) 37
  - D) 48
  - E) 61

19. The graph at the right shows a normal curve. The shaded area is 68% of the total area. What is the value of the standard deviation?



- A) 0.5
- B) 1
- C) 3
- D) 5
- E) 7

20. A small table of data pairs is shown to the right. A straight line is to be fit to these data by least squares. What is the value of the Residual (or Error) Sum of Squares? (THINK! Do not calculate!)

X	Y
1	2
2	4
3	6

- A) 0
- B) 1
- C) 2
- D) 3
- E) 4

21. A distribution of exam scores is strongly skewed toward the low values. Lots of people got good scores but a few got very low scores. The mean of the distribution is 85 (out of 100) and the standard deviation is 10. What percentage of the scores are 75 or below?

- A) 16%
- B) 32%
- C) 68%
- D) 95%
- E) Cannot be determined from the information given since the distribution of scores is skewed.

22. Data were collected from batches produced at a chemical plant. The variables measured were final product yield, set time, solid catalyst used, and liquid catalyst used. A portion of an ANOVA (analysis-of-variance) table from a multiple regression calculation is shown to the right. What is the Sum of Squares for Regression?

Source	SS	df
Regression		3
Residual Error	5010	11
Total	51533	14

- A) 5010
- B) 46523
- C) 51533
- D) 56543
- E) None of the above.

23. Refer to question 22. What is the residual standard deviation?

- A) 21.34
- B) 60.67
- C) 124.5
- D) 455.45
- E) 3680.93

24. Refer to question 22. How many batches were tested and measured?

- A) 11
- B) 12
- C) 13
- D) 14
- E) 15

25. The distribution of birthweight of babies in Singapore is approximately a normal distribution with mean 3135 grams and standard deviation 459 grams. What fraction of babies have birthweights from 2377.65 grams to 4007.10grams?

- A) 0.45
- B) 0.90
- C) 0.92
- D) 0.95
- E) 0.997

26. A company manufactures bolts with a mean length of 6 inches and a standard deviation of 0.06 inches. If a random sample 200 bolts is taken, how many bolts would you expect to be longer than 6.03 inches?
- A) 6
  - B) 12
  - C) 62
  - D) 100
  - E) 150
27. The Mountain Dew bottling company fills “32-ounce” bottles with a process that produces approximately a normal distribution of fill volume with a mean of 32.5 ounces and a standard deviation of 0.5 ounces. In the long run, what percentage of bottles are filled with less than 32 ounces?
- A) 2.5%
  - B) 5%
  - C) 16%
  - D) 32%
  - E) 68%
28. Refer to question 27. Joe purchases 4 “32-ounce” bottles of Mountain Dew for a big weekend party. Joe paid for 128 ounces (4 times 32) of Mountain Dew. What are the chances Joe got what he paid for (or even more)? (To the nearest whole percentage point.)
- A) 16%
  - B) 32%
  - C) 68%
  - D) 95%
  - E) 98%
29. The accounts receivable for a computer software company have a skewed distribution with a mean of \$125 and a standard deviation of \$25. An auditor decides to randomly sample 100 such accounts. What is the chance that the mean of these 100 accounts is between \$125 and \$128?
- A) 0.1151
  - B) 0.3849
  - C) 0.7698
  - D) 0.8849
  - E) Cannot be determined since the distribution of accounts receivable is skewed and not normal.
30. As long as a sequence of subgroup means remains within the control limits there is no evidence that special causes are influencing the process mean.
- A) True
  - B) False

31. Tampering with a process that is in control will usually lead to increased variability.

- A) True B) False

32. We used an example of measuring the variation in Attentiveness in Lecture. In that example “Michelle’s daydreaming about spring break in Mexico” would be considered a *common cause* of variation.

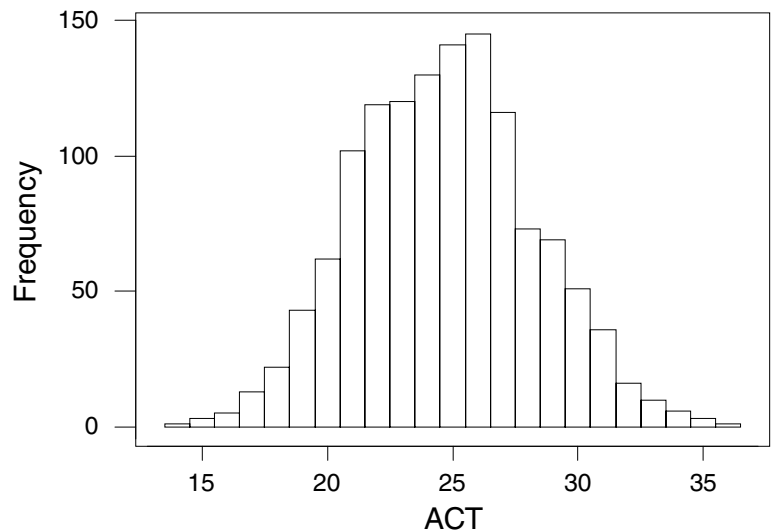
- A) True B) False

33. The results produced by a constant-cause system vary, and may vary over a wide band or a narrow band.

- A) True B) False

34. The histogram at the right displays the distribution of ACT composite scores for students entering the University of Iowa, Fall 2000. Which of the following is the standard deviation for this distribution?

- A) 0.05  
 B) 0.94  
 C) 3.6  
 D) 8.2  
 E) 12.1



35. A least squares regression line is fitted to the weights (in kilograms) of a group of children between the ages of 12 and 36 months. The equation of the line is  $\text{Weight} = 6.9 + 0.27 \text{ Age}$ . Weight is in kilograms and Age is in months. Which one of the following statements is true?

- A) The predicted weight for a two year old is 7.44 kilograms.  
 B) The predicted weight for a two year old is 13.38 kilograms.  
 C) It is not appropriate to use this equation for a two year old since 2 is not between 12 and 36.  
 D) A baby weighs about 6.9 kilograms at birth.  
 E) None of the above are true.

36. A model has been fitted by least squares to a set of 5 observations. Some of the values obtained from this fit are shown in the table at the right. From this information alone, you are able to construct all but which of the following?

- A) the missing residual  
 B) All the  $y$  values  
 C) the standard deviation of the  $x$  variable  
 D) the residual standard deviation  
 E) The correct answer is not given by A, B, C, or D.

$x$	RESI1	FITS1
3.5	-0.3713	5.3713
5.7	0.0570	6.9430
1.0	0.4147	3.5853
8.1		8.6575
5.0	-0.4429	6.4429

37. What is the area under a standard normal curve between  $z = -0.81$  and  $z = 0.5$ ?
- A) 0.2090
  - B) 0.2910
  - C) 0.4825
  - D) 0.6915
  - E) 0.9005
38. An aptitude test administered to management trainees results in scores that are normally distributed with a mean of 90 and a standard deviation of 20. If the top 3% of the test takers are given Certificates of Special Merit, how high must you score to receive that certificate?
- A) 109
  - B) 117
  - C) 121
  - D) 128
  - E) 134
39. What is the area under a standard normal curve between  $z = 0$  and  $z = 4.5$ ?
- A) 0.0000
  - B) 0.4500
  - C) 0.5000
  - D) 1.0000
  - E) Cannot be determined with our normal table.
40. A least squares regression line is to be used to predict students second exam score from the first exam score. On the first exam the mean score was 32 and the standard deviation was 3.4. The same students had a mean score on the second exam of 29 and the standard deviation was 4.2. The correlation coefficient between the two sets of scores was 0.63. What is the equation of the least squares line?
- A)  $\text{Exam2} = 4.1 + 0.7782 \text{ Exam1}$
  - B)  $\text{Exam2} = 29.0 + 0.7782 \text{ Exam1}$
  - C)  $\text{Exam2} = 19.0 + 0.6300 \text{ Exam1}$
  - D)  $\text{Exam2} = 29.0 + 0.6300 \text{ Exam1}$
  - E)  $\text{Exam2} = 32.0 + 0.6300 \text{ Exam1}$