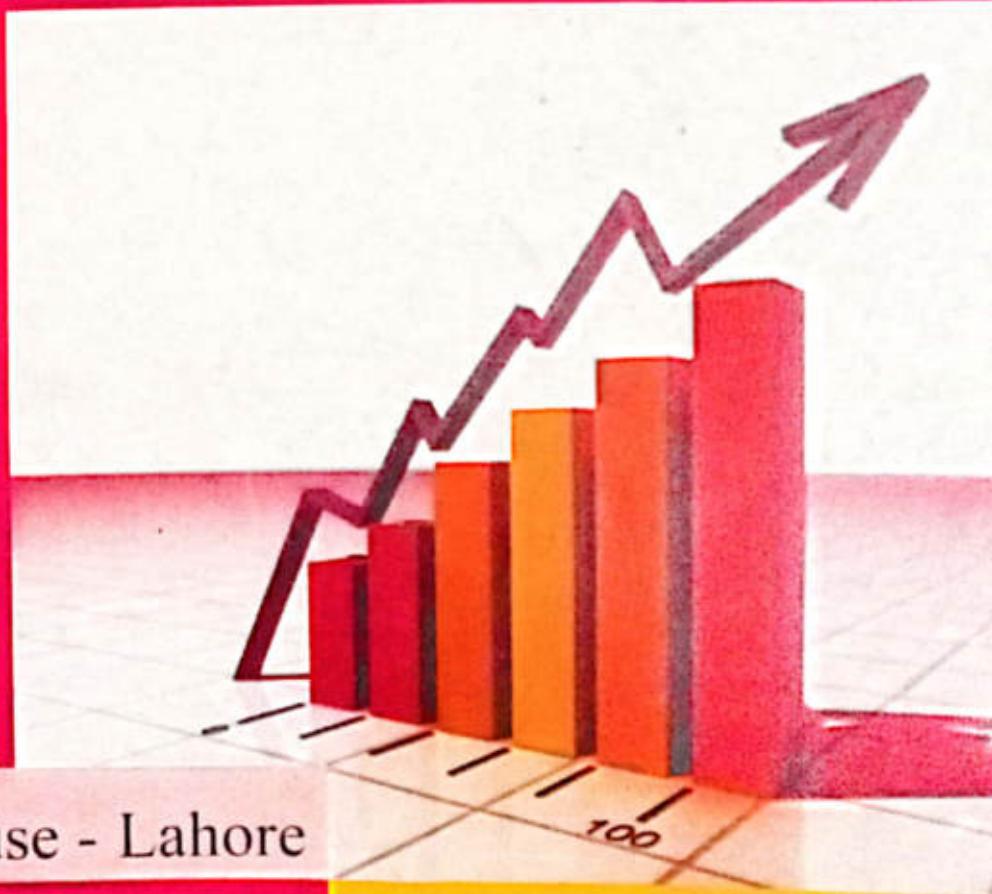




# Compendium of **STATISTICS**

## MCQs

Lectureship,  
Subject Specialist,  
Statistical Officer,  
S.S.S Statistics, M.A/M.Sc. Statistics  
and Economics, GRE Statistics &  
Other Competitive Examination



Caravan Book House - Lahore

# CONTENTS

1.	INTRODUCTION .....	1
	Answers .....	4
2.	PRESENTATION OF A DATA .....	5
	Answers .....	9
3.	MEASURES OF CENTRAL TENDENCY AND DISPERSION .....	10
	Answers .....	18
4.	PROBABILITY AND PROBABILITY DISTRIBUTIONS .....	19
	Answers .....	52
5.	REGRESSION ANALYSIS AND ECONOMETRICS .....	53
	Answers .....	85
6.	TIME SERIES AND FORECASTING .....	86
	Answers .....	93
7.	SURVEY SAMPLING AND RESEARCH METHODOLOGY .....	94
	Answers .....	123
8.	STATISTICAL INFERENCE .....	124
	Answers .....	154
9.	DESIGN OF EXPERIMENT .....	155
	Answers .....	166
10.	MULTIVARIATE ANALYSIS .....	157
	Answers .....	171
11.	NON PARAMETRIC METHODS .....	172
	Answers .....	176
12.	STATISTICAL QUALITY CONTROL .....	177
	Answers .....	204
13.	INDEX NUMBERS .....	205
	Answers .....	209
	LECTURER (STATISTICS) TEST PAPER, 2011 PPSC .....	210
	Answers .....	217
	PAST LECTURER (FPSC), STATISTICAL OFFICER AND GRE TESTS QUESTIONS .....	218
	Answers .....	232
	CSS, STATISTICS PAPER 2001 .....	232
	Answers .....	234
	CSS, STATISTICS PAPER 2003 .....	234
	Answers .....	235
	CSS, STATISTICS PAPER 2004 .....	235
	Answers .....	235
	CSS, STATISTICS PAPER 2005 .....	237
	Answers .....	237
	CSS, STATISTICS PAPER 2006 .....	238
	Answers .....	238
	CSS, STATISTICS PAPER 2007 .....	239
	Answers .....	239
	CSS, STATISTICS PAPER 2008 .....	241
	Answers .....	241
	CSS, STATISTICS PAPER 2009 .....	241
	Answers .....	243
	CSS, STATISTICS PAPER 2010 .....	243
	Answers .....	243
	CSS, STATISTICS PAPER 2011 .....	245
	Answers .....	245

## INTRODUCTION

1. The origin of statistics can be traced to  
a) State  
b) Commerce  
c) Economics  
d) Industry
2. Statistics may be called the science of counting' is the definition given by  
a) Croxton  
b) A.L. Bowley  
c) Boddington  
d) Webster.
3. Raw data means  
a) Primary data  
b) Secondary data  
c) Data collected for investigation  
d) Well classified data.
4. Which of the following are not the limitations of statistics?  
a) Statistics study the individuals  
b) Statistics tables may be misused  
c) Statistical laws are exact  
d) Both a & c but not b  
e) None of the above
5. Which of the following level of measurements in which zero is meaningful?  
a) Ordinal scale  
b) Ratio Scale  
c) Interval Scale  
d) Nominal Scale  
e) None of the above
6. Which of the following mathematical structure of ordinal scale?  
a) Permutation Group  
b) Isotonic Group  
c) General linear Group  
d) None of the above
7. Numbering of brands is an example of  
a) Ordinal Data  
b) Ratio Data
8. Interval Data  
d) Nominal Data  
e) None of the above
9. Advertising expenditure is an example of  
a) Ordinal Data  
b) Ratio Data  
c) Interval Data  
d) Nominal Data  
e) None of the above
10. A Qualitative data is  
a) Ordinal  
b) Nominal  
c) Both a) & b)  
d) None of the above
11. A Quantitative data is  
a) Interval Scale  
b) Ratio Scale  
c) Both a) & b)  
d) Nominal
12. ID number or workers, Zip code and telephone number are the examples of  
a) Ordinal data  
b) Ratio Scale data  
c) Interval scale data  
d) Nominal data
13. Which is a categorical factor?  
a) Temperature used within an oven  
b) The horizontal location of the logo on a web page  
c) Type of tire used on a motorcycle  
d) All of the above are categorical factors  
e) All of the above are correct except a) and d)

14. Which scale is the simplest form of measurement?  
a) Nominal  
b) Ordinal  
c) Interval  
d) Ratio
15. If a baseball coach calculates batting averages, what scale would be used?  
a) Interval scale  
b) Ratio scale  
c) Nominal scale  
d) Ordinal scale
16. Most of the outcome/dependent variable characteristics and attributes measured in educational research probably exist at the \_\_\_\_\_ level of measurement.  
a) Nominal  
b) Ordinal  
c) Interval  
d) Ratio
17. An ordinal scale is:  
a) The simplest form of measurement  
b) A rank-order scale of measurement  
c) A scale with equal intervals between adjacent numbers  
d) A scale with an absolute zero point  
e) A categorical scale
18. Which of the following is the correct order of Stevens' four levels of measurement?  
a) Ordinal, nominal, ratio, interval  
b) Nominal, ordinal, interval, ratio  
c) Interval, nominal, ordinal, ratio  
d) Ratio, interval, nominal, ordinal
19. A condition or characteristic that can take on different values or categories is called.  
a) A constant  
b) A variable  
c) A cause-and-effect relationship  
d) A descriptive relationship  
e) None of the above
20. Which of the following includes examples of quantitative variables?  
a) Age, temperature, income, height
- b) Grade point average, anxiety level, reading performance  
c) Gender, religion, ethnic group  
d) Both a) and b)
21. What is the opposite of a variable?  
a) A constant  
b) An extraneous variable  
c) A dependent variable  
d) A data set
22. A number derived from sample data which describes the data in some useful way is called a:  
a) Constant  
b) Statistic  
c) Parameter  
d) Critical value.
23. Which of the following are examples of a variable?  
a) Gender of a high school graduate  
b) Number of major credit cards a person has  
c) Type of automobile transmission  
d) All of the above
24. A variable's level of measurement indicates the:  
a) Number of questions used to measure the variable  
b) Number of categories which can be used to group scores on the variable  
c) Kinds of comparisons that can be made between cases in different categories  
d) Correspondence between conceptual and operational definitions
25. Which of the following is an example of nominal level measurement?  
a) Family size (number of children in a family)  
b) Political participation (number of times voted in last 10 years)  
c) Educational attainment (highest year or grade in school completed)  
d) Political party affiliation (Republican, Democrat, Independent, Other)

26. Suppose a survey item asks someone if something is very important, somewhat important, not very important, or unimportant. This is an example of what level of measurement?
- Nominal
  - Ordinal
  - Interval
  - Ratio
27. What level of measurement do most inferential statistics relies upon?
- Nominal
  - Ordinal
  - Interval
  - Ratio
28. An example of a measurement scale with an interval rather than a ratio scale is:
- 1 = blue, 2 = green, 3 = red.
  - Temperature in degrees centigrade.
  - Area in square millimeters.
  - Priority ranking, such as first, second, third.
29. The following data were collected on the diameters of turned shafts: 2.506 2.508 2.505 2.505. These values are:
- Attribute data.
  - Discrete data.
  - Variables data.
  - Continuous data.
- I and II
  - I only
  - II only
  - I and IV
  - III and IV
30. Suppose a large image file is downloaded from the Internet. The speed of the data, in bits per second (bps), is plotted as a function of time in seconds. In this situation, data speed is considered
- The dependent variable
  - The independent variable
  - A constant function
  - Nondecreasing
31. Which of the following is an example of a discrete variable?
- The direction of the wind as a tornado passes.
  - The number of car accidents per month in a certain town.
  - The overall loudness of sound during a symphony.
  - The speed of a car on a highway.
  - The thrust of a jet engine during an airline flight.
32. Which of the following statements are correct?
- Color of ten automobiles recently purchased at a certain dealership is an example of a univariate data set.
  - Height and weight for each basketball player on B.Z University team is an example of bivariate data set.
  - The systolic blood pressure, diastolic blood pressure, and serum cholesterol level for each patient participating in a research study is an example of multivariate data set.
  - None of the above statements is correct.
  - All of the above statements are correct.
33. Which of the following statements are correct?
- Probability reasons from the population to the sample (deductive reasoning), whereas inferential statistics reasons from the sample to the population (inductive reasoning).
  - Hypothesis testing and estimation by confidence intervals are the least important types of inferential statistical procedures.
  - In a probability problem, properties of the population under study are assumed to be unknown.
  - In a statistics problem, characteristics of a sample are assumed to be unknown.

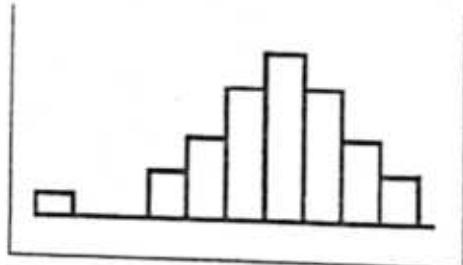
34. Which of the following statements are false?
- In enumerative studies, interest is focused on a finite, identifiable, unchanging collection of individuals or objects that make up a population.
  - "All companies listed on the Pakistan Stock Exchange" is an example of a sample.
  - A sampling frame is a listing of the individuals or objects to be sampled.
  - "All possible yields (in grams) from a certain chemical reaction carried out in a laboratory" is an example of conceptual or hypothetical population.
35. Which of the following statements are false?
- A variable is continuous if its set of possible values either is finite or else can be listed in an infinite sequence in which there is a first number, a second number, and so on.
  - Continuous variables arise from making measurements.
  - The frequency of any particular observation of a discrete variable  $x$  is the number of times that value occurs in the data set.
  - In theory, the relative frequencies should sum to 1.0, but in practice
- the sum may differ slightly from 1.0 due to rounding
- e) None of the above
36. A study was conducted in America to find out the proportion of blacks to white Americans in California. The variable chosen is.
- Nominal
  - Ordinal
  - Continuous
  - Discrete numerical
  - Dichotomous

## ANSWERS

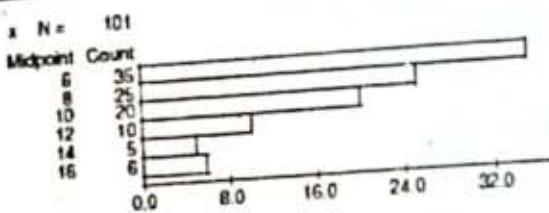
1	a	16	b	31	b
2	b	17	b	32	e
3	a	18	b	33	a
4	a	19	b	34	b
5	d	20	d	35	a
6	b	21	a	36	e
7	a	22	b		
8	b	23	d		
9	c	24	c		
10	c	25	d		
11	c	26	b		
12	c	27	c		
13	c	28	b		
14	a	29	e		
15	b	30	a		

## **PRESENTATION OF A DATA**

1. In an exclusive class interval  
a) The upper class limit is exclusive.  
b) The lower class limit is exclusive.  
c) The lower and upper class limits are exclusive.  
d) None of the above
2. Class intervals of the type 30-39, 40-49, 50-59 represents  
a) Inclusive type  
b) Exclusive type  
c) Open-end type  
d) None of these
3. Which of the following is one dimensional diagram.  
a) Bar diagram  
b) Pie diagram  
c) Cylinder  
d) Histogram
4. Percentage bar diagram has  
a) Data expressed in percentages  
b) Equal width  
c) Equal interval  
d) Equal width and equal interval
5. Frequency curve  
a) Begins at the origin  
b) Passes through the origin  
c) Begins at the horizontal line.  
d) Begins and ends at the base line.
6. With the help of histogram we can draw  
a) Frequency polygon  
b) Frequency curve  
c) Frequency distribution  
d) All the above
7. When the collected data is grouped with reference to time, we have  
a) Quantitative classification  
b) Qualitative classification  
c) Geographical Classification  
d) Chronological Classification
8. Most quantitative classifications are  
a) Chronological  
b) Geographical  
c) Frequency Distribution  
d) None of these
9. In tabulation caption stands for  
a) A numerical information  
b) The column headings  
c) The row headings  
d) The table headings
10. A simple table contains data on  
a) Two characteristics  
b) Several characteristics  
c) One characteristic  
d) Three characteristics
11. The headings of the rows given in the first column of a table are called  
a) Stubs  
b) Captions  
c) Titles  
d) Reference notes
12. The first step for constructing frequency distribution  
a) Record the number of data points  
b) Determine the number of classes  
c) Determine the range of data  
d) None of the above
13. A graph of cumulative distribution is called  
a) Histogram  
b) Ogive  
c) Probability plot  
d) None of the above



14. Which of the following best describes the shape of the histogram above.
- Approximately normal
  - Skewed left
  - Skewed right
  - Approximately normal with an outlier
  - Symmetric



15. For the histogram pictured above, what is the class interval (boundaries) for the class that contains the median of the data?
- (5, 7)
  - (9, 11)
  - (11, 13)
  - (15, 17)
  - (7, 9)
16. Boxplots have been used successfully to describe
- Center of a data set
  - Spread of a data set
  - The extent and nature of any departure from symmetry
  - Identification of "outliers"
  - All of the above

17. The five-number summary for a one-variable dataset is {5, 18, 20, 40, 75}. If you wanted to construct a modified boxplot for the dataset (that is, one that would show outliers if there are any), what would be the maximum possible length of the right side "whisker"?
- 35
  - 33
  - 5
  - 55
  - 53

18. In a frequency distribution:
- The frequency is the dependent variable
  - The median is always equal to the mean

- The mode represents the average value
- The mean is always equal to the mode

19. The largest cumulative absolute frequency in a set of numbers is equal to

- The sum of all the individual absolute frequency values
- Twice the mean
- Twice the median
- Twice the mode

20. The box plot was introduced by
- Gauss
  - Fisher
  - John Tukey
  - None of the above.

21. The purpose of histogram is to show
- Centering of data
  - Dispersion (spread) of data
  - Shape of the data
  - All of the above

22. The term fuzzy truth is used to describe
- A theory in which there are degrees of truth that span a range
  - Standard deviation
  - Cumulative frequency
  - The probability fallacy
  - Any continuous distribution

23. A graph that shows proportions that look like slices of a pizza is called
- A histogram
  - A slice graph
  - A pie graph.
  - A bar graph
  - A nomograph

24. A variable-width histogram is an excellent scheme for showing
- Proportions.
  - Correlation
  - Medians
  - Variances
  - Ranges

25. When the values of a function are shown on a coordinate system for selected points, and adjacent pairs of

- points are connected by straight lines, the resulting illustration is
- A quadratic graph
  - A bar graph
  - A venn diagram
  - A histogram
  - A point-to-point graph.
26. The stem-and-leaf plot has the following advantage over the histogram:
- It provides additional information
  - It provides a time reference
  - It provides a better indication of the shape of the data
  - It is more appropriate for discrete data
  - None of the above
27. A dimension has tolerance 15.100 to 15.110. To simplify the data collection sheets, the 15 is omitted so values such as .102 are recorded. This is an example of:
- Data coding.
  - Data manipulation.
  - Stem-and-leaf separation.
  - Data shifting.
  - Rational subgroups.
28. Maintaining a good calibration schedule will help with:
- Data accuracy.
  - Data precision.
  - Data integrity.
  - Data security.
  - Data coding.
29. A graph that uses vertical bars to represent data is called a:
- Line graph
  - Bar graph
  - Scatterplot
  - Vertical graph
30. Which of the following statements are correct?
- In general, a stem-and-leaf display based on between 15 and 25 stems is recommended.
  - The observation  $X_1$  in any numerical data set
- $X_1, X_2, X_3, \dots, X_n$  the smallest observation in the set, whereas the observation  $n x$  is the largest.
- A quick way to obtain an informative visual representation of a numerical data set is to construct a stem-and-leaf display.
  - A stem-and-leaf display shows the order in which observations were obtained.
  - All of the above
31. Which of the following statements are false?
- A time-series plot is a graphical representation of the observations in a data set over time.
  - A dotplot is an attractive summary of numerical data when the data set is reasonably large or there are relatively many distinct data values.
  - A dotplot gives information about location, spread, extremes, and gaps in a data set.
  - Multiplying a relative frequency by 100 gives a percentage.
32. Which of the following statements are correct?
- Constructing a histogram for continuous data (measurements) entails subdividing the measurement axis into a suitable number of class intervals or classes, such that each observation is contained in exactly one class.
  - The reaction time to a particular stimulus is an example of a discrete variable.
  - Constructing a histogram for discrete data is generally not different from constructing a histogram for continuous data.
  - The total area of all rectangles in a density histogram is 100.
33. Which of the following statements regarding histograms are correct?
- A unimodal histogram is one that rises to a single peak and then

## MEASURES OF CENTRAL TENDENCY AND DISPERSION

1. Which of the following measures the central tendency?
  - a) Mean
  - b) Median
  - c) Mode
  - d) a), b) and c)
2. For ordinal data, which of the following measure of central tendency is used?
  - a) Mean
  - b) Median
  - c) Mode
  - d) Geometric mean
3. Which of the following measure of central tendency cannot be affected by extreme values
  - a) Mean
  - b) Median
  - c) Mode
  - d) Both b) & c) but not a
  - e) None of the above
4. Which of the following measure of central tendency can be calculated for open-end classes
  - a) Geometric Mean
  - b) Mode
  - c) Harmonic Mean
  - d) Weighted Mean
  - e) None of the above
5. Ogives for more than type distribution intersect at
  - a) Mean
  - b) Median
  - c) Mode
  - d) Origin
6. Which of the following are the measures of dispersion?
  - a) Standard Deviation
  - b) Variance
  - c) Range
7. Which of the following is the absolute measure of dispersion?
  - a) Range.
  - b) Quartile Deviation.
  - c) Mean Deviation.
  - d) All of the above.
  - e) None of the above.
8. The goal of \_\_\_\_\_ is to focus on summarizing and explaining a specific set of data.
  - a) Inferential statistics
  - b) Descriptive statistics
  - c) None of the above
  - d) All of the above
9. Which measure of central tendency would be appropriate if you had mostly extremely high scores and extremely low scores on some item measured at the ratio level?
  - a) Mean
  - b) Median
  - c) Mode
  - d) Octal
10. Which of the following is the relative measure of dispersion?
  - a) Coefficient of range.
  - b) Coefficient Quartile Deviation.
  - c) Coefficient Mean Deviation.
  - d) All of the above
  - e) None of the above
11.  $CV[aX] =$  \_\_\_\_\_
  - a)  $a^2 CV(X)$
  - b)  $aCV(X)$
  - c)  $CV(X)$
  - d) Non of the above

12. The Standard deviation of marks of 100 students is 13. Every student is awarded later 5 or more marks. The standard deviation of the new set of marks is  
 a) 13  
 b) 18  
 c)  $13 + \sqrt{5}$   
 d)  $13 - \sqrt{5}$   
 e) None of the above
13. The Arithmetic Mean of two non-negative numbers is 5. Their standard deviation will be  
 a) Less than 4  
 b) At least 4  
 c) At most 5  
 d) Greater than 5  
 e) Nothing can be said based on the given information.
14. The arithmetic mean of various powers of deviation taken from the mean of distribution is called  
 a) Standard deviation  
 b) Skewness.  
 c) Deciles  
 d) Moments  
 e) None of the above
15. First moment about mean is always  
 a) Less than 1  
 b) Greater than 1  
 c) Zero  
 d) Less than zero  
 e) Greater than zero
16. For a set of observations, the mean is 58, the median is 62 and mode is 63. the data are  
 a) Skewed to left  
 b) Skewed to right  
 c) Symmetric about median  
 d) Bimodal
17. Suppose that the 34th percentile of Msc Statistics students marks is 34 this means that (SO, SBP, 2008)  
 a) Average marks of students is 34  
 b) 34 students obtained marks below 34  
 c) 34 students obtained marks above 34  
 d) All of the above
18. Which of the following is not a measure of location?  
 a) The mean  
 b) The median  
 c) The mode  
 d) The variance  
 e) All of the above
19. Suppose a large number of people take a test, and every single student gets exactly half of the answers right. In this case, the standard deviation is  
 a) Equal to the mean  
 b) Equal to the median  
 c) Equal to zero  
 d) Impossible to determine without more information
20. In a frequency distribution:  
 a) The frequency is the dependent variable  
 b) The median is always equal to the mean  
 c) The mode represents the average value  
 d) The mean is always equal to the mode
21. A tabulation of cumulative absolute frequency values is useful in determining  
 a) The mode  
 b) The variance  
 c) The median  
 d) The mean
22. Imagine that 11 people take a 10-question test. Suppose one student gets 10 correct answers, one gets 9 correct, one gets 8 correct, and so on, all the way down to one student getting none correct. What is the mean score, accurate to three decimal places?  
 a) 4.545  
 b) 5.000  
 c) 5.500  
 d) It is not defined.

23. Imagine that 11 people take a 10-question test. Suppose one student gets 10 correct answers, one gets 9 correct, one gets 8 correct, and so on, all the way down to one student getting none correct. What is the median score, accurate to three decimal places?
- 4.545
  - 5.000
  - 5.500
  - It is not defined.
24. Imagine that 11 people take a 10-question test. Suppose one student gets 10 correct answers, one gets 9 correct, one gets 8 correct, and so on, all the way down to one student getting none correct. What is the mode score, accurate to three decimal places?
- 4.545
  - 5.000
  - 5.500
  - It is not defined.
25. The largest cumulative absolute frequency in a set of numbers is equal to
- The sum of all the individual absolute frequency values
  - Twice the mean
  - Twice the median
  - Twice the mode
26. Which of the following is an expression of the extent to which values are spread out relative to the mean?
- The average
  - The mode
  - The median
  - None of the above
27. When a distribution is skew to the right:
- The median is greater than the mean
  - The distribution is always unimodal
  - The tail on the left is shorter than the tail on the right
  - The standard deviation is always less than the variance
- e) The majority of observations are less than the mean.
28. The difference between first and third quartile is called
- Range
  - Inter Quartile range
  - Median
  - Mode
29. The purpose of inter quartile range is to Measure the
- Spread around the mean
  - Spread around the median
  - Spread around the mode
  - Spread around the harmonic mean
30. Kurtosis determined the distribution
- Flatness
  - Peak ness
  - Both a) and b)
  - None of the above
31. Which of the following represents median?
- First Quartile
  - Fiftieth Percentile
  - Sixth decile
  - Third quartile
32. If the grouped data has open-end classes, one can not calculate
- Median
  - Mode
  - Mean
  - Quartile
33. In a symmetric distribution
- Mean = median = mode
  - Mean > median > mode
  - Mean < median < mode
  - None of these
34. The following list is ordered from smallest to largest: 25, 26, 26, 30, y, y, 33, 150. Which of the following statements is (are) true?
- The mean is greater than the median
  - The mode is 26
  - There are no outliers in the data
- I only
  - I and II only
  - III only

- d) I and III only  
e) II and III only
35. The mean and standard deviation of a normally distributed dataset are 19 and 4, respectively. 19 is subtracted from every term in the dataset and then the result is divided by 4. Which of the following best describes the resulting distribution?  
 a) It has a mean of 0 and a standard deviation of 1.  
 b) It has a mean of 0, a standard deviation of 4, and its shape is normal.  
 c) It has a mean of 1 and a standard deviation of 0.  
 d) It has a mean of 0, a standard deviation of 1, and its shape is normal.  
 e) It has a mean of 0, a standard deviation of 4, and its shape is unknown.
36. A tabulation of cumulative absolute frequency values is useful in determining  
 a) The mode  
 b) The variance  
 c) The median  
 d) The mean
37. If modal value is not clear in a distribution; it can be ascertained by the method of  
 a) Grouping  
 b) Guessing  
 c) Summarizing  
 d) Trial and error
38. The middle value of an ordered series is called:  
 a) 2nd quartile  
 b) 5th decile  
 c) 50th percentile  
 d) All the above
39. The variate values which divide a series (frequency distribution) into ten equal parts are called:  
 a) Quartiles  
 b) Deciles  
 c) Octiles  
 d) Percentiles
40. For percentiles, the total numbers of partition values are  
 a) 10  
 b) 59  
 c) 100  
 d) 99
41. The first quartile divides a frequency distribution in the ratio  
 a) 4 : 1  
 b) 1 : 4  
 c) 3 : 1  
 d) 1 : 3
42. Sum of the deviations about mean is  
 a) Zero  
 b) Minimum  
 c) Maximum  
 d) One
43. Histogram is useful to determine graphically the value of  
 a) Mean  
 b) Median  
 c) Mode  
 d) All the above
44. Median can be located graphically with the help of  
 a) Histogram  
 b) Ogives  
 c) Bar diagram  
 d) Scatter diagram
45. Sixth deciles is same as  
 a) Median  
 b) 50th percentile  
 c) 60th percentile  
 d) First quartile
46. Which of the following statements is (are) true?  
 I. The median is resistant to extreme values.  
 II. The mean is resistant to extreme values.  
 III. The standard deviation is resistant to extreme values.  
 a) I only  
 b) II only  
 c) III only  
 d) II and III only  
 e) I and III only

- 47 One of the values in a normal distribution is 43 and its z-score is 1.65. If the mean of the distribution is 40, what is the standard deviation of the distribution?
- 3
  - 1.82
  - 0.55
  - 1.82
  - 0.55
- 48 What percentage of values lies between 5<sup>th</sup> and 25<sup>th</sup> percentiles?
- 5%
  - 20%
  - 30%
  - 75%
- 49 Which of the following statements are true?
- Generally speaking, using a trimmed mean with a moderate trimming proportion (between 5% and 25%) will yield a measure that is neither as sensitive to outliers as the mean, nor as insensitive as the median.
  - When the data is categorical, a frequency distribution or relative frequency distribution provides an effective tabular summary of the data.
  - When the  $n$  observations in a sample are ordered from smallest to largest, and  $n$  is odd, then the sample median is the  $\left(\frac{n+1}{2}\right)^{th}$  ordered value.
  - None of the above is true.
  - All of the above are true.
- 50 Which of the following is a unitless measure of dispersion?
- Standard deviation
  - Mean deviation
  - Coefficient of variation
  - Range
- 51 Absolute sum of deviations is minimum from
- Mode
  - Median
- 52 In a distribution  $S.D = 6$ . All observation multiplied by 2 would give the result to S.D is
- 12
  - 6
  - 18
  - 6
- 53 The mean of squared deviations about the mean is called
- S.D
  - Variance
  - M.D
  - None
- 54 If the minimum value in a set is 9 and its range is 57, the maximum value of the set is
- 33
  - 66
  - 48
  - 24
- 55 Quartile deviation is equal to
- Inter quartile range
  - Double the inter quartile range
  - Half of the inter quartile range
  - None of the above
- 56 Which of the following measures is most affected by extreme values?
- S.D
  - Q.D
  - M.D
  - Range
- 57 Which measure of dispersion ensures highest degree of reliability?
- Range
  - Mean deviation
  - Q.D
  - S.D
- 58 For a negatively skewed distribution, the correct inequality is
- Mode < median
  - Mean < median
  - Mean < mode
  - None of the above

- 59 In case of positive skewed distribution, the extreme values lie in the

  - Left tail
  - Right tail
  - Middle
  - Any where

60 The square root of the variance as a measure of dispersion is always the

  - Half variance
  - Median
  - Z-score
  - Standard deviation

61 Suppose a large number of people take a test, and every single student gets exactly half of the answers right. In this case, the standard deviation is

  - Equal to the mean
  - Equal to the median
  - Equal to zero
  - Impossible to determine without more information

62 Which of the following is an expression of the extent to which values are spread out relative to the mean?

  - The average
  - The mode
  - The median
  - None of the above

63 Suppose a large number of people take a test. The 3rd decile point is determined by

  - Finding the highest score representing the "worst" 20% or fewer papers; the 3rd decile is at the top of that set
  - Finding the highest score representing the "worst" 30% or fewer papers; the 3rd decile is at the top of that set
  - Finding the lowest score representing the "best" 20% or fewer papers; the 3rd decile is at the bottom of that set
  - Finding the lowest score representing the "best" 30% or fewer papers; the 3rd decile is at the bottom of that set

64 Suppose many students take a 10-question quiz, and the mean turns out to be 7.22 answers correct. Suppose the standard deviation turns out to be 0.722. What is the coefficient of variation?

  - We can't answer this unless we know how many people take the quiz
  - 0.1
  - 10
  - 100

65 Imagine that you take a standardized test, and after you've finished you are told that you are at the 91st percentile. This means

  - 90 of the students taking the same test have scores higher than yours
  - 90% of all the students taking the same test have scores higher than yours
  - 90 of the students taking the same test have scores lower than yours
  - nothing, there is no such thing as the 91st percentile

66 If you take a standardized test and then you are told you are in the 50th percentile, this means

  - That your score is among the lowest in the range
  - That your score is among the best in the range
  - That your score is near the middle of the range
  - That your score has low correlation
  - Nothing in particular about your score

67 How many decile points are there in a set of 100,000 ranked data elements?

  - More information is necessary in order to answer this question.
  - 9.
  - 10
  - 99
  - 100

68 The average of the outcomes in an experiment is known as the

  - Continuous variable
  - Discrete variable
  - Random variable
  - Mean
  - Frequency

69. Which of the following pairs of characteristics are both measures of the extent to which the data in a distribution is spread out?
- The mean and the median.
  - The mean and the deviation.
  - The variance and the standard deviation.
  - The mode and the mean.
  - None of the above.
70. If the number of elements in a distribution is even, then the \_\_\_\_\_ is the value such that half the elements have values greater than or equal to it, and half the elements have values less than or equal to it."
- Mean
  - Average
  - Standard deviation
  - Median.
  - Variance
71. In a normal distribution, an element in the 35th percentile lies within
- The 1st quartile
  - The 2nd quartile.
  - The 3rd quartile
  - The 4th quartile
  - The middle quartile
72. In a ranked data set, the value of the 3rd quartile point minus the value of the 1st quartile point is called
- The interquartile range.
  - The standard deviation
  - The variance
  - The coefficient of variation
  - The z score
73. If  $X$  and  $Y$  have the variances 25 and 9 then  $SD(X-Y)$  is: (Subject GRE, 2010)
- 16
  - 34
  - 7
  - 0
74. Estimate the standard deviation of the population from which the following set of sample observations was drawn:  
26, 31, 31, 27, 24, 29, 29
- 6.81
  - 28.14
  - 2.61
  - 2.42
75. Estimate the variance of the population which produced the following sample data: 26, 31, 31, 27, 24, 29, 29
- 6.81
  - 5.84
  - 2.61
  - 2.42
76. The median is a better choice than the mean for a measure of central tendency if the data:
- Are bimodal.
  - Often have outliers.
  - Are normally distributed.
  - Are exponentially distributed.
77. Calculate the estimated variance of the population from which the following values have been randomly selected: 2.8, 2.7, 2.6, 2.9, 2.8, 2.8, 2.8.
- 0.095
  - 0.009
  - 0.088
  - 0.008
78. An advantage of using standard deviation rather than range for measuring dispersion of a large sample is that:
- Standard deviation has a simpler formula.
  - Calculators have a standard deviation key but not a range key.
  - Standard deviation uses information from each measurement.
  - Range calculations are not normally distributed.
79. If in a categorical data situation we focus attention on a particular category and code the sample results so that a "1" is recorded for an individual not in the category, then the sample proportion of individuals in the category is:
- larger than the sample mean of the sequence of 1's and 0's
  - equal to the sample mean of the sequence of 1's and 0's
  - smaller than the sample mean of the sequence of 1's and 0's
  - the number of 0's divided by the number of 1's

80. Which of the following is not a measure of variability?  
 a) Variance  
 b) Standard deviation  
 c) Mean absolute deviation  
 d) Median  
 e) All of the above
81. What is the median of the following set of scores? 18, 6, 12, 10, 14 ?  
 a) 10  
 b) 14  
 c) 18  
 d) 12
82. Approximately what percentage of scores fall within one standard deviation of the mean in a normal distribution?  
 a) 34%  
 b) 95%  
 c) 99%  
 d) 68%
83. The standard deviation is:  
 a) The square root of the variance  
 b) A measure of variability  
 c) An approximate indicator of how numbers vary from the mean  
 d) All of the above
84. The most frequently occurring number in a set of values is called the.  
 a) Mean  
 b) Median  
 c) Mode  
 d) Range
85. As a general rule, the \_\_\_\_\_ is the best measure of central tendency because it is more precise.  
 a) Mean  
 b) Median  
 c) Mode  
 d) Range
86. Focusing on describing or explaining data versus going beyond immediate data and making inferences is the difference between.  
 a) Central tendency and common tendency  
 b) Mutually exclusive and mutually exhaustive properties
87. Why are variance and standard deviation the most popular measures of variability?  
 a) They are the most stable and are foundations for more advanced statistical analysis  
 b) They are the most simple to calculate with large data sets  
 c) They provide nominally scaled data  
 d) None of the above
88. Which of the following statements are true?  
 a) The unit for the standard deviation is the same as the unit for the data values.  
 b) The average deviation of  $x_1, x_2, \dots, x_n$  from the mean  $\bar{x}$  may be positive or negative.  
 c) The average absolute deviation of  $x_1, x_2, \dots, x_n$  from the mean  $\bar{x}$  is always zero.  
 d) We use  $\sigma^2$  (square of the lowercase Greek letter sigma) to denote the sample variance.
89. Which of the following is used as a divisor in the sample variance  $s^2$ , where  $n$  is the sample size?  
 a)  $n+1$   
 b)  $n$   
 c)  $n-1$   
 d)  $n-2$
90. Given that  $n = 10$ ,  
 $\sum_{i=1}^{10} x_i = 25$ ,  $\sum_{i=1}^{10} x_i^2 = 512$ , then the sample standard deviation is  
 a) 7.012  
 b) 6.704  
 c) 7.067  
 d) None of these
91. Which of the following statements are correct if  $y_i = x_i + 5, \forall i = 1, 2, 3, \dots, n$ ?  
 a) The mean of the new data set is 5 more than the mean of the original data set.  
 b) The standard deviation of the new data set is 5 more than the standard deviation of the original data set.  
 c) The range of the new data set is 5 more than the range of the original data set.  
 d) The median of the new data set is 5 more than the median of the original data set.

- |      |  |                |
|------|--|----------------|
|      |  | dot plots.     |
| a)   | $\bar{y} = \bar{x}$  | b) 2           |
| b)   | $\bar{y} = \bar{x} + 5$  | c) 3           |
| c)   | $s_y^2 = s_x^2 + 5$  | d) 4           |
| d)   | $s_y^2 = s_x^2 + 25$   | 5              |
| e)   | None of the above  |                |
| 92.  | Which of the following statements are correct if $y_i = 2x_i$ , $\forall i = 1, 2, 3, \dots, n$ ?                              |                |
| a)   | $\bar{y} = 2\bar{x}$   | a) Mean        |
| b)   | $\bar{y} = 4\bar{x}$   | b) Median      |
| c)   | $s_y^2 = 2s_x^2$   | c) Mode        |
| d)   | $s_y^2 = s_x^2 + 4$  | d) Correlation |
| e)   | None of the above  | e) Modal class |
| 93   | How many outliers are there in this data set? 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024                                      |                |
| a)   | 0  | a) Left tail   |
| b)   | 1  | b) Right tail  |
| c)   | 2  | c) Middle      |
| d)   | 3  | d) Any where   |
| e)   | 4  |                |
| 94.  | Which of the following would be greatly affected by outliers?  |                |
| I.   | IQR  |                |
| II.  | Range  |                |
| III. | median   |                |
| a)   | I only   | a) 1           |
| b)   | II only  | b) 31          |
| c)   | I and II   | c) 61          |
| d)   | I and III  | d) 91          |
| e)   | II and III   | e) 92          |
| 95.  | If two distributions are symmetric and have the same range and the same mean, then which of the following statements are true? |                |
| I.   | They have the same standard deviation.   | a) 93          |
| II.  | They have the same median.   | b) 32          |
| III. | They have the same IQR.  | c) 62          |
| a)   | I only   | d) 95          |
| b)   | II only  | e) 96          |
| c)   | III only   | f) 97          |
| d)   | I and III  | g) 98          |
| e)   | I, II, and III   | h) 99          |
| 96.  | Two distributions have identical box plots. How many of the following may be true about the two distributions?                 |                |
| I.   | They have the same range.  | i) 40          |
| II.  | They have the same IQR.  | j) 41          |
| III. | They have the same mean.   | k) 42          |
| IV.  | They have the same median.   | l) 43          |
|      |  | m) 44          |
|      |  | n) 45          |
|      |  | o) 46          |
|      |  | p) 47          |
|      |  | q) 48          |
|      |  | r) 49          |
|      |  | s) 50          |
|      |  | t) 51          |
|      |  | u) 52          |
|      |  | v) 53          |
|      |  | w) 54          |
|      |  | x) 55          |
|      |  | y) 56          |
|      |  | z) 57          |
|      |  | aa) 58         |
|      |  | ab) 59         |
|      |  | ac) 60         |

## ANSWERS

1	d	31	b	61	c	91	b
2	b	32	c	62	d	92	a
3	d	33	a	63	b	93	b
4	b	34	a	64	b	94	b
5	b	35	d	65	d	95	b
6	e	36	c	66	c	96	e
7	d	37	d	67	b	98	a
8	b	38	d	68	d		
9	a	39	b	69	c		
10	d	40	c	70	d		
11	c	41	d	71	b		
12	a	42	a	72	a		
13	c	43	c	73	c		
14	d	44	b	74	c		
15	e	45	c	75	a		
16	b	46	a	76	b		
17	b	47	d	77	b		
18	d	48	b	78	c		
19	c	49	e	79	b		
20	a	50	c	80	d		
21	c	51	c	81	d		
22	b	52	a	82	d		
23	b	53	b	83	d		
24	d	54	b	84	c		
25	a	55	b	85	a		
26	d	56	d	86	c		
27	a	57	d	87	a		
28	b	58	c	88	a		
29	b	59	b	89	c		
30	b	60	d	90	c		

## PROBABILITY AND PROBABILITY DISTRIBUTIONS

---

1. Experiments that results in unique outcomes are known as
- Deterministic experiment
  - Random experiment
  - Both a) & b)
  - None of these
2. Experiments that do not results in unique outcomes are known as
- deterministic experiment
  - Random experiment
  - Both a) & b)
  - None of these
3. Which of the following is not a type of probability?
- Classical
  - Empirical
  - Subjective
  - Finite
4. Which of the following is true for events  $M$  and  $N$ ?
- $(M \cup N)' = M' \cap N'$
  - $(M \cap N)' = M' \cup N'$
  - Neither a) nor b) is true
  - Both a) and b) are true
5. Rolling a die or tossing a coin is called a
- Sample experiment
  - Probability experiment
  - Infinite experiment
  - Repeated experiment
6. Two outcomes are mutually exclusive if and only if
- They are non-disjoint
  - They have no elements in common
  - They have at least one element in common
  - They have identical sets of outcomes
7. The intersection of two disjoint sets contains
- No elements
  - One element
  - All the elements of one set
  - All the elements of both sets
  - An infinite number of elements
8. The probability, expressed as a percentage, of a particular occurrence can never be
- less than 100
  - Less than 0
  - greater than 1
  - anything but a whole number
9. The difference between permutations and combinations lies in the fact that
- Permutations take order into account, but combinations do not
  - Combinations take order into account, but permutations do not
  - Combinations involve only continuous variables, but permutations involve only discrete variables
  - Permutations involve only continuous variables, but combinations involve only discrete variables
10. Which of the following statements is true?
- The number of combinations of size  $k$  from a particular set is larger than the number of permutations.
  - The number of combinations of size  $k$  from a particular set is smaller than the number of permutations.
  - The number of combinations of size  $k$  from a particular set is equal to the number of permutations.
  - None of the above

11. A homeowner doing some remodeling requires the services of both a plumbing contractor and an electrical contractor. If there are 15 plumbing contractors, and 10 electrical contractors available in the area, in how many ways can the contractors be chosen?
- a) 15                          b) 10  
c) 25                          d) 150
12. How many permutations of size 3 can be constructed from the set (A, B, C, D, E)?
- a) 60                          b) 20  
c) 15                          d) 8
13. How many combinations of size 4 can be formed from a set of 6 distinct objects?
- a) 36                          b) 24  
c) 15                          d) 10
14. An experiment consists of three stages. There are five ways to accomplish the first stage, four ways to accomplish the second stage, and three ways to accomplish the third stage. The total number of ways to accomplish the experiment is
- a) 60                          b) 20  
c) 15                          d) 12
15. What is  $0!$ ?
- a) 0                                  b) 1  
c) Undefined                          d) Infinite
16. In how many ways can the letters in the word UNIVERSITY can be arranged randomly
- a) 9!                                  b) 10!  
c)  $\binom{10}{2}$                                   d)  $\frac{10!}{2!}$   
e) None of the above
17. Seventeen teams can take a part in the Football championship of a country. In how many ways can the Gold, Silver and bronze medals be distributed among the teams
- a)  $17^3$                                   b)  $\frac{17!}{3!14!}$
- c)  $\frac{17!}{14!}$                                   d)  $14!$
18. The number of 3-digit telephone area codes that can be made if repetitions are not allowed is
- a) 100                                  b) 720  
c) 1000                                  d) 604
19. In how many different ways can a person select one book from 3 novels, one book from 6 biographies and one book from 7 self help books?
- a) 15                                  b) 105  
c) 3    d) 22
20. In how many ways can 7 different calculators be displayed in a row on a shelf?
- a) 7    b) 49  
c) 823,543                                  d) 5,040
21. If a board of directors consists of 10 people, in how many ways can a chief executive officer, a director, a treasurer, and a secretary be selected?
- a) 5040                                  b) 210  
c) 40    d) 14
22. How many different eggs signals can be made from 3 red eggs, 2 green eggs, and 2 white eggs?
- a) 49    b) 84  
c) 210    d) 320
23. In how many ways can 3 boxes of cereal be selected for testing from 12 boxes?
- a) 36    b) 220  
c) 480    d) 1320
24. In how many ways can a jury of 5 men and 7 women be selected from 10 men and 10 women?
- a) 120    b) 252  
c) 372    d) 30,240
25. What is the number of possible permutations of 7 objects taken 3 at a time?
- a) 10    b) 21  
c) 35    d) 210

26. Given six books, how many sets can be arranged in lots of three—but always in a different order?  
 a) 18                    b) 54  
 c) 108                d) 120
27. The result of an event is called  
 a) An experiment  
 b) A trial  
 c) An outcome  
 d) A variable
28. A set of all possible outcomes of a random experiment is called  
 a) Event  
 b) Compound event  
 c) Sample space  
 d) None of these
29. A sample space together with a probability measure provides a mathematical model for an experiment. This model is often called  
 a) A probability space  
 b) Event space  
 c) Random experiment  
 d) None of the above
30. A Borel sigma field is generated by all  
 a) Close intervals  
 b) Open intervals  
 c) Close-open intervals  
 d) Open-close intervals
31. When an event cannot occur, its probability is  
 a) 1                    b) 0  
 c)  $\frac{1}{2}$                 d) 0.01
32. The type of probability that uses sample spaces is called  
 a) Classical probability  
 b) Empirical probability  
 c) Subjective probability  
 d) Relative probability
33. The probability of an event can be any number from \_\_\_\_\_ to \_\_\_\_\_.  
 a) -1, 1              b) 0, 1  
 c) 1, 100            d) 0, infinity
34. When two events cannot occur at the same time, they are said to be events  
 a) Independent  
 b) Mutually exclusive  
 c) Random  
 d) Inconsistent
35. If the probability that an event will happen is 0.48, then the probability that the event will not happen is  
 a) 0.52              b) -0.48  
 c) 0                    d) 1
36. When two dice are rolled, the sample space consists of \_\_\_\_\_ outcomes  
 a) 6                    b) 12  
 c) 18                  d) 36
37. When an event is certain to occur, its probability is  
 a) 0                    b) 1  
 c)  $\frac{1}{2}$                 d) -1
38. How many outcomes are there in the sample space, when two coins are tossed?  
 a) 1                    b) 2  
 c) 3                    d) 4
39. When a coin is tossed and then a die is rolled, the probability of getting a tail on the coin and an odd number on the die is  
 a)  $\frac{1}{2}$                 b)  $\frac{1}{4}$   
 c)  $\frac{3}{4}$                 d)  $\frac{1}{12}$
40. When a coin is tossed and a die is rolled, the probability of getting a head and a number less than 5 on the die is  
 a)  $\frac{1}{3}$                 b)  $\frac{2}{3}$   
 c)  $\frac{1}{2}$                 d)  $\frac{5}{6}$

41. When three coins are tossed, the probability of getting at least one tail is
- $\frac{3}{8}$
  - $\frac{1}{8}$
  - $\frac{7}{8}$
  - $\frac{5}{8}$
42. When three coins are tossed, the probability of getting two or more heads is
- $\frac{3}{8}$
  - $\frac{1}{8}$
  - $\frac{1}{2}$
  - $\frac{7}{8}$
43. A box contains a penny, a nickel, a dime, and a quarter. If two coins are selected without replacement, the probability of getting an amount greater than 11¢ is
- $\frac{5}{72}$
  - $\frac{2}{3}$
  - $\frac{3}{4}$
  - $\frac{5}{6}$
44. A bag contains a red bead, a green bead, and a blue bead. If a bead is selected and its color noted, and then it is replaced and another bead is selected, the probability that both beads will be of the same color is
- $\frac{1}{8}$
  - $\frac{3}{4}$
  - $\frac{1}{16}$
  - $\frac{1}{3}$
45. A card is selected at random from an ordinary deck of 52 cards. The probability that the 7 of diamonds is selected is
- $\frac{1}{13}$
  - $\frac{1}{4}$
  - $\frac{1}{52}$
  - $\frac{1}{26}$
46. A card is selected at random from a deck of 52 cards. The probability that it is a 7 is
- $\frac{1}{4}$
  - $\frac{1}{52}$
  - $\frac{7}{52}$
  - $\frac{1}{13}$
47. A card is drawn from an ordinary deck of 52 cards. The probability that it is a spade is
- $\frac{1}{4}$
  - $\frac{1}{13}$
  - $\frac{1}{52}$
  - $\frac{1}{26}$
48. A card is drawn from an ordinary deck of 52 cards. The probability that it is a 9 or a club is
- $\frac{17}{52}$
  - $\frac{5}{8}$
  - $\frac{4}{13}$
  - $\frac{3}{4}$
49. A card is drawn from an ordinary deck of 52 cards. The probability that it is a face card is
- $\frac{3}{52}$
  - $\frac{1}{4}$
  - $\frac{9}{13}$
  - $\frac{3}{13}$
50. Two dice are rolled. The probability that the sum of the spots on the faces will be nine is
- $\frac{1}{9}$
  - $\frac{5}{36}$
  - $\frac{1}{6}$
  - $\frac{3}{13}$
51. Two dice are rolled. The probability that the sum of the spots on the faces is greater than seven is
- $\frac{2}{3}$
  - $\frac{7}{36}$

- c)  $\frac{3}{4}$       d)  $\frac{5}{12}$
52. Two dice are rolled. The probability that one or both numbers on the faces will be 4 is
- a)  $\frac{1}{3}$       b)  $\frac{4}{13}$   
c)  $\frac{11}{36}$       d)  $\frac{1}{6}$
53. Two dice are rolled. The probability that the sum of the spots on the faces will be even is
- a)  $\frac{3}{4}$       b)  $\frac{5}{6}$   
c)  $\frac{1}{2}$       d)  $\frac{1}{6}$
54. When two coins are tossed, the sample space is
- a) H, T  
b) HH, HT, TT  
c) HH, HT, TH, TT  
d) H, T and HT
55. When a die is rolled, the probability of getting a number greater than 4 is
- a)  $\frac{1}{6}$       b)  $\frac{1}{3}$   
c)  $\frac{1}{2}$       d) 1
56. When two coins are tossed, the probability of getting 2 tails is
- a)  $\frac{1}{2}$       b)  $\frac{1}{3}$   
c)  $\frac{1}{4}$       d)  $\frac{1}{8}$
57. If a letter is selected at random from the word "Mississippi," and the probability that it is an "s" is
- a)  $\frac{1}{8}$       b)  $\frac{1}{2}$   
c)  $\frac{3}{11}$       d)  $\frac{4}{11}$
58. When a die is rolled, the probability of getting an 8 is
- a)  $\frac{1}{6}$       b) 0  
c) 1      d)  $\frac{1}{2}$
59. In a survey of 180 people, 74 were over the age of 64. If a person is selected at random, what is the probability that the person is over 64?
- a)  $\frac{16}{45}$       b)  $\frac{32}{37}$   
c)  $\frac{37}{90}$       d)  $\frac{53}{90}$
60. In a classroom of 24 students, there were 20 freshmen. If a student is selected at random, what is the probability that the student is not a freshman?
- a)  $\frac{2}{3}$       b)  $\frac{5}{6}$   
c)  $\frac{1}{3}$       d)  $\frac{1}{6}$
61. What is the mathematical probability that a coin, tossed 10 times in a row, will come up "tails" on all 10 tosses?
- a) 1      b) 1/10  
c) 1/1024      d) 1/4096
62. A random variable that assume a finite or countably infinite number of values is called
- a) A discrete random variable  
b) A continuous random variable  
c) Both a & b  
d) None of the above
63. Let  $X$  be a discrete random variable with  $V(X) = 3.70$ , then  $V(2X)$  is
- a) 13.69      b) 5.70  
c) 14.80      d) 7.40
64. Let  $X$  be a discrete random variable with  $V(X) = 8.6$ , then  $V(3X + 5.6)$  is
- a) 77.4      b) 14.2  
c) 83.0      d) 31.4

35. A random variable that assume an uncountably infinite number of values is called  
 a) A discrete random variable  
 b) A continuous random variable  
 c) Both a) & b)  
 d) None of the above
36. Something that might or might not happen depending on outcome of the experiment is called  
 a) Trial  
 b) An event  
 c) Sample space  
 d) Compound event  
 e) None of the above
37. What is the mathematical probability that a coin, tossed 13 times in a row, will come up "heads" on all 13 tosses?  
 a) 1 in 512  
 b) 1 in 1024  
 c) 1 in 2048  
 d) 1 in 4096  
 e) 1 in 8192.
38. A member of a set is also called  
 a) A dependent variable of the set  
 b) An independent variable of the set  
 c) A random variable of the set  
 d) A discrete variable of the set  
 e) None of the above.
39. A phone extension consists of 3 digits. If all the digits have a probability of being selected, what is the probability that the extension consists of the digits 1, 2, and 3 in any order? Repetitions are allowed.  
 a) 0.555  
 b) 0.006  
 c) 0.233  
 d) 0.125
40. Three cards are selected at random, what is the probability that they will all be clubs?  
 a) 0.002  
 b) 0.034  
 c) 0.013  
 d) 0.127
41. At a used book sale, there are 6 novels and 4 biographies. If a person selects 4 books at random, what is the probability that the person selects two novels and two biographies?
72. To win a lottery, a person must select 4 numbers in any order from 20 numbers. Repetitions are not allowed. What is the probability that the person wins?  
 a) 0.0002  
 b) 0.0034  
 c) 0.0018  
 d) 0.0015
73. Empirical probability is based on  
 a) Observation or experimentation  
 b) Theoretical models only  
 c) Continuous outcomes  
 d) Standard deviations
74. What is the mathematical probability that an "unweighted" die, tossed four times, will show the face with 6 dots on all four occasions?  
 a) 1 in 6  
 b) 1 in 36  
 c) 1 in 64  
 d) 1 in 1296.  
 e) 1 in 46,656
75. Let  $N(A)$  be the number of elements in set A then  $N(A \cup B)$  is  
 a)  $N(A) + N(B)$   
 b)  $N(A) + N(B) - N(A \cap B)$   
 c)  $N(A) - N(A \cap B)$   
 d)  $N(A) + N(B) + N(A \cap B)$
76. If  $P(A \cap B) = \phi$ , then  $P(A \cup B) =$   
 a) 0  
 b)  $P(A) + P(B) - P(A \cap B)$   
 c)  $P(A) + P(B)$   
 d) None of the above
77. Probability of sample space is equal to  
 a) 1  
 b) 0  
 c) 0.5  
 d) 0.75
78. If  $P(A/B) = P(A)$  then  
 a) A and B are dependent  
 b) A and B are independent  
 c) None of the above

79. Two outcomes are independent if and only if  
 a) They always occur simultaneously  
 b) One occurs only when the other does not  
 c) They sometimes occur simultaneously, but usually they do not  
 d) The occurrence of one does not affect the probability that the other will occur.  
 e) They rarely occur simultaneously, but once in a while they do
80. If A & B are two independent events then  
 a)  $P(A \setminus B) = P(A)$   
 b)  $P(A \cap B) = P(A)P(B)$   
 c)  $P(B \setminus A) = P(B)$   
 d) All of the above
81. If A & B are two independent events then  
 a)  $\bar{A}$  and  $B$  are independent  
 b)  $A$  and  $\bar{B}$  are independent  
 c)  $\bar{A}$  and  $\bar{B}$  are independent  
 d) All of the above
82. If  $P(A) \geq 0.8$  and  $P(B) \geq 0.8$  then which of the following is true?  
 a)  $P(A \cap B) \geq 0.6$   
 b)  $P(A \cap B) \geq 0.64$   
 c)  $P(A \cap B) \geq 0.8$   
 d)  $P(A \cap B) < 0.64$   
 e) None of the above
83. Which of the two events are not mutually exclusive?  
 a) Rolling a die and getting a 6 or a 3  
 b) Drawing a card from a deck and getting a club or an ace  
 c) Tossing a coin and getting a head or a tail  
 d) Tossing a coin and getting a head and rolling a die and getting an odd number
84. Which of the two events are mutually exclusive?  
 a) Drawing a card from a deck and getting a king or a club  
 b) Rolling a die and getting an even number or a 6  
 c) Tossing two coins and getting two heads or two tails  
 d) Rolling two dice and getting doubles or getting a sum of eight
85. Which of the following statements is always true?  
 a) The complement of an event A, denoted by  $A'$ , is the set of all outcomes in the sample space that are not contained in A.  
 b) The union of two events A and B, denoted by  $A \cup B$ , is the event consisting of all outcomes that are in both events.  
 c) The intersection of two events A and B, denoted by  $A \cap B$ , is the event consisting of all outcomes that are either in A or in B.  
 d) All of the above
86. Which of the following statements is correct?  
 a) The term probability refers to the study of certainty.  
 b) With sample space  $S = \{1, 2, 3, 4, 5, 6\}$ , and event  $A = \{3, 4, 5, 6\}$ , then  $A' = \{1, 2, 3, 4\}$   
 When events A and B have no outcomes in common, they are said to be mutually exclusive or disjoint events.  
 c) None of the above
87. Possessing a boat and renting an apartment are:  
 a) Independent events  
 b) Mutually exclusive  
 c) Both independent and mutually exclusive  
 d) Neither independent nor mutually exclusive

88. If  $A$  and  $B$  are disjoint events with  $P(A) > 0$  and  $P(B) > 0$ , then
- $P(A \cup B) = P(A) + P(B)$
  - $P(A \cup B) = P(A)P(B)$
  - $P(A \cap B) = P(A) + P(B)$
  - $P(A \cap B) = P(A)P(B)$
89. For events  $A$  and  $B$  with  $P(B) > 0$ , which of the following is always true?
- $P(A/B) \cdot P(B/A) = P(A \cap B)$
  - $P(A/B) = P(B/A)$
  - $P(A/B) + P(B/A) = 1$
  - $P(A/B) + P(A'/B) = 1$
90. An experiment consists of three stages. There are five ways to accomplish the first stage, four ways to accomplish the second stage, and three ways to accomplish the third stage. The total number of ways to accomplish the experiment is
- 60
  - 20
  - 15
  - 12
91. In many counting and probability problems, a useful configuration that can be used to represent pictorially all the possibilities (sample space and probabilities) is the
- Product rule for ordered pairs
  - Histogram
  - Venn diagram
  - Tree diagram
92. If  $P(A) = 0.30$ ,  $P(B) = 0.60$ , and  $P(A \cap B) = 0.18$ , then events  $A$  and  $B$  are
- Dependent
  - Independent
  - Mutually exclusive
  - Complementary
93. In a box there are 6 white marbles, 3 blue marbles, and 1 red marble. If a marble is selected at random, what is the probability that it is red or blue?
94. When a single die is rolled, what is the probability of getting a prime number (2, 3, or 5)?
- $\frac{5}{6}$
  - $\frac{2}{3}$
  - $\frac{1}{2}$
  - $\frac{1}{6}$
95. A storeowner plans to have his annual "Going Out of Business Sale." If each month has an equal chance of being selected, and the probability that the sale will be in a month that begins with the letter J or M,
- $\frac{1}{4}$
  - $\frac{1}{6}$
  - $\frac{5}{8}$
  - $\frac{5}{12}$
96. A card is selected from a deck of 52 cards. Find the probability that it is a red queen or a black ace.
- $\frac{2}{13}$
  - $\frac{1}{13}$
  - $\frac{5}{13}$
  - $\frac{8}{13}$
97. At a high school with 300 students, 62 play football, 33 play baseball, and 14 play both sports. If a student is selected at random, and the probability that the student plays football or baseball,
- $\frac{27}{100}$
  - $\frac{109}{300}$
  - $\frac{19}{60}$
  - $\frac{14}{300}$

98. A card is selected from a deck. Find the probability that it is a face card or a diamond.

- a)  $\frac{25}{52}$       b)  $\frac{3}{52}$   
 c)  $\frac{11}{26}$       d)  $\frac{13}{52}$

99. A single card is selected from a deck. Find the probability that it is a queen or a black card.

- a)  $\frac{11}{26}$       b)  $\frac{7}{13}$   
 c)  $\frac{1}{13}$       d)  $\frac{15}{26}$

100. Two dice are rolled. What is the probability of getting doubles or a sum of 10?

- a)  $\frac{11}{18}$       b)  $\frac{2}{9}$   
 c)  $\frac{1}{4}$       d)  $\frac{11}{36}$

101. The probability that a family visits Safari Zoo is 0.65, and the probability that a family rides on the Mt. Pleasant Tourist Railroad is 0.55. The probability that a family does both is 0.43. Find the probability that the family visits the zoo or the railroad.

- a) 0.77      b) 0.22  
 c) 0.12      d) 0.10

102. If a card is drawn from a deck, what is the probability that it is a king, queen, or an ace?

- a)  $\frac{5}{13}$       b)  $\frac{7}{13}$   
 c)  $\frac{6}{13}$       d)  $\frac{3}{13}$

103. Which of the following events are dependent?

- a) Tossing a coin and rolling a die

- b) Rolling a die and then rolling a second die  
 c) Sitting in the sun all day and getting sunburned  
 d) Drawing a card from a deck and rolling a die

104. Three dice are rolled. What is the probability of getting three 4s?

- a)  $\frac{1}{216}$       b)  $\frac{1}{6}$   
 c)  $\frac{1}{36}$       d)  $\frac{1}{18}$

105. What is the probability of selecting 4 spades from a deck of 52 cards if each card is replaced before the next one is selected?

- a)  $\frac{4}{13}$       b)  $\frac{1}{2561}$   
 c)  $\frac{1}{52}$       d)  $\frac{1}{13}$

106. A coin is tossed four times; what is the probability of getting 4 heads?

- a)  $\frac{1}{2}$       b)  $\frac{1}{16}$   
 c) 1      d)  $\frac{1}{4}$

107. If 25% of U.S. prison inmates are not U.S. citizens, what is the probability of randomly selecting three inmates who will not be U.S. citizens?

- a) 0.75      b) 0.421875  
 c) 0.015625      d) 0.225

108. If three people are randomly selected, and the probability that they will all have birthdays in June.

- a)  $\frac{1}{1728}$       b)  $\frac{1}{36}$   
 c)  $\frac{1}{12}$       d)  $\frac{1}{4}$

109 In a sample of 10 telephones, 4 are defective. If 3 are selected at random and tested, what is the probability that all will be non defective?

- a)  $\frac{1}{30}$
- b)  $\frac{8}{125}$
- c)  $\frac{1}{6}$
- d)  $\frac{27}{125}$

110. A bag contains 4 blue marbles and 5 red marbles. If 2 marbles are selected at random without replacement, what is the probability that both will be blue?

- a)  $\frac{16}{81}$
- b)  $\frac{1}{6}$
- c)  $\frac{1}{2}$
- d)  $\frac{1}{15}$

111. If two people are selected at random from the phone book of a large city, and the probability that they were both born on a Sunday.

- a)  $\frac{1}{7}$
- b)  $\frac{1}{49}$
- c)  $\frac{2}{365}$
- d)  $\frac{2}{7}$

112. The numbers 1 to 12 are placed in a hat, and a number is selected. What is the probability that the number is 4 given that it is known to be an even number?

- a)  $\frac{2}{3}$
- b)  $\frac{3}{4}$
- c)  $\frac{1}{6}$
- d)  $\frac{1}{8}$

113. Three coins are tossed; what is the probability of getting 3 heads if it is known that at least two heads were obtained?

- a)  $\frac{1}{4}$
- b)  $\frac{2}{3}$
- c)  $\frac{1}{2}$
- d)  $\frac{3}{8}$

114. In a certain group of people, it is known that 40% of the people take Vitamins C and E on a daily basis. It is known that 75% take Vitamin C on a daily basis. If a person is selected at random, what is the probability that the person takes Vitamin E given that the person takes Vitamin C?

- a)  $\frac{5}{13}$
- b)  $\frac{8}{15}$
- c)  $\frac{3}{11}$
- d)  $\frac{7}{12}$

115. Two dice are tossed; what is the probability that the numbers are the same on both dice if it is known that the sum of the spots is 6?

- a)  $\frac{2}{3}$
- b)  $\frac{1}{6}$
- c)  $\frac{5}{4}$
- d)  $\frac{1}{5}$

116. Three coins are tossed. What are the odds in favor of getting 3 heads?

- a) 1:7
- b) 3:8
- c) 7:1
- d) 8:3

117. When two dice are rolled, what are the odds against getting doubles?

- a) 1:5
- b) 1:6
- c) 5:1
- d) 6:1

118. When a card is selected from a deck of 52 cards, what are the odds in favor of getting a face card?

- a) 5:2
- b) 3:10
- c) 12:1
- d) 2:5

119. When a die is rolled, what are the odds in favor of getting a 5 or a 6?

- a) 2:3
- b) 1:2
- c) 3:2
- d) 2:1

120. On a roulette wheel, there are 38 numbers: 18 numbers are red, 18 numbers are black, and 2 are green. What are the odds in favor of getting a red number when the ball is rolled?

- a) 19:9
- b) 9:10
- c) 10:9
- d) 9:19

121. If the odds in favor of an event occurring are 3:5, what are the odds against the event occurring?  
 a) 5:3      b) 2:5  
 c) 5:8      d) 8:5
122. If the odds against an event occurring are 8:3, what are the odds in favor of the event occurring?  
 a) 3:8      b) 11:3  
 c) 11:8      d) 8:5
123. The probability of an event occurring is 5:13. What are the odds in favor of the event occurring?  
 a) 8:5      b) 13:5  
 c) 5:13      d) 5:8
124. The probability of an event occurring is 7:12. What are the odds against the event occurring?  
 a) 5:12      b) 12:5  
 c) 5:7      d) 7:5
125. What are the odds for a fair game?  
 a) 0:0      b) 1:1  
 c) 2:1      d) 1:2
126. Expectations are also called  
 a) Probabilistic averages  
 b) Statistical averages  
 c) Mean  
 d) All of the above
127. When a game is fair the expected value would be  
 a) 1      b) 0  
 c) -1      d) 0.5
128. When three coins are tossed, what is the expected value of the number of heads?  
 a) 1      b) 2  
 c) 1.5      d) 2.5
129. A special die is made with 1 one, 2 twos, 3 threes. What is the expected number of spots for one roll?  
 a)  $3\frac{1}{2}$       b)  $1\frac{2}{3}$   
 c) 2      d)  $2\frac{1}{3}$
130. The expected number of trials in 200 tosses of a fair coin is  
 a) 100      b) 200  
 c) 50      d) 120
131. When two dice are rolled, the probability of getting a sum of 8 is  
 a)  $\frac{5}{36}$       b)  $\frac{1}{6}$   
 c)  $\frac{8}{36}$       d) 0
132. A Gallup poll found that 78% of Americans worry about the quality and healthfulness of their diets. If three people are selected at random, the probability that all three will worry about the healthfulness and quality of their diets is  
 a) 2.34      b) 0.78  
 c) 0.22      d) 0.47
133. When a die is rolled, the probability of getting a number less than 5 is  
 a)  $\frac{5}{36}$       b)  $\frac{1}{3}$   
 c)  $\frac{2}{3}$       d)  $\frac{5}{6}$
134. When a card is drawn from a deck of 52 cards, the probability of getting a heart is  
 a)  $\frac{1}{13}$       b)  $\frac{1}{2}$   
 c)  $\frac{1}{4}$       d)  $\frac{5}{52}$
135. When a die is rolled, the probability of getting an odd number less than three is  
 a)  $\frac{1}{6}$       b) 0  
 c)  $\frac{1}{3}$       d)  $\frac{1}{2}$
136. A survey conducted at a local restaurant found that 18 people preferred orange juice, 12 people preferred grapefruit juice, and 6

people preferred apple juice with their breakfasts. If a person is selected at random, the probability that the person will select apple juice is

- a)  $\frac{1}{2}$       b)  $\frac{1}{3}$   
 c)  $\frac{1}{4}$       d)  $\frac{1}{6}$

137 During a sale at a men's store, 16 white sweaters, 3 red sweaters, 9 blue sweaters, and 7 yellow sweaters were purchased. If a customer is selected at random, find the probability that the customer purchased a yellow or a white sweater.

- a)  $\frac{23}{35}$       b)  $\frac{9}{35}$   
 c)  $\frac{19}{35}$       d)  $\frac{7}{35}$

138 A card is selected from an ordinary deck of 52 cards. The probability that it is a 7 or a heart is

- a)  $\frac{17}{52}$       b)  $\frac{1}{4}$   
 c)  $\frac{4}{13}$       d)  $\frac{1}{13}$

139 Two dice are rolled; the probability of getting a sum greater than or equal to 9 is

- a)  $\frac{1}{6}$       b)  $\frac{5}{18}$   
 c)  $\frac{1}{4}$       d)  $\frac{1}{9}$

140 A card is selected from an ordinary deck of 52 cards. The probability that it is a red face card is

- a)  $\frac{3}{26}$       b)  $\frac{3}{13}$   
 c)  $\frac{1}{2}$       d)  $\frac{2}{13}$

141 Three cards are drawn from an ordinary deck of 52 cards without replacement. The probability of getting three queens is

- |    |                  |    |                 |
|----|------------------|----|-----------------|
| a) | $\frac{1}{2197}$ | b) | $\frac{3}{51}$  |
| c) | $\frac{1}{5525}$ | d) | $\frac{1}{169}$ |

142 An automobile license plate consists of 3 letters followed by 2 digits. The number of different plates that can be made if repetitions are not permitted is

- a) 7800  
 b) 1,404,000  
 c) 1,757,600  
 d) 6318

143 The number of different arrangements of the letters of the word next is

- a) 256      b) 24  
 c) 18      d) 16

144 A psychology quiz consists of 12 true-false questions. The number of possible different answer keys that can be made is

- a) 24  
 b) 144  
 c) 47,900,600  
 d) 4096

145 How many different ways can 4 boys be selected from 7 boys?

- a) 210      b) 35  
 c) 28      d) 840

146 The number of different ways 5 boys and 4 girls can be selected from 7 boys and 9 girls is

- a) 147      b) 2646  
 c) 635,040      d) 43,286

147 The number of different ways 8 children can be seated on a bench is

- a) 8      b) 256  
 c) 6720      d) 40,320

148. A card is selected from an ordinary deck of 52 cards. The probability that it is a three given that it is a red card is

- a)  $\frac{1}{13}$
- b)  $\frac{2}{13}$
- c)  $\frac{1}{4}$
- d)  $\frac{1}{2}$

149. Three cards are selected from an ordinary deck of 52 cards without replacement. The probability of getting all diamonds is

- a)  $\frac{1}{64}$
- b)  $\frac{1}{12}$
- c)  $\frac{3}{52}$
- d)  $\frac{11}{850}$

150. A coin is tossed and a card is drawn from an ordinary deck of 52 cards. The probability of getting a head and a club is

- a)  $\frac{1}{8}$
- b)  $\frac{3}{4}$
- c)  $\frac{1}{6}$
- d)  $\frac{1}{4}$

151. When two dice are rolled, the probability of getting a sum of 5 or 7 is

- a)  $\frac{1}{3}$
- b)  $\frac{2}{3}$
- c)  $\frac{5}{18}$
- d)  $\frac{1}{4}$

152. The odds in favor of an event when

$$P(E) = \frac{3}{7}$$
 are

- a) 3 : 7
- b) 7 : 3
- c) 3 : 4
- d) 10 : 3

153. The odds against an event when

$$P(E) = \frac{5}{9}$$
 are

- a) 4 : 9
- b) 4 : 5
- c) 9 : 4
- d) 13 : 4

154. The probability of an event when the odds against the event are 4 : 9 are

- a)  $\frac{9}{13}$
- b)  $\frac{4}{9}$
- c)  $\frac{5}{9}$
- d)  $\frac{4}{13}$

155. A person selects a card at random from a box containing 5 cards. One card has a 5 written on it. Two cards have a 10 written on them, and two cards have a 3 written on them. The expected value of the draw is

- a) 4.5
- b) 18
- c) 3.6
- d) 6.2

156. When a game is fair, the odds of winning will be

- a) 1 : 2
- b) 1 : 1
- c) 2 : 1
- d) 3 : 2

157. A person has 2 pennies, 3 nickels, 4 dimes, and 1 quarter in her purse. If she selects one coin at random, the expected value of the coin is

- a) 4.7 cents
- b) 6.3 cents
- c) 8.2 cents
- d) 12.4 cents

158. The number of outcomes of a binomial experiment is

- a) 1
- b) 2
- c) 3
- d) Unknown

159. A survey found that one in five Americans say that he or she has visited a doctor in any given month. If 10 people are selected at random, the probability that exactly 3 visited a doctor last month is

- a) 0.101
- b) 0.201
- c) 0.060
- d) 0.304

160. A survey found that 30% of teenage consumers received their spending money from a part time job. If 5 teenagers are selected at random, the probability that 3 of them have income from a part time job is

- a) 0.132
- b) 0.471
- c) 0.568
- d) 0.623

161. A box contains 4 white balls, 3 red balls, and 3 blue balls. A ball is

selected at random and its color is noted. It is replaced and another ball is selected. If 5 balls are selected, the probability that 2 are white, 2 are red, and 1 is blue is

- |                     |                     |
|---------------------|---------------------|
| a) $\frac{72}{365}$ | b) $\frac{41}{236}$ |
| c) $\frac{52}{791}$ | d) $\frac{1}{14}$   |

- 162 If there are 200 typographical errors randomly distributed in a 500-page manuscript, the probability that a given page contains exactly 3 errors is
- a) 0.0063
  - b) 0.0028
  - c) 0.0072
  - d) 0.0014

- 163 A recent study found that 4 out of 10 houses were undermined. If 5 houses are selected, the probability that exactly 2 are undermined is

- |                      |                     |
|----------------------|---------------------|
| a) $\frac{4}{25}$    | b) $\frac{27}{125}$ |
| c) $\frac{216}{625}$ | d) $\frac{8}{25}$   |

- 164 Two hundred tickets are sold for \$1.00 each. One prize of \$75 is awarded. What is the expected value if a person purchases one ticket?
- a) \$0.625
  - b) -\$0.625
  - c) \$1.75
  - d) -\$1.75

- 165 According to the 2000 United States Census, 12.3% of the population is Black or African American. The probability that a randomly selected U.S. resident is NOT Black or African American is
- a) 0.123
  - b) 0.877
  - c) 0.754
  - d) Cannot determine

- 166 Assume the statistics final is a multiple choice test with 40 questions. Each question has four choices with one correct answer per question. If you were to randomly guess on each of the questions, what is the

probability of getting exactly the expected number of correct answers?

- a) 0.5839
- b) 0.5605
- c) 0.25
- d) 0.1444

- 167 If the probability that an event will occur is .83, then the probability that the event will not occur is:
- a) 0.17
  - b) 0.07
  - c) 0.6889
  - d) 1.20
  - e) 83%

- 168 If the probability that event A occurs is .51, the probability that event B occurs is .64, and the probability that both A and B occur is .23, then:

- a) Events A and B are complementary.
- b) Events A and B are mutually exclusive.
- c) Events A and B are supplementary.
- d) Events A and B are not mutually exclusive.
- e) Events A and B are statistically independent.

- 169 If the probability that event A occurs is .51, the probability that event B occurs is .64, and events A and B are statistically independent, then:

- a) A and B are mutually exclusive.
- b) The probability that both A and B occur is .3264.
- c) A and B can't both occur.
- d) The probability that A occurs is 1 - (probability that B occurs).
- e) A and B have different standard deviations.

- 170 In an exponential distribution, the mean is larger than the median.

- a) True
- b) False.

- 171 A function  $p(x,y)$  of two discrete random variables  $X$  and  $Y$  can be used as a joint probability mass function provided that:

- a)  $p(x,y) \geq 0 \forall x \text{ and } y$
- b)  $\sum_{x} \sum_{y} p(x,y) = 1$

- c) Both A and B are required conditions  
d) Either A or B is a required condition
172. A function  $f(x,y)$  of two continuous random variables  $X$  and  $Y$  can be used as a joint probability density function provided that:  
a)  $f(x,y) \geq 0 \forall x \text{ and } y$   
b)  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = 1$   
c) Both A and B are required conditions  
d) Either A or B is a required condition
173. Which of the following statements are correct about the independence of two random variables?  
a) Two random variables are independent if their joint probability mass function (pmf) or their joint probability density function (pdf) is the product of the two marginal pmf's or pdf's.  
b) Independence of two discrete random variables  $X$  and  $Y$  require that every entry in the joint probability table be the product of the corresponding row and column marginal probabilities.  
c) Independence of two random variables is most useful when the description of the experiment under study tells us that  $X$  and  $Y$  have no effect on one another.  
d) All of the above are correct statements  
e) Only A and B are correct statements
174. Which of the following statements are correct if  $a$  and  $c$  are either both positive  
a)  $\text{Corr}(aX+b, cY+d) = ab \text{ Corr}(X, Y)$   
b)  $\text{Corr}(aX+b, cY+d) = \text{Corr}(aX, cY)$   
c)  $\text{Corr}(aX+b, cY+d) = \text{Corr}(X, Y)$
- d)  $\text{Corr}(aX+b, cY+d) = ab \text{ Corr}(X, Y) + bd$   
e) None of the above statements is correct
175. For any two variables  $X$  and  $Y$ ,  $\text{Corr}(2X+1, 3Y+4)$  is the same as  
a) 6  $\text{Corr}(X+1, Y+4)$   
b) 5  $\text{Corr}(X, Y) + 4$   
c) 5  $\text{Corr}(X, Y) + 5$   
d)  $\text{Corr}(X, Y)$   
e) None of the above statements is true
176. An accident occurs when both fault event A and fault event B occur. The equation for joint probability of two fault events under any circumstances is given by:  $P(A \& B) = P(A|B) \times P(B)$  where  $P(A|B)$  means the probability of fault A given that fault B has occurred. If the occurrence of fault B does not affect the probability of the occurrence of fault A, the probability of an accident, when  $P(A) = 0.1$  and  $P(B) = 0.05$ , is given by:  
a) 0.00025      b) 0.10000  
c) 0.05000      d) 0.00500
177. Alfred and Ben don't know each other but are each considering asking the lovely Charlene to the school prom. The probability that at least one of them will ask her is 0.72. The probability that they both ask her is 0.18. The probability that Alfred asks her is 0.6. What is the probability that Ben asks Charlene to the prom?  
a) 0.78      b) 0.30  
c) 0.24      d) 0.48  
e) 0.54
178. If  $A_1, A_2, A_3$  are independent events such that  $P(A_i) = p_i, i = 1, \dots, 3$  the probability that at least one of the three events will occur is  
a)  $\min\{p_1, p_2, p_3\}$   
b)  $\sum_{i=1}^3 p_i$

c)  $1 - \prod_{i=1}^3 P_i$

d)  $1 - \prod_{i=1}^3 (1 - P_i)$

e) None of the above

179. The relation

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

is called

- a) Bayes theorem
- b) The law of total probability
- c) Multiplication rule
- d) None of the above

180. The relation

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)} = \frac{P(A|B_j)P(B_j)}{P(A)}$$

is called

- a) The law of total probability
- b) Conditional probability
- c) Bayes' theorem
- d) None of these

181. Which of the following is the waiting time distribution?

- a) Poisson distribution
- b) Binomial distribution
- c) Gamma distribution
- d) Cauchy distribution

182. If two outcomes  $H_1$  and  $H_2$  are complementary, then the probability, expressed as a ratio, of one outcome is equal to

- a) the probability of the other outcome
- b) 1 times the probability of the other outcome
- c) 1 divided by the probability of the other outcome
- d) 1 plus the probability of the other outcome
- e) 1 minus the probability of the other outcome

183. Probabilities from continuous medium can be calculated by which of the following distribution

- a) Negative binomial
- b) Poisson
- c) Geometric
- d) Normal

184. In a normal distribution, a \_\_\_\_\_ is a number that divides a data set into 10 intervals, each interval containing about 1/10 or 10% of the elements in the data set."

- a) Range
- b) Coefficient
- c) Deciles
- d) Median
- e) Mode

185. A normal distribution curve is based mainly on:

- a) Mean and sample size
- b) Mean and standard deviation
- c) Range and sample size
- d) Range and standard deviation
- e) Mean and range

186. The graph showing the age of getting a driver's license in California starts and peaks at age 16, and decreases from there. This shape most closely resembles what type of distribution?

- a) Normal
- b) Binomial
- c) Uniform
- d) Exponential

187. When the posterior distribution belongs to the same family as the prior distribution is called

- a) Prior family of distributions
- b) Jeffry prior
- c) Conjugate family of distribution
- d) None of the above

188. In Bayesian inference gamma distribution is used a conjugate prior density with known mean of the following distributions

- a) Poisson
- b) Exponential
- c) Normal
- d) All of the above
- e) None of the above

189. If  $X \sim Poisson(\lambda)$  then its skewness is

- a)  $\frac{E(X, - \lambda)^3}{\sigma^3}$
- b)  $\sqrt{\lambda}$
- c)  $\frac{1}{\sqrt{\lambda}}$
- d) Both a) & c)
- e) None of these

190. For binomial distribution if  $np$  and  $nq$  are greater than 5 then binomial approaches

- a) Poisson distribution
- b) Hypergeometric distribution
- c) Normal Distribution
- d) Chi-square distribution
- e) None of the above

191. Binomial distribution approximated to Poisson distribution when

- a)  $p$  approaches to zero
- b)  $q$  approaches to one
- c)  $np$  remains fixed
- d) All of the above
- e) None of the above

192. Eighteen trials of a binomial random variable  $X$  are conducted. If the probability of success for any one trial is 0.4, then  $P(X = 7)$ .

- a)  $\binom{18}{7} (0.40)^7 (0.60)^{11}$
- b)  $\binom{18}{11} (0.40)^7 (0.60)^{11}$
- c)  $\binom{18}{7} (0.40)^7 (0.60)^{11}$
- d)  $\binom{18}{7} (0.40)^7 (0.60)^{18}$
- e)  $\binom{18}{7} (0.40)^{18} (0.60)^7$

193. Which of the following statements are true?

- a) For any discrete random variable  $X$  and constants  $a$  and  $b$ ,  $E(aX+b) = (a+b)E(X)$
- b) For any discrete random variable  $X$  and constants  $a$  and  $b$ ,  $V(ax+b) = (a+b)^2 V(X)$
- c) If a constant  $c$  is added to each possible value of a discrete random variable  $X$ , then the variance of  $X$  will be shifted by that same constant amount.
- d) If a constant  $c$  is added to each possible value of a discrete random variable  $X$ , then the expected value of  $X$  will be shifted by that same constant amount

194. In a binomial distribution we

- a) Count the number of successes until a failure is obtained
- b) Count the number of trials until a success is obtained
- c) Count the number of successes in a finite number of trials
- d) Count the number of trials until the number of successes equals the number of failures

195. How many outcomes are there for a binomial experiment?

- a) 0
- b) 1
- c) 2
- d) It varies

196. Which one is not a requirement of a binomial experiment?

- a) There are 2 outcomes for each trial.
- b) There are a fixed number of trials.
- c) The outcomes must be dependent.
- d) The probability of a success must be the same for all trials.

197. The formula for the mean of a binomial distribution is

- a)  $np$
- b)  $np(1-p)$
- c)  $n(1-p)$
- d)  $\sqrt{np(1-p)}$

189. If  $If X, \square Poisson(\lambda)$  then its skewness is

  - $\frac{E(X, -\lambda)^3}{\sigma^3}$
  - $\sqrt{\lambda}$
  - $\frac{1}{\sqrt{\lambda}}$
  - Both a) & c)
  - None of these

190. For binomial distribution if  $np$  and  $nq$  are greater than 5 then binomial approaches

  - Poisson distribution
  - Hypergeometric distribution
  - Normal Distribution
  - Chi-square distribution
  - None of the above

191. Binomial distribution approximated to Poisson distribution when

  - $p$  approaches to zero
  - $q$  approaches to one
  - $np$  remains fixed
  - All of the above
  - None of the above

192. Eighteen trials of a binomial random variable  $X$  are conducted. If the probability of success for any one trial is 0.4, then  $P(X = 7)$ .

  - $\binom{18}{7}(0.40)^7(0.60)^{11}$
  - $\binom{18}{11}(0.40)^7(0.60)^{11}$
  - $\binom{18}{7}(0.40)^7(0.60)^{11}$
  - $\binom{18}{7}(0.40)^7(0.60)^{18}$
  - $\binom{18}{7}(0.40)^{18}(0.60)^7$

193. Which of the following statements are true?

  - For any discrete random variable  $X$  and constants  $a$  and  $b$ ,  $E(aX+b) = (a+b)E(X)$
  - For any discrete random variable  $X$  and constants  $a$  and  $b$ ,  $V(aX+b) = (a+b)^2 V(X)$
  - If a constant  $c$  is added to each possible value of a discrete random variable  $X$ , then the variance of  $X$  will be shifted by that same constant amount.
  - If a constant  $c$  is added to each possible value of a discrete random variable  $X$ , then the expected value of  $X$  will be shifted by that same constant amount

194. In a binomial distribution we

  - Count the number of successes until a failure is obtained
  - Count the number of trials until a success is obtained
  - Count the number of successes in a finite number of trials
  - Count the number of trials until the number of successes equals the number of failures

195. How many outcomes are there for a binomial experiment?

  - 0
  - 1
  - 2
  - It varies

196. Which one is not a requirement of a binomial experiment?

  - There are 2 outcomes for each trial.
  - There are a fixed number of trials.
  - The outcomes must be dependent.
  - The probability of a success must be the same for all trials.

197. The formula for the mean of a binomial distribution is

  - $np$
  - $np(1-p)$
  - $n(1-p)$
  - $\sqrt{np(1-p)}$

198. The formula for a standard deviation of a binomial distribution is

- a)  $np$
- b)  $np(1-p)$
- c)  $n(1-p)$
- d)  $\sqrt{np(1-p)}$

199. Which of the following statements are correct for any two random variables  $X_1$  and  $X_2$  with mean values  $\mu_1$  and  $\mu_2$ ?

- a)  $E(2X_1 + 3X_2) = 2\mu_1 + 3\mu_2$
- b)  $E(2X_1 + 3X_2) = \mu_1 + \mu_2 + 5$
- c)  $E(2X_1 + 3X_2) = \mu_1\mu_2 + 6$
- d)  $E(2X_1 + 3X_2) = 5\mu_1\mu_2$
- e) None of the above are correct

200. If  $X_1, X_2, X_3$  are three independent random variables with variances of 2, 4, and 5, respectively, then

$$V(2X_1 + 3X_2 + 4X_3)$$

- a) 32
- b) 12
- c) 108
- d) 140

- e) None of the above answers are correct

201. If we perform a large number of independent binomial experiments, each with  $n = 10$  trials, and  $p = .60$ , then the average number of successes per experiment will be close to

- a) 10
- b) 8
- c) 6
- d) 4

202. What are the mean and standard deviation of a binomial experiment that occurs with probability of success 0.76 and is repeated 150 times?

- a) 114, 27.35
- b) 100.5, 5.23
- c) 114, 5.23
- d) 100.5, 27.35
- e) The mean is 114, but there is not enough information given to determine the standard deviation.

203. If 30% of commuters ride to work on a bus, and the probability that if 8 workers are selected at random, 3 will ride the bus.

- a) 0.361
- b) 0.482
- c) 0.254
- d) 0.323

204. If 10% of the people who take a certain medication get a headache and the probability that if 5 people take the medication, one will get a headache.

- a) 0.328
- b) 0.136
- c) 0.472
- d) 0.215

205. A survey found that 30% of all Americans have eaten pizza for breakfast. If 500 people are selected at random, the mean number of people who have eaten pizza for breakfast is

- a) 100
- b) 150
- c) 200
- d) 230

206. A survey found that 10% of older Americans have given up driving. If a sample of 1000 Americans is taken, the standard deviation of the sample is

- a) 10
- b) 8.42
- c) 9.49
- d) 5

207. A survey found that 50% of adults get the daily news from radio. If a sample of 64 adults is selected, the approximate range of the number of people who get their news from the radio is

- a) 24 and 40
- b) 30 and 34
- c) 28 and 32
- d) 26 and 36

208. Poisson distribution with parameter  $\lambda$  approaches the normal distribution when

- a)  $\lambda \rightarrow 1$
- b)  $\lambda \rightarrow 0$
- c)  $\lambda \rightarrow \infty$
- d)  $\lambda < 0$
- e)  $\lambda > 0$

209. If  $X \sim U(a, b)$  then  $Y = \frac{X-a}{b-a}$  is distributed as

- a)  $U(0, 1)$
- b)  $F(0, 1)$
- c)  $\chi^2(1)$
- d) All of the above
- e) None of the above

210. The pdf

$$f_w(w) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{w^2}{2\sigma^2}}; w \in [0, \infty]$$

is called

- a) Chi-Square distribution with "w" degree of freedom
- b) Chi-Square distribution with " $\frac{1}{2}w$ " degree of freedom
- c) Chi-Square distribution with "1" degree of freedom
- d) Standard normal distribution

211. If  $X$  follows F distribution with  $v_1$  and  $v_2$  d.f. and  $v_2 \rightarrow \infty$  then the distribution of  $v_2 X$  is

- a) Chi-square
- b) Normal
- c) T
- d) F
- e) None of these

212. If  $X$  follow a chi-square distribution with  $v$  df and  $c > 0$ , then distribution of  $cX$  is

- a) Chi-square
- b) Gamma
- c) Exponential
- d) Weibull
- e) Rayleigh

213. If  $X$  follows Rayleigh (1) distribution then the distribution of  $X^2$  is

- a) Chi-square (1)
- b) Chi-square (2)
- c) Chi-square (3)
- d) Chi-square (6)

214. If  $X_i \sim B(n, p)$  then  $\sum X_i$  is distributed as

- a) Poisson( $n\lambda$ ) distribution

- b) Binomial( $mn, p$ ) distribution
- c) Normal( $np, n\sigma^2$ ) distribution
- d) None of these

215. Probability distribution of sum of squares of  $n$  independent normal variables is

- a) Binomial distribution
- b) Normal distribution
- c) Chi-Square distribution
- d) None of the above

216. If  $X$  has a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ , and  $Z$  is the standard normal random variable whose cumulative distribution function is  $P(Z \leq z) = \varphi(z)$ , then which of the following statements are not correct?

- a)  $Z = \frac{X-\mu}{\sigma}$
- b)  $P(a \leq X \leq b) = \varphi\left(\frac{b-\mu}{\sigma}\right) - \varphi\left(\frac{a-\mu}{\sigma}\right)$
- c)  $P(X \leq a) = 1 - \varphi\left(\frac{(a-\mu)}{\sigma}\right)$
- d)  $P(X \geq b) = 1 - \varphi\left(\frac{(b-\mu)}{\sigma}\right)$
- e) All of the above statements are not correct

217. Which of the following statements about the percentiles for the standard normal distribution are correct?

- a) The 90<sup>th</sup> percentile is approximately -1.28
- b) The 10<sup>th</sup> percentile is approximately 1.28
- c) The 75<sup>th</sup> percentile is approximately .67
- d) The 15<sup>th</sup> percentile is approximately 1.04
- e) None of the above answers is correct

218. If  $X \sim B(n, p)$  and  $Y \sim B(m, p)$  are independent binomial variables, then  $X + Y$  is distributed as
- Poisson( $n\lambda$ )
  - Poisson( $m\lambda$ )
  - Binomial( $nm, p$ )
  - Binomial( $n+m, p$ )
219. If  $n$  approaches infinity  $p$  approaches zero and  $np$  remain fixed then Binomial( $n, p$ ) approaches to
- Hypergeometric distribution
  - Normal distribution
  - Poisson distribution
  - None of the above
220. If  $X \sim N(1,1)$ ;  $Y \sim N(2,2)$  and are independent, then  $X+Y$
- Is not symmetric with mean 3
  - Is not symmetric with variance 3
  - Has mean 3 and is symmetric about it
  - None of the above
221. The distribution which has variance and skewness zero is
- Normal
  - Cauchy
  - Degenerate
  - None of these
222. Geometric distribution is a special case of
- Binomial distribution
  - Hyper geometric distribution
  - Negative binomial distribution
  - None of the above
223. Which distribution requires that sampling be done without replacement?
- Geometric
  - Multinomial
  - Hypergeometric
  - Poisson
224. The mean of the hypergeometric random variable  $X$  with parameters  $n = 10$ ,  $M = 50$ , and  $N = 100$  is
- 10
  - 2
- c) 5  
d) None of the above
225. Which distribution can be used, when there are 3 or more outcomes?
- Geometric
  - Multinomial
  - Hypergeometric
  - Poisson
226. Which distribution can be used, when the variable occurs over time?
- Geometric
  - Multinomial
  - Hypergeometric
  - Poisson
227. When monitoring the status of a computer system over time, with breakdowns constituting the events of interest, the appropriate probability distribution is
- Binomial distribution
  - Negative binomial distribution
  - Poisson distribution
  - Hypergeometric distribution
228. Which distribution can be used to determine the probability of an outcome occurring on a specific trial?
- Geometric
  - Multinomial
  - Hypergeometric
  - Poisson
229. The expected value of the negative binomial random variable  $X$  with parameters  $r = 5$  and  $p = 0.8$  is
- 1.00
  - 1.25
  - 4.00
  - 5.80
230. The standard deviation of the negative binomial random variable  $X$  with parameters  $r = 10$  and  $p = 0.6$  is
- 3.33
  - 6.67
  - 2.40
  - 7.75
231. The probabilities that a page of a training manual will have 0, 1, 2, or 3 typographical errors are 0.75, 0.15, 0.10, and 0.05 respectively. If 6 pages are randomly selected, the probability that 2 will contain no errors, 2 will contain one error, one will contain 2 errors, and one will contain 3 errors is



- a) 0.078      b) 0.063  
 c) 0.042      d) 0.011
- 232 A die is rolled four times. The probability of getting two 3s, one 4, and one 5 is  
 a)  $\frac{1}{36}$       b)  $\frac{1}{72}$   
 c)  $\frac{1}{108}$       d)  $\frac{5}{36}$
- 233 If 5 cards are drawn from a deck without replacement, the probability that exactly three clubs will be selected is  
 a) 0.08      b) 0.02  
 c) 0.16      d) 0.003
- 234 Of the 20 automobiles in a used car lot, 8 are white. If 6 are selected at random to be sold at an auction, the probability that exactly 2 are white is  
 a) 0.267      b) 0.482  
 c) 0.511      d) 0.358
- 235 A board of directors consists of 7 women and 5 men. If a slate of 4 officers is selected at random, the probability that exactly 2 officers are men is  
 a) 0.375      b) 0.424  
 c) 0.261      d) 0.388
- 236 The number of boating accidents on a large lake follows a Poisson distribution. The probability of an accident is 0.01. If there are 500 boats on the lake on a summer day, the probability that there will be exactly 4 accidents will be  
 a) 0.192      b) 0.263  
 c) 0.175      d) 0.082
- 237 About 5% of the population carries a genetic trait. Assume the distribution is Poisson. The probability that in a group of 100 randomly selected people 7 people carry the gene is  
 a) 0.256      b) 0.104  
 c) 0.309      d) 0.172
- 238 When a coin is tossed, the probability of getting the first tail on the fourth toss is  
 a)  $\frac{1}{2}$       b)  $\frac{3}{4}$   
 c)  $\frac{8}{7}$       d)  $\frac{1}{16}$
- 239 An eight-sided die is rolled. The average number of tosses that it will take to get a 6 is  
 a) 6      b) 16  
 c) 8      d) 12
- 240 A ten-sided die is rolled; the average number of tosses that it will take to get three 3s is  
 a) 30      b) 8  
 c) 18      d) 40
- 241 On average how many rolls of an eight-sided die will it take to get all the faces at least once?  
 a) 28.62      b) 64  
 c) 21.74      d) 32
- 242 Which of the following distribution is used for both parametric and non parametric methods  
 a) T-distribution  
 b) Chi-Square distribution  
 c) Gamma distribution  
 d) None of the above
- 243 The mode of Geometric distribution is  
 a) 2      b) 1  
 c) 0      d) 3  
 e) May b) or c)
- 244 The sum of independent Geometric distributed random variables follow a  
 a) Binomial distribution  
 b) Hyper geometric distribution  
 c) Negative binomial distribution  
 d) None of the above
- 245 If  $X \sim \text{Gamma}(k, \theta)$  distribution, then mode is  
 a)  $k$       b)  $(k - 1)$   
 c)  $\theta$       d)  $(k - 1)\theta$

246 If  $X \sim \text{Gamma}(k, \theta)$  distribution, then skewness is

- a)  $k$
- b)  $\frac{2}{k}$
- c)  $\frac{\theta}{2}$
- d)  $\frac{2}{\sqrt{k}}$

247 For Gaussian distribution the shape parameter is

- a) Mean
- b) Variance
- c) Standard deviation
- d) Both b & c

$$248. M(t_1, t_2) = \sum_{x=0}^{n-k} \frac{n!}{x! y! (n-x-y)!} (Pe^y)^x (Pe^b)^y I_1^{t_1} I_2^{t_2}$$

is the MGF of

- a) Binomial distribution
- b) Bernoulli distribution
- c) Trinomial distribution
- d) Multinomial distribution

249. The mean, median, and mode of a distribution have the same value. What can be said about the distribution?

- a) It is exponential
- b) It is normal
- c) It is uniform
- d) None of the above

250.  $X$  and  $Y$  are independent random variables with  $\mu_X = 3.5$ ,

$\mu_Y = 2.7$ ,  $\sigma_X = 0.8$  and  $\sigma_Y = 0.65$ .

What are  $\mu_{X+Y}$  and  $\sigma_{X+Y}$ ?

- a)  $\mu_{X+Y} = 6.2$ ,  $\sigma_{X+Y} = 1.03$
- b)  $\mu_{X+Y} = 6.2$ ,  $\sigma_{X+Y} = 1.0625$
- c)  $\mu_{X+Y} = 3.1$ ,  $\sigma_{X+Y} = 0.725$
- d)  $\mu_{X+Y} = 6.2$ ,  $\sigma_{X+Y} = 1.45$
- e)  $\mu_{X+Y} = 6.2$ ,  $\sigma_{X+Y}$  cannot be determined from the information given

251. Approximately what percent of the data values are smaller than the median?

- a) 25%
- b) 50%
- c) 75%
- d) Between 0% and 99% inclusive

252. A normal probability plot is used to

- a) Determine whether the distribution is normal
- b) Plot z-values.
- c) Determine process capability.
- d) Find percent out of specification

253. The mean of a Poisson distribution is 2.94. Its standard deviation is:

- a) Not enough information is given
- b) 1.71
- c) 8.64
- d) 74.7
- e) 1.31

254. The mean of the standard normal distribution is

- a) 1
- b) 100
- c) 0
- d) Variable

255. The percent of the area under the normal distribution curve that falls within 2 standard deviations on either side of the mean is approximately

- a) 68
- b) 95
- c) 99
- d) Variable

256. The total area under the standard normal distribution curve is

- a) 50%
- b) 65%
- c) 95%
- d) 100%

257. In the graph of the standard normal distribution, the values of the horizontal axis are called

- a) x values
- b) y values
- c) z values
- d) None of the above

258. The scores on a national achievement test are approximately normally distributed with a mean of 120 and a standard deviation of 10. The probability that a randomly selected student scores between 100 and 130 is
- 68.2%
  - 34.1%
  - 48.8%
  - 81.8%
259. The heights of a group of adult males are approximately distributed with a mean of 70 inches and a standard deviation of 2 inches. The probability that a randomly selected male from the group is between 68 and 72 inches is
- 34.1%
  - 68.2%
  - 81.8%
  - 95.4%
260. The average time it takes an express bus to reach its destination is 32 minutes with a standard deviation of 3 minutes. Assume the variable is normally distributed. The probability that it will take the bus between 30.5 minutes and 33.5 minutes to arrive at its destination is
- 38.3%
  - 93.3%
  - 34.1%
  - 6.7%
261. The average time it takes pain medicine to relieve pain is 18 minutes. The standard deviation is 4 minutes. The variable is approximately normally distributed. If a randomly selected person takes the medication, the probability that the person experiences pain relief within 25.2 minutes is
- 50%
  - 68.3%
  - 75.2%
  - 96.4%
262. The average electric bill for the month of October in a residential area is \$72. The standard deviation is \$5. If a resident of the area is randomly selected, the probability that his or her electric bill for the month of October is greater than \$75 is
- 27.4%
  - 42.3%
  - 72.6%
  - 88.1%
263. What percent of the area under the standard normal distribution lies within two standard deviations of the mean?
- 68%
  - 95%
  - 100%
  - Unknown
264. When the value of a variable is transformed into a standard normal variable, the new value is called a
- x value
  - y value
  - 0 value
  - z value
265. Two people who developed simulation techniques are
- Fermat and Pascal
  - Laplace and DeMoivre
  - Von Neuman and Ulam
  - Plato and Aristotle
266. Mathematical simulation techniques use \_\_\_\_\_ numbers.
- Prime
  - Odd
  - Even
  - Random
267. A coin can be used as a simulation device when there are two outcomes and each outcome has a probability of
- $\frac{1}{4}$
  - $\frac{1}{2}$
  - $\frac{1}{3}$
  - $\frac{1}{6}$
268. Which device will not generate random numbers?
- Computer
  - Abacus
  - Dice
  - Calculator
269. The  $\chi^2$  distribution is:
- Symmetric.
  - Left skewed.
  - Right skewed.
  - Normal.
  - Uniform.
270. Given that  $X \sim U(3, 15)$ ,  $f(x) =$
- $1/15$
  - $1/18$
  - $1/12$
  - 1
271. Samples of 100 units of blood are tested for the presence or absence of

the hepatitis C antibody. The resulting data (number of units per sample testing positive) will form which type of distribution?

- a) Normal
- b) Poisson
- c) Hypergeometric
- d) Binomial
- e) Exponential

272. The number of errors per invoice for an accounts receivable function are of interest. The most appropriate probability distribution is:

- a) Normal.
- b) Poisson.
- c) Hypergeometric.
- d) Binomial.
- e) Exponential.

273. A population is bimodal. One hundred samples of size 30 are randomly collected and the 100 sample means are calculated. The distribution of these sample means is:

- a) Bimodal.
- b) Approximately exponential.
- c) Approximately Poisson.
- d) Approximately normal.
- e) Approximately uniform.

274. A population is bimodal with a variance of 5.77. One hundred samples of size 30 are randomly collected and the 100 sample means are calculated. The standard deviation of these sample means is approximately:

- |         |         |
|---------|---------|
| a) 5.77 | b) 2.40 |
| c) 1.05 | d) 0.44 |
| e) 0.19 |         |

275. Which of the following is NOT an assumption of the Binomial distribution?

- a) All trials must be identical.
- b) All trials must be independent.
- c) Each trial must be classified as a success or a failure.
- d) The number of successes in the trials is counted.
- e) The probability of success is equal to 0.5 in all trials.

276. Cans of soft drinks cost \$0.30 in a certain vending machine. What is the expected value and variance of daily revenue ( $Y$ ) from the machine, if  $X$ , the number of cans sold per day has  $E(X) = 125$ , and  $V ar(X) = 50$ ?

- a)  $E(Y) = 37.5$ ,  $V ar(Y) = 50$
- b)  $E(Y) = 37.5$ ,  $V ar(Y) = 4.5$
- c)  $E(Y) = 37.5$ ,  $V ar(Y) = 15$
- d)  $E(Y) = 37.5$ ,  $V ar(Y) = 15$
- e)  $E(Y) = 125$ ,  $V ar(Y) = 4.5$

277. Which of the following statements are true?

- a) The gamma family is a family of probability density functions that yields a wide variety of skewed distributional shapes.
- b) The expected value of a random variable  $X$  having the gamma distribution  $f(x; \alpha, \beta)$  with parameter values  $\alpha = 2$  and  $\beta = 2$  is  $E(X) = 2$ .
- c) The standard deviation of a random variable  $X$  having the gamma distribution with parameter values  $\alpha = 1$  and  $\beta = 2$  is 4.
- d) A continuous random variable  $X$  having a gamma distribution with parameters  $\alpha$  and  $\beta$  satisfy  $\alpha > 0$  and  $\beta > 0$  is said to have a standard gamma distribution if  $\alpha = 1$  and  $\beta = 0$ .
- e) All of the above statements are true.

278. Which of the following distributions have identical mean, median and mode?

- a) Normal distribution
- b) T-distribution
- c) Laplace distribution
- d) All of the above
- e) None of the above

279. If  $X \sim Laplace(\mu, b)$  parameters, then kurtosis of the distribution is

- a) 3
- b) 2
- c) 1
- d) 0
- e) None of these

- 280 The probability distribution for which its mgf does not exist while moments are exist  
 a) Hypergeometric distribution  
 b) Pareto distribution  
 c) Gamma Distribution  
 d) Both a) and b)  
 e) None of these
- 281 The probability distribution for which its mgf exist while it's mean and variance does not exist  
 a) Gamma distribution  
 b) Weibull distribution  
 c) Cauchy distribution  
 d) Pareto distribution
- 282 Gamma distribution is the analogue of discrete distribution  
 a) Uniform  
 b) Negative binomial  
 c) Geometric  
 d) Poisson  
 e) None of these
- 283 Which of the following statements are not true?  
 a) The exponential distribution is frequently used as a model for the distribution of times between the occurrences of successive events such as calls coming in a switchboard.  
 b) The exponential distribution is closely related to the Poisson process.  
 c) The chi-squared distribution with parameter  $v$  is a special case of the gamma distribution with parameters  $\alpha$  and  $\beta$  for which  $\alpha = \frac{v}{2}$  and  $\beta = 2$   
 d) All of the above statements are true.  
 e) All of the above statements are not true.
- 284 Negative binomial distribution is arises as a continuous, is a mixture of  
 a) Binomial and Poisson  
 b) Exponential and Poisson  
 c) Gamma and Poisson  
 d) Erlange and Poisson
- 285 For negative Binomial i.e.  $-NB(r, \frac{\lambda}{\lambda+r})$ , then Negative binomial converges to  
 a) Binomial  
 b) Gamma  
 c) Poisson  
 d) None of the above
- 286 For a Poisson distribution with parameter  $\lambda$   
 a) Mean and Square of S.D are Equal  
 b) Mean and Mode are equal  
 c) Mode and variance are equal  
 d) All of the above
- 287 The expression  $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$  is the general term for the  
 a) Poisson distribution.  
 b) Pascal distribution  
 c) Hypergeometric distribution  
 d) Binomial distribution.
- 288 If  $X_1 \sim Poisson(\lambda_1)$  and  $X_2 \sim Poisson(\lambda_2)$ , then the difference  $Y = X_1 - X_2$  follows a  
 a) Gamma distribution  
 b) Exponential distribution  
 c) Skellam distribution  
 d) Binomial distribution
- 289 If  $X_1 \sim Poisson(\lambda_1)$  and  $X_2 \sim Poisson(\lambda_2)$  are independent, and  $Y = X_1 + X_2$ , then the distribution of  $X_1$  conditional on  $Y = y$  is  
 a)  $Poisson(\lambda_1 + \lambda_2)$   
 b)  $Binomial(y, \frac{\lambda_1}{\lambda_1 + \lambda_2})$   
 c)  $Exponential(\lambda_1 + \lambda_2)$   
 d) None of the above

290. The numbers of light bulbs that burn out in a certain amount of time follow

- a) Exponential distribution
- b) Gamma distribution
- c) Poisson distribution
- d) Uniform distribution

291. Chi-Square distribution is a special case of

- a) Beta
- b) Normal
- c) Gamma
- d) Exponential

292. For a chi-Square distribution, if mean is 11 then variance

- a) 11
- b) 33
- c) 22
- d) 5.5

293. The MGF of chi-Square distribution with  $r$  degree of freedom, is

- a)  $(1-2t)^r$
- b)  $(1-2t)^{\frac{r}{2}}$
- c)  $(1-2t)^{\frac{r}{4}}$
- d)  $(1-2r)^{\frac{r}{2}}$

294. If random variable  $X$ , follow exponential distribution then kurtosis

- a) 2
- b) 4
- c) 6
- d) 1

295. If  $X \sim \text{Beta}(\alpha, \beta)$ , if  $\alpha = 1, \beta = 1$ , then the distribution of  $X$  is

- a) Bernoulli
- b) Uniform
- c) Gamma
- d) None of the above

296. A continuous random variable  $X$  is uniformly distributed on the interval  $[A, B]$ , then the probability density function between  $A$  and  $B$  has a value of

- a)  $1/(A-B)$
- b)  $(A+B)/2$
- c)  $A-B$
- d)  $B-A$
- e)  $A+BX$

297. A continuous random variable  $X$  is uniformly distributed over the interval  $[12, 18]$ . The variance of  $X$  is

- a) 15
- b) 6
- c) 3
- d) 9

298. The probability that " $X$  is at most 3" is interpreted as:

- a)  $P(X < 3) + P(X=3)$
- b)  $P(X < 3)$
- c)  $P(X > 3)$
- d)  $P(X > 3) + P(X=3)$

299. Let  $X$  be a continuous random variable with probability density function  $f(x)$  and cumulative distribution function  $F(x)$ . Then for any two numbers  $a$  and  $b$  with  $a < b$ , which of the following inequalities is true?

- a)  $P(a \leq X \leq b) = F(b) - F(a)$
- b)  $P(X > a) = 1 - F(a)$
- c)  $F(x) = \frac{x-a}{b-a}$
- d)  $P(X > b) = F(b)^{-1}$

300. The cumulative distribution function  $F(x)$  for a continuous random variable  $X$  is defined for every number  $x$  by which of the following inequalities?

- a)  $F(x) = P(X \geq x)$
- b)  $F(x) = 1 - P(X \leq x)$
- c)  $F(x) = P(X \leq x)$
- d)  $F(x) = P(X \geq x) - P(X \leq x)$

301. If  $X \sim \text{B}(\alpha, \theta)$  and  $Y \sim \text{G}(\beta, \theta)$  then  $\frac{X}{X+Y}$  is distributed as

- a) Bernoulli
- b) Beta
- c) Binomial
- d) None of the above

302. The Beta distribution is a special case of

- a) Wishart distribution
- b) Dirichlet distribution
- c) Multivariate normal distribution
- d) None of the above

- 303 The mode of beta distribution with parameter  $\alpha$  and  $\beta$  is exist only if
- $\alpha + \beta = 2$
  - $\alpha + \beta < 2$
  - $\alpha + \beta \geq 2$
  - $\alpha + \beta > 2$
  - Both c) and d)
- 304 The mode of beta distribution with parameter  $\alpha$  and  $\beta$  is
- $\frac{\alpha}{\alpha + \beta}$
  - $\frac{\alpha - 1}{\alpha + \beta - 2}$
  - $\frac{\alpha + 1}{\alpha + \beta - 2}$
  - $\frac{\alpha - 1}{\alpha + \beta + 2}$
- 305 For beta distribution with parameter  $\alpha$  and  $\beta$ , if  $\alpha = \beta$  then density becomes
- Positively skewed
  - Symmetric
  - Asymmetric
  - Negatively skewed
- 306 If  $X \sim Beta(\alpha, \beta)$ , then probability distribution of  $1-X$  is
- Beta
  - Binomial
  - Uniform
  - Exponential
- 307 If  $X \sim Beta(\alpha, 1)$ , then probability distribution of  $X^2$  is
- Kumaraswamy  $(\alpha, \beta)$
  - Kumaraswamy  $(1, \beta)$
  - Kumaraswamy  $(\alpha, 1)$
  - Kumaraswamy  $(1, 1)$
  - None of these
- 308 The skewness of the t-distribution with  $v$  degree of freedom is
- $\frac{6}{v-4}$
  - $\frac{6}{v+4}$
  - $\frac{3}{v-4}$
  - $\frac{3}{v+4}$
  - $\frac{6}{v-2}$
- 309 The t-distribution first derived as a
- Prior density
  - Posterior density
  - Both a) and b)
  - None of these
- 310 The t-distribution was first published by William Sealy Gosset in
- 1900
  - 1906
  - 1908
  - 1921
- 311 The distribution of Sum of squared centered independent normal variate is
- Chi-square
  - Gamma
  - F
  - Normal
  - T
- 312 For t-distribution with parameter  $v$ , if  $v=1$  then t-distribution is converted to
- Normal distribution
  - Chi-square distribution
  - Cauchy distribution
  - F distribution
  - T distribution
- 313 If  $X$  and  $Y$  follow a Standard normal distribution, then the ratio of these two variates is
- Standard normal
  - Standard Cauchy
  - Standard chi-square
  - None of the above
- 314 If  $X$  follow beta distribution, then mean of beta distribution is
- Equal to variance
  - Greater than variance
  - Smaller than variance
  - None of the above

315. For a beta distribution the range of  $X$

- a)  $X \in (0,1)$
- b)  $X \in (1,0)$
- c)  $X \in (1,\infty)$
- d)  $X \in (0,\infty)$

316. The Chi-Square distribution is used to test

- a) The variance
- b) The goodness of fit
- c) Independence
- d) All of the above

317. For Cauchy distribution which of the following is true?

- a) Mean does not exist
- b) Variance does not exist
- c) First moment does not exist
- d) Second moment does not exist
- e) All of the above

318. For a chi-square distribution the range of  $X$  is

- a)  $X \in (1,\infty)$
- b)  $X \in (1,0)$
- c)  $X \in (-\infty,\infty)$
- d)  $X \in (0,\infty)$

319. If  $X \sim \chi^2(2)$ , then equivalent distribution

- a) Exponential  $\left(\frac{1}{2}\right)$
- b) Poisson  $\left(\frac{1}{2}\right)$
- c) Exponential(2)
- d) None of the above

320. If  $X_i \sim N(\mu, 1)$ , then

$$Y = \sum_{n=1}^t X_n^2$$

- is distributed as
- a) Non central F
  - b) Chi-Square
  - c) Non central Chi-Square
  - d) None of the above

321. If  $Z_1 \sim \chi^2(V_1)$  and  $Z_2 \sim \chi^2(V_2)$  then  $\gamma$

$= \frac{Z_1/V_1}{Z_2/V_2}$  is distributed as

- a)  $F(V_1, V_2)$
- b) Non central F
- c)  $t(V_1 + V_2 - 2)$
- d) None of the above

322. The Erlang distribution is a special case of

- a) Exponential distribution
- b) Gamma distribution
- c) Chi-square distribution
- d) Beta distribution

323. For Erlang distribution the range of  $X$  is

- a)  $X \in (1,\infty)$
- b)  $X \in (1,0)$
- c)  $X \in (-\infty,\infty)$
- d)  $X \in (0,\infty)$

324. The Erlang distribution is used to measure

- a) Traffic load
- b) Time between incoming calls
- c) Probability of packet loss or delay
- d) All of the above

325. For exponential distribution the range of  $X$  is

- a)  $X \in (1,\infty)$
- b)  $X \in (1,0)$
- c)  $X \in (-\infty,\infty)$
- d)  $X \in (0,\infty)$

326. The Skewness of Exponential distribution is

- a) 1
- b) 6
- c) 2
- d) 3

327. Given that  $X \sim \text{Exp}(0.4)$ , then median is

- a) 2.5
- b) 1.25
- c) 1.73
- d) 1/4

328. Which of the following distributions is not waiting time distribution?

- a) Weibull distribution
- b) Exponential distribution
- c) Binomial distribution
- d) Gamma distribution

329. Which of the following distribution have larger variance than its mean

- a) Binomial
- b) Hypergeometric
- c) Negative binomial
- d) None of the above

330. In Reliability Analysis if time is zero then reliability

- a) 0
- b) 1
- c) -1
- d) None of the above

331. For exponential life time distribution with parameter  $\lambda$  the reliability  $R(t)$  is

- a)  $e^{-t}$
- b)  $e^{-\lambda t}$
- c)  $e^{-\lambda}$
- d) None of these

332. What is the reliability of a system at 850 hours, if the average usage of the system was 400 hours for 1650 items and the total number of failures was 145? Assume an exponential distribution.

- |        |        |
|--------|--------|
| a) 0%  | b) 36% |
| c) 18% | d) 83% |

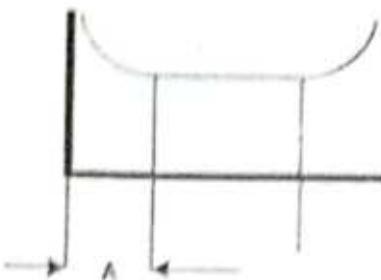
333. In a reliability test, 135 nonrepairable items are tested for two hours. Eleven failures are observed at the end of the test. The best expression for MTTF is:

- a) 11/270
- b) 270/11
- c) 1/135
- d) 11
- e) 1/11

334. If the time to failure distribution is exponential, the formula for the reliability function is. If MTBF = 285, then  $R(100)$ .

- a) 0.704
- b) 0.058
- c) 28,500
- d) 77,471

335. Refer to the diagram of the bathtub curve. Failures in the area labeled A are usually caused by:



- a) Product design failure.

- b) Fatigue.

- c) Warranty cost.

- d) Manufacturing and/or quality control errors

- e) Random failure

336. The bathtub curve can be divided into three regions. Which region is most impacted by changes in manufacturing process control?

- a) Early life region
- b) Constant failure rate region
- c) Wear-out region
- d) Within  $\pm 2s$  of the mean
- e) none of the above

337. Appropriate labels for the horizontal and vertical axes of a graph of the bathtub curve are, respectively:

- a) MTBF and time.
- b) Time and failure rate.
- c) Failure rate and MTTF.
- d) Time and number of failures
- e) Failure rate and time.

338. On the flat part (floor) of the bathtub curve:

- a) The failure rate is zero
- b) The failure rate is increasing
- c) The MTBF is constant
- d) None of the above

339. A system has five components in a series, each with a reliability of 99%. The system reliability is closest to

- a) 0.995
- b) 4.995
- c) 1.01
- d) 0.1998
- e) 1.000

340. A dual brake system on an automobile allows it to stop safely even if the line to one of the wheels leaks. This is an example of a

- a) Parallel system



353. Suppose  $X$  has density  $f(x) = \frac{1}{2^4 \sqrt{\pi}} \exp\left(-\frac{x^2}{\sqrt{2}}\right)$ ,  $-\infty < x < \infty$

then mean and variance of  $X$  is

- a)  $\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}$
- b) 1, 1
- c)  $1, \frac{1}{\sqrt{2}}$
- d)  $0, \sqrt{2}$
- e) None of the above

354. For frequency data with 7 classes, a normal distribution is fitted after estimating the parameters. If chi-square goodness of fit test is to be used without combining the classes the degree of freedom associated with chi-square test is

- a) 4
- b) 5
- c) 6
- d) 7
- e) None of the above

355. Which of the following statement is true for normal distribution?

- a)  $P(X \geq 6) = 1 - P(X < 5)$
- b)  $P(X \geq 6) = P(X > 6)$ .
- c)  $P(X \geq 2) = 1 - P(X \leq 1)$
- d)  $P(X > 2) = P(X \geq 3)$
- e) None of the above

356. Moment Generating functions of two independent random variables  $X_1$  and  $X_2$  are  $M_1(t)$  &  $M_2(t)$  then the MGF of  $X_1 + 2X_2$

- a)  $M_1(t) + 2M_2(t)$
- b)  $M_1(t) + M_2(2t)$
- c)  $2M_1(t)M_2(t)$
- d)  $M_1(t)M_2(2t)$
- e) none of the above

357. Lognormal Distribution is \_\_\_\_\_ distribution

- a) Symmetric
- b) Skewed
- c) Left skewed
- d) Right skewed
- e) None of these

358. In reliability analysis when the failure rate of constant, then which of the following distribution is used

- a) Log normal
- b) Weibull
- c) Exponential
- d) All of the above
- e) None of the above

359. The Weibull distribution first introduced by Frechet in

- a) 1915
- b) 1918
- c) 1921
- d) 1924
- e) 1927

360. The Weibull described by Waloddi Weibull in

- a) 1950
- b) 1951
- c) 1954
- d) 1957

361. In reliability analysis when the Failure rate is not constant, then which of the following distribution is used?

- a) Log normal
- b) Weibull
- c) Exponential
- d) All of the above
- e) None of the above

362. For logistic distribution with parameter  $\alpha$  &  $\beta$ ,

$\beta > 0$  and  $x \rightarrow \infty$  then  $\alpha + \beta x$

becomes

- a) 0
- b)  $\infty$
- c)  $-\infty$
- d) None of the above

363. For generalized gamma distribution with three parameters  $\mu$ ,  $\sigma$  &  $\lambda$ , if

$\lambda = 1$  this is identical to

- a) Exponential distribution
- b) Weibull distribution
- c) Normal distribution
- d) Log normal distribution
- e) None of these

364. For generalized gamma distribution with three parameters  $\mu$ ,  $\sigma$  &  $\lambda$ , if  $\lambda = 1$  &  $\sigma = 1$  this is identical to
- Exponential distribution
  - Weibull distribution
  - Normal distribution
  - Log normal distribution
  - None of these
365. Which of the following statements are true?
- The Weibull distribution is a special case of the exponential distribution
  - The exponential distribution is a special case of both the gamma and Weibull distributions.
  - There are gamma distributions that are not Weibull distributions.
  - There are Weibull distributions that are not gamma distributions.
  - All of the above statements are true
366. The shape of logistic distribution is similar to
- Gamma Distribution
  - Normal Distribution
  - Exponential Distribution
  - None of the above
367. The person who developed the concepts of game theory was
- Garry Kasparov
  - Leonhard Euler
  - John Von Neumann
  - Blaise Pascal
368. The reward for winning the game is called the
- Bet
  - Payoff
  - Strategy
  - Loss or win
369. In a game where one player pays the other player and vice versa, the game is called —— game
- A payoff
  - An even sum
  - No win
  - A zero sum
370. When both players use an optimal strategy, the amount that on average is the payoff over the long run is called the —— of the game.
- Value
  - Winnings
  - Strength
  - Odds
371. If a game is fair, the value of the game will be
- 0
  - 1
  - 1
  - Undetermined
372. Which of the following statements are not correct?
- The essence of a probability plot is that if the distribution on which the plot is based is correct, the points in the plot will fall close to a straight line.
  - If the actual distribution is quite different from the one used to construct the probability plot, the points should depart substantially from a linear pattern.
  - Sample percentiles are defined in a different way than percentiles of a population distribution are defined.
  - All of the above statements are correct.
  - None of the above statements are correct.
373. Which of the following statements are not correct?
- If the sample observations are in fact drawn from a normal distribution with mean value  $\mu$  and standard deviation  $\sigma$ , the points on the probability plot should fall close to a straight line with slope  $\sigma$  and intercept  $\mu$ .
  - Minitab statistical software provides an alternative version of a normal probability plot in which the z percentile axis is replaced by a nonlinear probability axis

- d) Lognormal and Weibull distributions are symmetric distributions  
 d) All of the above statements are correct  
 e) None of the above statements is correct
- 374 Which of the following statements are correct about a nonnormal population distribution?  
 a) It can be symmetric and has "lighter tails" than does a normal distribution, that is, the density curve declines more rapidly out in the tails than does a normal curve  
 b) It can be symmetric and has "heavy tails" compared to a normal population  
 c) It can be skewed  
 d) All of the above statements are correct  
 e) None of the above statements is correct.

375 If  $X_1, X_2, \dots, X_n$  are iid  $N(\mu, \sigma^2)$ , then  $\Sigma(X_i - \bar{X})$  is distributed as

- a)  $\chi^2(n)$   
 b)  $\chi^2(n-1)$   
 c)  $\sigma^2 \chi^2(n-1)$   
 d)  $\sigma^2 \chi^2(n)$

376 For Poisson distribution if  $\mu = 2$ , then mean of the distribution is

- a) 2/3      b) 1/3  
 c) 3/2      d) 1/2

377 If random variable  $X$  has the density function  $f(x) = \frac{1}{2} e^{-\frac{|x|}{2}}$ , then the mean and variance are

- a)  $\left(1, \frac{1}{2}\right)$       b)  $\left(2, \frac{1}{2}\right)$   
 c)  $(2, 4)$       d)  $\left(\frac{1}{2}, \frac{1}{4}\right)$

- 378 The probability distribution of the sum of chi-square variables is  
 a) Gamma  
 b) Chi-square  
 c) F  
 d) Exponential  
 e) Erlange

- 379 The Mean deviation of a normal distribution is  
 a)  $\frac{5}{4}\sigma$       b)  $\frac{4}{5}\sigma$   
 c)  $\frac{2}{5}\sigma$       d) None of these

- 380 The point of inflection of normal distribution is  
 a)  $\mu - \sigma, \mu + \sigma$   
 b)  $\mu - 2\sigma, \mu + 2\sigma$   
 c)  $\mu - 3\sigma, \mu + 3\sigma$   
 d) None of these

- 381 The maximum ordinate of a normal distribution is  
 a) 0      b) 1  
 c)  $\sigma$       d)  $\mu$

#### ANSWERS

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----

89	d	90	a	91	d	92	b	237	b	238	d	239	b	240	a
93	a	94	c	95	d	96	b	241	c	242	b	243	b	244	b
97	a	98	c	99	b	100	b	245	d	246	d	247	b	248	b
101	a	102	d	103	c	104	a	249	b	250	a	251	b	252	a
105	b	106	b	107	c	108	a	253	b	254	c	255	b	256	a
109	c	110	b	111	b	112	c	257	c	258	d	259	b	260	a
113	a	114	b	115	d	116	a	261	d	262	a	263	b	264	d
117	c	118	b	119	b	120	b	265	c	266	d	267	b	268	b
121	a	122	a	123	d	124	c	269	c	270	c	271	d	272	b
125	b	126	d	127	b	128	c	273	d	274	d	275	b	276	b
129	d	130	a	131	a	132	d	277	a	278	d	279	a	280	d
133	c	134	c	135	a	136	d	281	c	282	b	283	d	284	b
137	a	138	c	139	b	140	a	285	c	286	d	287	a	288	b
141	c	142	b	143	b	144	d	289	b	290	a	291	c	292	b
145	b	146	b	147	d	148	a	293	b	294	d	295	b	296	a
149	d	150	a	151	c	152	c	297	c	298	a	299	b	300	c
153	b	154	a	155	d	156	b	301	b	302	a	303	d	304	b
157	c	158	b	159	b	160	a	305	b	306	a	307	c	308	a
161	d	162	c	163	c	164	b	309	b	310	c	311	b	312	c
165	b	166	d	167	a	168	d	313	b	314	b	315	a	316	d
169	b	170	a	171	c	172	c	317	e	318	d	319	b	320	c
173	d	174	c	175	d	176	d	321	a	322	b	323	d	324	d
177	b	178	c	179	b	180	c	325	d	326	c	327	c	328	b
181	c	182	e	183	b	184	c	329	c	330	b	331	b	332	d
185	b	186	d	187	c	188	d	333	b	334	a	335	d	336	a
189	d	190	c	191	d	192	c	337	b	338	c	339	a	340	a
193	d	194	c	195	c	196	c	341	c	342	a	343	a	344	a
197	a	198	d	199	a	200	d	345	b	346	b	347	d	348	c
201	c	202	c	203	c	204	a	349	c	350	c	351	c	352	c
205	b	206	c	207	a	208	c	353	d	354	a	355	c	356	b
209	a	210	c	211	a	212	b	357	d	358	c	359	e	360	b
213	b	214	b	215	c	216	c	361	b	362	b	363	b	364	a
217	c	218	d	219	c	220	c	365	e	366	b	367	c	368	b
221	c	222	c	223	c	224	c	369	d	370	a	371	a	372	c
225	b	226	d	227	c	228	a	373	b	374	d	375	c	376	a
229	b	230	a	231	d	232	c	377	c	378	b	379	b	380	a
233	a	234	d	235	b	236	c	381	d						

## **REGRESSION ANALYSIS AND ECONOMETRICS**

---

1. Suppose a large image file is downloaded from the Internet. The speed of the data, in bits per second (bps), is plotted as a function of time in seconds. In this situation, data speed is considered
  - a) The dependent variable
  - b) The independent variable
  - c) A constant function
  - d) Nondecreasing
2. Suppose the path of a hurricane's center is plotted as a set of points on a map at 12-hour intervals, with the latitude and longitude lines as graphic coordinates. A continuous plot of the path of the storm center can be approximated using
  - a) Interpolation
  - b) Extrapolation
  - c) Curve fitting
  - d) Correlation
3. If independent and dependent both are numerical variables then we use
  - a) Simple regression
  - b) Multiple Regression
  - c) Logistic regression
  - d) Both a) and b)
4. If independent is Nominal/Numerical and dependent is Nominal then we use
  - a) Cox regression
  - b) Multiple Regression
  - c) Logistic regression
  - d) Factor analysis
5. If independent is Nominal/Numerical and dependent is numerical (censored) then we use
  - a) Cox regression
  - b) Multiple Regression
  - c) Logistic regression
  - d) Factor analysis
6. Regression model used for estimating growth rate is
  - a) Log-Log
  - b) Linear-Log
  - c) Log-Linear
  - d) Linear
7. The constant in Cobb-douglas production function indicates
  - a) Elasticity
  - b) Return to a scale
  - c) Efficiency
  - d) None of the above
8. In a linear regression of  $y$  on  $x$ ,  $y = a + bx$ ,
  - a) Both variables have the same variance
  - b) A fitted straight line goes through  $(0, 0)$
  - c) The variance of  $x$  is constant for all values of  $y$
  - d)  $x$  increases by  $b$  units for every unit increase in  $y$
  - e)  $x$  is the explanatory variable.
9. The slope of a line can be calculated by:
  - a) Picking one point along the line, and calculating the  $Y$  coordinate of that point divided by the  $X$  coordinate of that point.
  - b) Picking any two points along the line and calculating the difference in the  $X$  coordinates divided by the difference in the  $Y$  coordinates.
  - c) Picking any two points along the line and calculating the RISE over the RUN, where RISE is the vertical distance between the two points and RUN is the horizontal distance between the two points (and sign matters)
  - d) picking any two points along the line and calculating the RUN over the RISE, as defined in c. above

10. For a two-variable population regression function (PRF),
- The conditional expected value of  $Y$ , given  $X$ , is assumed to lie along a straight line given by the formula  $\hat{Y} = \beta_1 + \beta_2 X$ .
  - The marginal expected value of  $Y$  is assumed to lie along a straight line given by the formula  $\hat{Y} = \beta_1 + \beta_2 X$ .
  - The actual value of  $Y$ , given  $X$ , is assumed to lie along a straight line given by the formula  $\hat{Y} = \beta_1 + \beta_2 X$ .
  - The variance of  $Y$  around its expected value is assumed to be zero.
11. In order to estimate the parameters of a simple ordinary least squares regression model, we need all but which of the following conditions to be met?
- Expected value of the error term is zero.
  - Homoscedasticity (constant variance of the errors).
  - No correlations among the error terms.
  - No correlations between the error term and the explanatory variables.
  - Normally distributed error term.
12. Why do we usually minimize the sum of squared deviations from the regression line, rather than the sum of just the deviations themselves?
- It is difficult to take derivatives if you do not have the squared term in the function to be minimized.
  - It is important to attach greater damage to the objective function to be minimized if you have an observation that is a great distance from the line, and the use of the square of the deviation accomplishes this.
  - The sum of the deviations is zero, so you would not get very far with this sort of objective function to be minimized.
13. If we minimize the sum of the squared deviations, we get the same parameter estimates for the line as we would get if we minimized the sum of the absolute deviations, and the calculus for the minimization is easier.
14. When we refer to the deviation form of  $X$  and  $Y$ , we mean:
- The difference between  $Y$  and  $X$ .
  - The differences between  $Y$  and  $X$  and their respective sample means.
  - The distance between  $Y$  and the regression line, and the difference between  $X$  and the regression line.
  - The differences between  $Y$  and  $X$  and their respective population means.
15. Which of the following constitutes an ordered list of two input variables, an intermediate variable, and an output variable?
- Lemonade stand (% sugar, % lemons, taste, profit)
  - Shoe sales (comfort rating, material, color, shoe sizes)
  - Sandwich making for lunch (peanut butter, jelly, weight, transportation cost)
  - All of the above fit the definition.
  - Both a) and c)
16. Which of the following constitutes an ordered list of two input variables, two intermediate variables, and an output variable?
- Lemonade stand (% sugar, % lemons, taste rating, material cost, total profit)
  - Chair manufacturing (wood type, saw type, stylistic appeal, waste, profit)
  - Chip manufacturing (time in acid, % silicon, % dopant, % acceptable, profit)
  - All of the above fit the definition.
  - Both a) and b)

16. Within the class of linear unbiased estimators, the OLS estimator have minimum variance is called  
 a) Consistency  
 b) Sufficiency  
 c) Gauss Markov  
 d) None of these
17. In Gauss Markov OLS estimator are BLUE, here B mean  
 a) Best  
 b) Unbiased  
 c) Minimum variance  
 d) Both a) and c)  
 e) None of these
18. The regression coefficient of Y on X is 0.5 and that of X on Y is 0.6. Hence the correlation coefficient between X and Y is  
 a) 0.5  
 b) 0.09  
 c) 0.25  
 d) -0.3  
 e) None of the above
19. A simple linear regression equation:  
 a) Always describes a line with zero slope.  
 b) Is not affected by changes in the units in which variables are measured.  
 c) Describes a line which goes through the point defined by the means of the two variables.  
 d) Is affected by the choice of dependent variable.
20. In regression analysis the explanatory and dependant variable must be following respectively  
 a) Deterministic and Stochastic  
 b) Stochastic and fixed  
 c) Fixed and deterministic  
 d) Random and stochastic
21. OLS techniques involves  
 a) Maximizing slope  
 b) Minimizing constant  
 c) Minimizing squared sum of errors  
 d) Minimizing errors
22. Which of the following statements are true?  
 a) One way to study the fit of a model is to superimpose a graph of the best-fit function on the scatter plot of the data.  
 b) An effective approach to assessment of model adequacy is to compute the fitted or predicted values  $\hat{y}_i$  and the residuals  $\hat{e}_i = y_i - \hat{y}_i$ , then plot various functions of these computed quantities, and examine the plots either to confirm our choice of model or for indications that the model is not appropriate.  
 c) Multiple regression analysis involves building models for relating the dependent variable y to two or more independent variables.  
 d) All of the above statements are true.  
 e) None of the above statements are true.
23. Which of the following statements are not true?  
 a) If a particular standardized residual is 1.5, then the residual itself is 3 estimated standard deviations larger than what would be expected from fitting the correct model.  
 b) Plotting the fitted or predicted values  $\hat{y}_i$  on the vertical axis versus the actual values on the horizontal axis is a diagnostic plot that can be used for assessing model validity and usefulness.  
 c) A normal probability plot of the standardized residuals is a basic plot that many statisticians recommend for an assessment of model validity and usefulness.  
 d) All of the above statements are true.  
 e) None of the above statements are true.

- 24 Which of the following statements are not true?
- Provided that the model is correct, no residual plot should exhibit distinct patterns.
  - Provided that the model is correct, the residuals should be randomly distributed about 0 according to a normal distribution, so all but a very few standardized residuals should lie between -2 and +2 ( i.e., all but a few residuals are within 2 standard deviations of their expected value 0 ).
  - If we plot the fitted or predicted values on the vertical axis versus the actual values  $\hat{y}$ , on the horizontal axis, and the plot yields points close to the  $60^{\circ}$  line, then the estimated regression function gives accurate predictions of the values actually observed.
  - All of the above statements are true.
  - None of the above statements are true.
- 25 Quite frequently, residual plots as well as other plots of the data will suggest some difficulties or abnormality in the data. Which of the following statements are not considered difficulties?
- Nonlinear probabilistic relationship between  $x$  and  $y$  is appropriate.
  - The variance of the error term  $\epsilon$  (and of  $Y$ ) is a constant  $\sigma^2$ .
  - The error term  $\epsilon$  does not have a normal distribution.
  - The selected model fits the data well except for very few discrepant or outlying data values, which may have greatly influenced the choice of the best-fit function.
  - One or more relevant independent variables have been omitted from the model.
- 26 The Pearson product moment correlation measures the degree of \_\_\_\_\_ relationship present between two variables.
- Curvilinear
  - Nonlinear
  - Linear and quadratic
  - Linear
- 27 The coefficient of multiple determination  $R$  is
- $SSE/SST$
  - $SST/SSE$
  - $1-SSE/SST$
  - $1-SST/SSE$
  - $(SSE + SST) / 2$
- 28 If the value of the coefficient of multiple determination  $R^2$  is .80 for a quadratic regression model, and that  $n = 11$ , then the adjusted  $R^2$  value is
- 0.75
  - 0.80
  - 0.85
  - 0.90
  - 0.95
- 29 Which of the following statements are not true?
- The exponential function  $y = \alpha e^{\beta x}$  is intrinsically linear.
  - The power function  $y = \alpha x^{\beta}$  can be linearized by the transformations  $y' = \log(y)$  and  $x' = \log(x)$ .
  - The function  $y = \alpha + \gamma x^{\beta}$  is intrinsically linear.
  - All of the above statements are true.
  - None of the above statements are true.
- 30 Which of the following statements are true?
- The function  $y = \alpha + \beta \log(x)$  is intrinsically linear.

- b) The reciprocal function  $y = \alpha + \frac{\beta}{x}$  can be linearized by the transformation  $x' = \frac{1}{x}$
- c) For an exponential function relationship  $y = \alpha e^{\beta x}$  only  $y$  is transformed to achieve linearity.
- d) All of the above statements are true.
- e) None of the above statements are true.
31. Which of the following statements are not true?
- The function is  $y = \alpha + \gamma e^{\beta x}$  intrinsically linear.
  - Intrinsically linear functions lead directly to probabilistic models which, though not linear in  $x$  as a function, have parameters whose values are easily estimated using ordinary least squares.
  - The multiplicative exponential model  $y = \alpha e^{\beta x}$  is intrinsically linear probabilistic model.
  - All of the above statements are true.
  - None of the above statements are true.
32. Which of the following statements are not true?
- To balance the cost of using more parameters against the gain in the coefficient of multiple determination  $R^2$ , many statisticians use the adjusted  $R^2$ .
  - It is always true  $SSE_k \geq SSE_{k+1}$  for any  $k^{th}$  degree polynomial regression model.
  - It is always true  $R_k^2 > R_{k+1}^2$  when  $k > k'$  for any  $k^{th}$  degree polynomial regression model.
- d) All of the above statements are true.
- e) None of the above statements are true
33. In an Analysis of Variance for linear regression of  $y$  on  $x$  using  $n$  pairs of observations
- Degrees of freedom for residual are  $(n - 1)$
  - $r^2$  (or  $R^2$ ) is near to 100% when the regression line is almost straight
  - The sum of squares for regression has 2 degrees of freedom
  - A confidence interval for  $b$  may be calculated from the residual sum of squares
  - The F-test examines the hypothesis  $\beta \neq 0$ .
34. In multiple linear regressions
- A response variable is explained using two or more explanatory variables
  - Explanatory variables must all be measured in the same units
  - Explanatory variables must not be correlated with one another
  - Stepwise fitting methods should be used
  - A "best subset" of explanatory variables is bound to have the greatest possible value of  $R^2$ .
35. The derivative with respect to  $X$  of the square of  $(a - bX)$  is:
- $b$  times  $X$ -squared
  - $-2b$
  - $2(a - bX)b$
  - $-2b(a - bX)$
36. Keeping in mind the formula for the ordinary least squares slope estimate, what can be said about the slope for a regression of  $Y$  on  $X$ , versus the slope for a regression of  $X$  on  $Y$ ?
- The slopes are reciprocals.
  - The slopes are not reciprocals.
  - The slopes are not reciprocals.
  - The slopes are identical.

- 37 Given a set of ordered pairs  $(x, y)$  so that  $S_x = 1.6$ ,  $S_y = 0.75$ ,  $r = 0.55$ . What is the slope of the least-square regression line for these data?
- 1.82
  - 1.17
  - 2.18
  - 0.26
  - 0.78
- 38 A study found a correlation of  $r = 0.58$  between hours per week spent watching television and hours per week spent exercising. That is, the more hours spent watching television, the less hours spent exercising per week. Which of the following statements is most accurate?
- About one-third of the variation in hours spent exercising can be explained by hours spent watching television.
  - A person who watches less television will exercise more.
  - For each hour spent watching television, the predicted decrease in hours spent exercising is 0.58 hrs.
  - There is a cause-and-effect relationship between hours spent watching television and a decline in hours spent exercising.
  - 58% of the hours spent exercising can be explained by the number of hours watching television.
- 39 A least-squares regression line for predicting performance on a college entrance exam based on high school grade point average (GPA) is determined to be Score = 273.5 + 91.2 (GPA). One student in the study had a high school GPA of 3.0 and an exam score of 510. What is the residual for this student?
- 26.2
  - 43.9
  - 37.1
  - 26.2
  - 37.1
- 40 Once you have calculated the ordinary least squares slope coefficient, the intercept estimate can be found by:
- Using the inverse of the slope coefficient.
  - Taking each value of  $Y$  and subtracting the slope times each corresponding value of  $X$ .
  - Since the sample regression function always goes through the means of the sample data, find the intercept by taking the mean of  $Y$  minus the slope times the mean of  $X$ .
  - Using the slope times the mean of  $X$ .
- 41 An association between two variables (like marked crosswalks and pedestrian accidents) implies that one variable is the cause of the other.
- True
  - False
- 42 For any observation, the amount of  $Y$  at a given value of  $X$ , say  $X_0$ , consists of
- A basic amount of  $Y$  that is always present, regardless of the magnitude of  $X$  (i.e. the intercept, or constant term)
  - Some amount of  $Y$  that is induced by the magnitude of  $X$  (equal to the slope coefficient times the amount of  $X$  at that point)
  - Some amount of  $Y$  that represents random error, and differs randomly from observation to observation
  - All of the above
  - None of the above
- 43 Which is NOT a property of the sample regression function (SRF)?
- It always passes through the means of  $X$  and  $Y$ .
  - The mean of the fitted  $Y$  values will equal the mean of the actual  $Y$  values.
  - The mean of the sample residuals (regression errors) will be zero.
  - There are only as many different possible sample regressions.

- functions as there are observations in the sample.
- e) None of these
44. Using deviation notation (e.g.  $x = X - \text{mean of } X$ ), the correct formula for the slope of an ordinary least squares regression line is:
- (sum of the  $xy$ )/(sum of the  $x$ -squareds).
  - (sum of the  $xy$ )/(sum of the  $x$ ).
  - (sum of the  $y$ -squareds)/(sum of the  $x$ -squareds).
  - (sum of the  $x$ )/(sum of the  $y$ ).
45. Which of the four maintained hypotheses of ordinary least squares regression makes it easy to calculate the variance of the slope estimator in a simple regression (across all possible samples that might be drawn)?
- Expected value of the regression error term is zero.
  - Covariances among different error terms are zero.
  - Variance of the error term is constant.
  - Covariance between error term and the explanatory variable is zero.
  - Both b) and c)
46. In a simple regression model, we use  $(n-2)$  instead of  $(n-1)$  as the degrees of freedom correction in the calculation of the sample variance for the errors. Why?
- There are two variables in the model, not just one.
  - Two parameters now need to be estimated before we can calculate the error term for each observation.
  - Because two of anything is better than one.
  - Because we just like to confuse students with a lot of silly formulas that must be memorized.
47. Taking into account the formula for the variance of the slope estimator in a simple regression model, how might you choose to design your data collection to ensure that you have more-accurate slope estimates?
- Make sure all of your  $X$  observations are close to the mean of  $X$ .
  - Make sure your  $X$  observations are widely dispersed around the mean of  $X$ .
  - Make sure that you have a lot of large and small values of  $Y$  in your sample.
  - Make sure that the values of  $Y$  in your sample are all very close to the mean of  $Y$ , so you have good resolution due to dense data.
48. Suppose the path of a hurricane's center is plotted as a set of points on a map at 12-hour intervals, with the latitude and longitude lines as graphic coordinates. A continuous plot of the path of the storm center can be approximated using
- Interpolation
  - Extrapolation
  - Curve fitting
  - Correlation
49. Which of the following statements is false?
- All relations are constant functions.
  - All constant functions are relations.
  - The values in a bar graph do not always add up to 100%.
  - Zero correlation is indicated by widely scattered points on a graph
50. Which of the following conditions will lead to a smaller variance for the intercept estimator for your simple regression model?
- $X$  values closely packed around the mean of  $X$  in your sample.
  - $X$  values clustered far from the origin of the  $X$  axis.
  - Small sample sizes.
  - Small error variance in the population regression function.
51. Statistical software calculates that the correlation between mold temperature

- and shrinkage is 0.92. This means that:
- The best way to control shrinkage is to control mold temperature.
  - Changes in mold temperature influence the quality of the product.
  - The software is calculating incorrectly.
  - Mold temperature changes cause shrinkage changes.
  - None of the above.
52. The first step in determining whether there is a linear association or correlation between two variables is to:
- Calculate the correlation coefficient.
  - Construct a Pareto analysis.
  - Construct a scatter diagram.
  - Calculate the standard deviation.
  - Construct a control chart.
53. Data are collected in  $xy$  pairs and a scatter diagram shows that the points are grouped very close to a straight line that tips down on its right-hand end. A reasonable value for the coefficient of correlation is:
- 0.8
  - 0
  - 0.9
  - 1
  - 1.3
  - 1.8
54. Rank correlation should NOT be used when
- The pairs of data are measured as continuous variables
  - There are no outliers in the data
  - The pairs of data are jointly normally distributed
  - Data are measures as integer (discrete uniform) variables sample size is small.
55. Paired data collected from a process are (2.3, 9.7), (2.4, 10.6), (3.5, 12.8), (4.1, 14.2). Use this data to calculate the coefficient of linear correlation. Its value is approximately:
- 0.58
  - 0.77
  - 0.42
  - 1.29
  - None of the above
56. Paired data collected from a process are: (2.3, 9.7), (2.4, 10.6), (3.5, 12.8), (4.1, 14.2). Calculate the slope of the best-fitting "least squares" linear regression equation. Its value is approximately:
- 0.58
  - 0.77
  - 0.42
  - 1.29
  - None of the above
57. Assume that a regression equation has slope  $b_1 = 2.5$  and intercept  $b_0 = -77.3$ . Estimate the  $y$ -value that would result if  $x$  is 12:
- 47.3
  - 925.1
  - 79.8
  - 62.8
  - None of the above
58. Two variables have a linear coefficient of -0.97. One concludes from this information that:
- The two variables are not related
  - Changes in one of the variables could not be causing changes in the other.
  - There is strong evidence that changes in one of the variables causes changes in the other.
  - The Two variables are strongly correlated.
59. Why is the so-called zero hypothesis usually the first hypothesis of interest regarding the slope parameter(s) of a regression model?
- In specifying a model, we hope to find zero slopes.
  - We expect the t-test statistic for a null hypothesis about a slope to be zero if the null hypothesis is true.
  - If the true value of a slope could plausibly be zero, we need to reconsider our model, because

this suggests that the explanatory variable we have proposed does not really explain any of the variations in  $Y$  that we observe in the data.

- d) We call it a zero hypothesis because convention dictates that the null hypothesis be denoted by a zero subscript on the parameter of interest.
- 60 In multiple regression analysis with  $n$  observations and  $k$  predictors (or equivalently  $k+1$  parameters), inferences concerning a single parameter  $\beta_i$  are based on the standardized variable

$$T = \frac{\hat{\beta}_i - \beta_i}{S_{\hat{\beta}_i}}$$
 which has a t-distribution

with degrees of freedom equal to

- a)  $n-k+1$   
 b)  $n-k$   
 c)  $n-k-1$   
 d)  $n+k-1$   
 e)  $n+k+1$
- 61 Which of the following is not a necessary condition for doing inference for the slope of a regression line?
- a) For each given value of the independent variable, the response variable is normally distributed.  
 b) The values of the predictor and response variables are independent.  
 c) For each given value of the independent variable, the distribution of the response variable has the same standard deviation.  
 d) The mean response values lie on a line.
- 62 A first-order no-interaction model has the form  $\hat{Y} = 5 + 3x_1 + 2x_2$ . As  $x_1$  increases by 1 unit, while  $x_2$  holding fixed, then  $y$  will be expected to

- a) Increase by 10  
 b) Increase by 5  
 c) Increase by 3  
 d) Decrease by 3  
 e) Decrease by 6

63. In terms of the data on a dependent variable  $Y$ , and two explanatory variables  $X$  and  $Z$  in the observed sample, it is possible to construct formulas for the OLS multiple regression slope coefficients on  $X$  and  $Z$ . These formulas are proportionately messier than the formulas for slopes in the simple regression case
- a) In practice, it is much harder to determine the slopes in a multiple regression than in a simple regression.  
 b) As you add more explanatory variables, the algebra for calculating the slopes become more and more intractable.  
 c) It takes several more steps to determine slopes in a multiple regression context than in a simple regression context.  
 d) None of the above.
64. In multiple regression models, the error term  $\epsilon$  is assumed to have
- a) A mean of 1.  
 b) A standard deviation of 1.  
 c) A variance of 0.  
 d) Negative values.  
 e) Normal distribution.
65. Incorporating a categorical variable with 4 possible categories into a multiple regression model requires the use of
- a) 4 indicator variables  
 b) 3 indicator variables  
 c) 2 indicator variables  
 d) 1 indicator variable  
 e) no indicator variables at all
66. In a simple regression, the point estimates of the slope and intercept parameters from a single sample are
- a) Jointly distributed and uncorrelated

- b) Jointly distributed and positively correlated  
 c) Jointly distributed and negatively correlated  
 d) Not jointly distributed at all; they are two separate random variables
67. For comparing goodness of fit across two models that have the same dependent variable, but differing numbers of explanatory variables, we do not usually rely on ordinary R-squared because  
 a) It is not valid  
 b) It can be expected to be larger if more variables are used, so of course the bigger model would always be preferred  
 c) It does not apply to multiple regression, only simple regression  
 d) None of the above
68. If  $X$  and  $Y$  are independent and "a, b" are constant then  $\text{Var}(aX+bY)$  is  
 a)  $\text{Var}(X)+\text{Var}(Y)$   
 b)  $a^2\text{Var}(X)+b^2\text{Var}(Y)$   
 c)  $\text{Var}(X)+\text{Var}(Y)+2\text{COV}(X,Y)$   
 d) None of these
69. The random variables  $X$  and  $Y$  are such that  $X+Y$  and  $X-Y$  are positively correlated then  
 a)  $V(X+Y) > V(X-Y)$   
 b)  $V(X+Y) < V(X-Y)$   
 c)  $V(X) < V(Y)$   
 d)  $V(X) > V(Y)$   
 e) None of the above
70. Two variables are uncorrelated if  
 a)  $V(X+Y)=V(X-Y)$   
 b)  $V(XY)=V(X)V(Y)$   
 c)  $V(X-Y)=V(X)-V(Y)$   
 d)  $V(X)=V(Y)$   
 e) None of the above
71. If  $V(X+Y)=V(X)+V(Y)$ , then value of correlation coefficient is  
 a) 1  
 b) 0  
 c) 0.5  
 d) -1  
 e) None of the above
72. If  $V(X+Y)=V(X)+V(Y)$ , then value of correlation coefficient is  
 a) 1  
 b) 0  
 c) 0.5  
 d) -1  
 e) None of the above
73. If  $X$  and  $Y$  are independent then correlation ( $X, Y$ ) is  
 a) 1  
 b)  $\frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y}$   
 c) 0  
 d) None of these
74. Suppose the covariance between  $Y$  and  $X$  is 12, the variance of  $Y$  is 25, and the variance of  $X$  is 36. What is the correlation coefficient ( $r$ ), between  $Y$  and  $X$ ?  
 a) 0.160  
 b) 0.400  
 c) 0.013  
 d) None of the above
75. Which of the following statements regarding the coefficient of determination is least accurate? The coefficient of determination:  
 a) Cannot decrease as independent variables are added to the model.  
 b) May range from -1 to +1  
 c) Is the percentage of the total variation in the dependent variable that is explained by  
 d) None of the above
76. If "R-sq 98.1%." Which of the following is/are true?  
 I. There is a strong positive linear relationship between the explanatory and response variables.  
 II. There is a strong negative linear relationship between the explanatory and response variables.

- III. About 98% of the variation in the response variable can be explained by the regression on the explanatory variable.
- I and III only
  - I or II only
  - I or II (but not both) and III
  - II and III only
  - I, II, and III
77. Determine and interpret the correlation coefficient for the two variables  $X$  and  $Y$ . The standard deviation of  $X$  is 0.05, the standard deviation of  $Y$  is 0.08, and their covariance is -0.003.
- +0.75 and the two variables are positively associated.
  - 0.75 and the two variables are negatively associated
  - 1.33 and the two variables are negatively associated
  - None of the above
78. Which of the following statements regarding a correlation coefficient of 0.60 for two variables  $Y$  and  $X$  is least accurate? This correlation:
- Is significantly different from zero.
  - Indicates a positive covariance between the two variables.
  - Indicates a positive linear relation between the two variables.
  - None of the above
79. Which of the following statements about covariance and correlation is least accurate?
- There is no relation between the sign of the covariance and the correlation.
  - The covariance and correlation are always the same sign, positive or negative
  - A zero covariance implies a zero correlation.
  - None of the above
80. Which term is least likely to apply to a regression model?
- Coefficient of variation.
  - Goodness of fit
  - Coefficient of determination
- d) All of the above  
e) Both b) and c) not a)
81. Which of the following statements regarding scatter plots is most accurate? Scatter plots:
- Illustrate the scatterings of a single variable
  - Are used to examine the third moment of a distribution (skewness).
  - Illustrate the relationship between two variables
  - None of the above
82. One of the limitations of correlation analysis of two random variables is the presence of outliers, which can lead to which of the following erroneous assumptions?
- The absence of a relationship between the two variables, when in fact, there is a linear relationship.
  - The presence of a nonlinear relationship between the two variables, when in fact, there is a linear relationship.
  - The presence of a nonlinear relationship between the two variables, when in fact, there is no relationship whatsoever between the two variables.
  - None of the above
83. Limitations of regression analysis include all of the following except
- Regression results do not indicate anything about economic significance.
  - Parameter instability.
  - Outliers may affect the estimated regression line.
  - None of the above
84. If the correlation between two variables is -1.0, the scatter plot would appear along a
- Straight line running from southwest to northeast.
  - A curved line running from southwest to northeast.

- c) Straight line running from northwest to southeast  
d) Inverse relation between two variables  
e) Both c) and d)
85. Consider the regression equation  $\hat{Y} = 2.83 + 1.5X$ , then the predicted value of the dependent variable when the value of an independent variable equals 2?  
a) 5.83  
b) 2.83  
c) -0.55  
d) None of the above
86. Two variables,  $x$  and  $y$ , seem to be exponentially related. The natural logarithm of each  $y$  value is taken and the least-squares regression line of  $\ln(y)$  on  $x$  is determined to be  $\ln(\hat{y}) = 3.2 + 0.42x$ , then the predicted value of  $y$  when  $x = 7$ ?  
a) 464.05  
b) 1380384.27  
c) 521.35  
d) 6.14  
e) 1096.63
87. R-Square statistic indicates  
a) Significance of a coefficient  
b) Significance of all coefficients  
c) Goodness of fit  
d) Interval estimate of the coefficient
88. Consider the following analysis of variance (ANOVA) table:
- | Source     | Sum of squares | Degrees of freedom | Mean square |
|------------|----------------|--------------------|-------------|
| Regression | 556            | 1                  | 556         |
| Error      | 679            | 50                 | 13.5        |
| Total      | 1,235          | 51                 |             |
- The  $R^2$  for this regression is:  
a) 0.55  
b) 0.82  
c) 0.45  
d) 0.85  
e) None of the above
89. If the regressors do not include a constant, then the  $R^2$  can be negative.  
a) True  
b) False
90. A dependent variable is regressed against a single independent variable across 100 observations. The mean squared error is 2.807, and the mean regression sum of squares is 117.5. What is the correlation coefficient between the two variables?  
a) 0.55  
b) 0.30  
c) 0.99  
d) 0.75  
e) None of the above
91. Consider the following analysis of variance (ANOVA) table:
- | Source     | Sum of squares | Degrees of freedom | Mean square |
|------------|----------------|--------------------|-------------|
| Regression | 200            | 1                  | 200         |
| Error      | 400            | 40                 | 10          |
| Total      | 600            | 41                 |             |
- The  $R^2$  and the F-statistic are, respectively:  
a)  $R^2 = 33\%$  and  $F = 20.0$ .  
b)  $R^2 = 33\%$  and  $F = 2.0$ .  
c)  $R^2 = 50\%$  and  $F = 2.0$ .  
d) None of the above
92. Regression analysis has a number of assumptions. Violations of these assumptions include which of the following?  
a) Independent variables that are not normally distributed  
b) Residuals that are not normally distributed  
c) A zero mean of the residuals  
d) Residuals are dependent  
e) Both b) and d)

93. The most appropriate measure of the degree of variability of the actual Y-values relative to the estimated Y-values from a regression equation is the:
- Standard error of the estimate (SEE).
  - Sum of squared errors (SSE).
  - Coefficient of determination ( $R^2$ ).
  - None of these
94. The standard error of estimate is closest to the:
- Standard deviation of the independent variable.
  - Standard deviation of the residuals.
  - Standard deviation of the dependent variable.
  - None of the above
95. The standard error of the estimate measures the variability of the:
- Predicted y-values around the mean of the observed y-values.
  - Actual dependent variable values about the estimated regression line.
  - Values of the sample regression coefficient.
  - None of the above
96. The standard error of the estimate in a regression is the standard deviation of the:
- Differences between the actual values of the dependent variable and the mean of the dependent variable.
  - Dependent variable.
  - Residuals of the regression.
  - None of the above
97. Which of the following statements about the standard error of the estimate (SEE) is least accurate?
- The SEE will be high if the relationship between the independent and dependent variables is weak.
  - The SEE may be calculated from the sum of the squared errors and the number of observations.
  - The larger the SEE the larger the  $R^2$ .
  - None of the above
98. If X and Y are perfectly correlated, regressing Y onto X will result in which of the following:
- The regression line will be sloped upward.
  - The alpha coefficient will be zero.
  - The standard error of estimate will be zero.
  - None of the above
99. One major limitation of the correlation analysis of two random variables is when two variables are highly correlated, but no economic relationship exists. This condition most likely indicates the presence of
- Spurious correlation.
  - Outliers.
  - Nonlinear relationships.
  - None of the above
100. One of the limitations of correlation analysis of two random variables is the presence of outliers, which can lead to which of the following erroneous assumptions?
- The absence of a relationship between the two variables, when in fact, there is a linear relationship.
  - The presence of a nonlinear relationship between the two variables, when in fact, there is a linear relationship.
  - The presence of a nonlinear relationship between the two variables, when in fact, there is no relationship whatsoever between the two variables.
  - None of the above
101. In order to have a negative correlation between two variables, which of the following is most accurate?
- Either the covariance or one of the standard deviations must be negative.
  - The covariance must be negative

- c) The covariance can never be negative.  
d) None of the above
102. Which model does not lend itself to correlation coefficient analysis?  
a)  $Y = X + 2$ .  
b)  $Y = X - 2$ .  
c)  $Y = X^3$   
d) None of the above
103. A sample covariance of two random variables is most commonly utilized to  
a) Identify and measure strong nonlinear relationships between the two variables.  
b) Calculate the correlation coefficient, which is a measure of the strength of their linear relationship.  
c) Estimate the "pure" measure of the tendency of two variables to move together over a period of time.  
d) None of the above
104. If  $X$  and  $Y$  are dependent and  $\text{Var}(X) = 4$ , then covariance is  
a)  $4\rho\sigma$ ,  
b)  $2\rho\sigma$ ,  
c)  $\frac{4\rho}{\sqrt{\sigma^2}}$ ,  
d) 0  
e) None of these
105. Limits for correlation coefficient.  
a)  $-1 \leq r \leq +1$   
b)  $0 \leq r \leq +1$   
c)  $-1 \leq r \leq 0$   
d)  $-1 \leq r \leq +2$   
e) None of these
106. The coefficient of correlation.  
a) Cannot be negative  
b) Cannot be positive  
c) Always positive  
d) Can either be positive or negative?

107. The product moment correlation coefficient is obtained by
- a)  $r = \frac{\sum XY}{xy}$   
b)  $r = \frac{\sum XY}{n\sigma_x\sigma_y}$   
c)  $r = \frac{\sum XY}{n\sigma_x}$   
d) All of these  
e) None of these
108. If  $\text{cov}(x,y) = 0$ , then  
a)  $x$  and  $y$  are correlated  
b)  $x$  and  $y$  are uncorrelated  
c)  $x$  and  $y$  are linearly related  
d) None of these
109. The covariance between two random variables  $X$  and  $Y$  is  $C_{XY}$  then the covariance between  $aX$  and  $bY$  is  
a)  $abC_{XY}$ .  
b)  $|ab|C_{XY}$   
c)  $\frac{|a|}{|b|}C_{XY}$   
d)  $\frac{|b|}{|a|}C_{XY}$   
e) None of the above
110. If you have two data sets, one for June, and one for July, for which you fit regression models as follows: the dependent variable is SALES and the explanatory variables are PRICE and ADVERT (advertising expenditure). Then:  
a) You can compare the goodness of fit of the June data and that of the July data using ordinary R-squared  
b) You can compare the goodness of fit of the June data and that of the July data using adjusted R-squared

- c) You cannot compare the goodness of fit in the two data sets using R-squared at all  
d) Either a) or b)
111. Ordinary R-squared will equal adjusted R-squared.  
a) Only when there are no explanatory variables  
b) Only when there is just one explanatory variable  
c) Only when there are the same number of regressors in each model  
d) None of these
112. A 95% joint confidence ellipse is  
a) The full set of null hypotheses about population regression function  $(\beta_1, \beta_2)$  pairs that cannot be rejected statistically by the data  
b) Centered on the observed  $(b_1, b_2)$  pair for the estimating sample  
c) Will capture the true but unknown  $(\beta_1, \beta_2)$  values for the PRF providing the sample upon which the  $(b_1, b_2)$  point estimates are based is not so rare as to happen less than 5% of the time  
d) All of the above
113. Joint confidence ellipses are the only way econometricians have to test hypotheses about sets of coefficients in a regression model  
a) True  
b) False
114. Where does the Central Limit Theorem (Normal Approximation Rule) figure into hypothesis testing in regression models?  
a) The conditional distribution of the errors around the population regression function is always approximately normal, so we can use t-tests for hypothesis testing  
b) The marginal distribution of the dependent variable ( $Y$ ) is approximately normal, so we can use t-tests for hypothesis testing
- c) The marginal distribution of the explanatory variable ( $X$ ) is approximately normal, so we can use t-tests for hypothesis testing  
d) Even if the conditional distribution of the errors around the population regression function is not normal, the distributions of the slope and intercept estimators are approximately normal, so we can use t-tests for hypothesis testing
115. The most common regression statistic for goodness-of-fit is called R-squared because  
a) In simple regression, the square of the sample correlation ( $r$ ) between  $Y$  and  $X$  gives a measure of the degree of linear association between the dependent and explanatory variable that is greater as this statistic approaches one and less as this statistic approaches zero.  
b) Using lower case r-squared is less impressive.  
c) Squaring R gives a larger number, and therefore a greater goodness of fit.  
d) Early econometricians could not think of a Latin term to use.
116. The abbreviations ESS and RSS stand for:  
a) Error sum-of-squares and Regression sum-of-squares  
b) Expanded sum-of-squares and Restricted sum-of-squares  
c) Explained sum-of-squares and Residual sum-of-squares  
d) Estimated sum-of-squares and Revised sum-of-squares
117. If your regression model is  $Y_i = b_1 + b_2 \cdot X_i + e_i$ , then the expected (predicted) value of  $Y$  given  $X$  (when  $b_1$  and  $b_2$  have been estimated by OLS) is given by  
a)  $Y_i + e_i$   
b)  $Y_i - b_1 - b_2 \cdot X_i$   
c)  $b_1 + b_2 \cdot X_i + e_i$   
d)  $b_1 + b_2 \cdot X_i$

8. In order to produce an estimate of the standard error of the predicted value of  $Y_i$ , we need to use the rule for calculating  
 a) The variance of a linear combination of two random variables, namely  $b_1$  and  $b_2$   
 b) The variance of a linear function of a single random variable, namely  $X$ .  
 c) The variance of a constant times a random variable, namely  $b_2 \cdot X$ .  
 d) The variance of a random variable, namely  $Y$ .
119. A researcher is hoping to find a predictive linear relationship between the explanatory and response variables in her study. Accordingly, as part of her analysis she plans to generate a 95% confidence interval for the slope of the regression line for the two variables. The interval is determined to be (0.45, 0.80). Which of the following is /are true?  
 I. She has good evidence of a predictive linear relationship between the variables.  
 II. It is likely that there is a non-zero correlation ( $r$ ) between the two variables.  
 III. It is likely that the true slope of the regression line is 0.  
 a) I and II only  
 b) I and III only  
 c) II and III only  
 d) I only  
 e) II only
120. A confidence interval for mean prediction, in the simple regression context, is narrower  
 a) The closer to the sample mean value of  $X$  is the value of  $X$  at which the prediction is being made  
 b) The larger the sample size  
 c) The greater the dispersion of  $X$  in the sample  
 d) All of the above
121. How does Variation differ from Variance?  
 a) Variation divides a sum of squared deviations by sample size ( $n$ ) and
- variance divides it by degrees of freedom ( $n-1$ ).  
 b) Variation divides a sum of deviations by ( $n-1$ ) whereas variance divides a sum of squared deviations by ( $n-1$ ).  
 c) Variation is a sum of squared deviations, whereas sample variance is variation divided by ( $n-1$ ), provided we are talking about a single variable, not a regression error term.  
 d) Variation is a synonym for sample range.
122. In the case of simple regression (only one explanatory variable), the coefficient of determination,  $R^2$ , is the same as  
 a) The adjusted  $R^2$ -value  
 b) The square of the standard error of the slope estimate  
 c) The squared value of the correlation between the dependent and explanatory variables  
 d) The sum of squared deviations of the values of  $Y$  from the marginal sample mean of  $Y$
123. Midterm grades are determined partly by time spent studying and partly by efficiency of studying time. Studying time is measured, but efficiency is not. If you fail to control for efficiency when regression midterm grades on studying time, and if studying time is inversely related to efficiency of studying, what bias can you expect in the coefficient on studying time?  
 a) the coefficient on studying time will be biased downwards, perhaps even becoming negative  
 b) the coefficient on studying time will be biased downward, but cannot become negative  
 c) the coefficient on studying time will be biased upwards  
 d) there will be no bias in the coefficient on studying time
124. A researcher finds that by including quantity of rainfall in an agricultural yield regression the positive

- coefficient on irrigation roughly quintuples. How can this be explained?
- If cannot be explained. Adding more variables means the explanatory power in the model has to be shared by more variables, so the coefficients can only go down in absolute value.
  - Before including rainfall, greater amounts of irrigation were proxying for lower rainfall, since the two are likely to be negatively correlated.
  - There must be other omitted variables, such as fertilizer, variations in which were being captured by the irrigation variable.
  - Irrigation amounts were actually five times greater than the original data suggested.
- 125 Suppose you work for an unscrupulous attorney for whom it is very important that you get rid of a positive coefficient on age in a model to explain the number of clients assigned to different employees at XYZ Company, which has been hit with an age discrimination suit. What sort of thing might this attorney ask you to do?
- Find some variable negatively correlated with age that might plausibly have a positive effect on client loads, and put it in the model too.
  - Find something completely uncorrelated with age that could also explain the lower client loads, and put it into the model as well.
  - Find something positively correlated with age that we can tell a story might have a positive effect on client loads, and put it in the model too.
  - Find something positively correlated with age that would tend to increase client loads and include it in the model too.
- 126 If  $X$  has a negative effect on  $Y$  and  $Z$  has a positive effect upon  $Y$ , and  $X$  and  $Z$  are negatively correlated, what is the expected consequence of omitting  $Z$  from a regression of  $Y$  on  $X$ ?
- The estimated coefficient on  $X$  will be biased downwards (too negative).
  - The estimated coefficient on  $X$  will be biased upwards (insufficiently negative).
  - The estimated coefficient on  $X$  will be biased upwards to the point of becoming positive.
  - The estimated coefficient on  $X$  (which should be positive) will be biased downwards to the point of becoming negative.
- 127 The basic automatic F-test statistic produced in conjunction with OLS estimation is designed to test:
- The null hypothesis that all regression parameters are simultaneously zero.
  - The null hypothesis that all of the slope parameters are simultaneously zero.
  - The null hypothesis that all slopes are identical.
  - The null hypothesis that the unrestricted model is true.
- 128 If we wish to test restrictions on a subset of regression parameters, in general, we run one model with the restrictions imposed and one model without the restrictions imposed and we focus our attention on the:
- t-test statistics in the unrestricted model.
  - F-test statistic for the unrestricted model.
  - The analysis-of-variance-from-means results for both the restricted and unrestricted models.
  - The maximized value of the log-likelihood for the unrestricted model.

129. If we wish to test the null hypothesis that  $\beta_1 = \beta_2 = \beta_3$ , in a model with 30 observations and five explanatory variables other than the intercept term, we will need to compare the F-test statistic for this hypothesis with the tabulated (or computer-generated) critical values of an F-distribution with:
- 3 and 25 degrees of freedom
  - 3 and 24 degrees of freedom
  - 2 and 24 degrees of freedom
  - 2 and 25 degrees of freedom
130. The adjusted coefficient of multiple determination is adjusted for
- The value of the error term  $\epsilon$
  - The number of dependent variables in the model
  - The number of parameters in the model
  - The number of outliers
  - The level of significance  $\alpha$
131. For a particular sample, if you wish to choose between a model that uses  $Y$  and one that uses  $LY = \log(Y)$  as the dependent variable, you should
- Prefer the model with the highest R-squared
  - Prefer the model with the highest adjusted R-squared
  - Prefer the model with the most negative log-likelihood value
  - Prefer the model with the largest log-likelihood value
132. The log-likelihood value for a regression model is
- The square of the F-test statistic
  - The prob-value associated with the R-squared value for this sample
  - The log of the midpoint of the confidence interval for the slope
  - The log of the joint conditional probability density function for the set of  $Y$  values observed in the sample
133. Econometricians often prefer fully log-log specifications extant because
- The always provide a better fit to the data
- b) They are easier to estimate  
 c) Their slopes can be interpreted as (constant) elasticities  
 d) You cannot take a logarithm of a negative number
134. We might consider introducing a quadratic term (or terms) in some  $X$  variable in a regression model
- If theory or intuition tells us that the effect on  $Y$  of a given change in  $X$  will not always be the same at all values of  $X$
  - Because the quadratic term might be a good replacement for the linear term in  $X$
  - Because R-squared only works if all the variables in the model are squared
  - All of the above
135. If you wish to use a set of dummy variables to capture six different categories of an explanatory variable, you should use six different dummy variables, each equal to one when the observation is a member of a specific category and equal to zero otherwise.
- True
  - False
  - Uncertain
136. Dummy variables can be used in the process of deseasonalization of time-series data that have regular periodic fluctuations (such as days of the week, months or quarters over the course of a year).
- True
  - False
  - Uncertain
137. In comparing linear models with log-log models, we take logarithms of all non-negative continuous variables, but we do not use the logs of dummy variables. This is because:
- Log of 1 is zero and log of 0 is undefined.
  - Adam Smith decreed that economists should never log a dummy variable.

- c) Only dummies attempt to log dummy variables.  
d) None of the above.
138. Which of the following is not a characteristic of the *F* distribution?  
a) It is a continuous distribution.  
b) The test statistic  $F_{\text{calc}}$  can never be negative.  
c) Its degrees of freedom vary, depending on  $\alpha$ .  
d) It is used to test for overall significance in a regression.
139. The unexplained sum of squares measures variation in the dependent variable about  
a) The mean of the *Y* values.  
b) The estimated *Y* values.  
c) The mean of the *X* values.  
d) The *Y*-intercept.
140. Which of the following is most useful in checking the normality assumption of the errors in a regression model?  
a) The *t* statistic for each coefficient.  
b) The leverage statistic for each residual.  
c) The histogram of all residuals.  
d) The VIF statistic for each predictor.
141. A multiple regression analysis with two independent variables yielded the following results.  $SS(\text{Total}) = 798$ ,  $SS(\text{Regression}) = 738$ ,  $SS(\text{Error}) = 60$ . The multiple correlation coefficient is  
a) 0.2742  
b) 0.0752  
c) 0.9248  
d) 0.9617
142. In a least squares multiple regression all of the following are true regarding residuals except  
a) They may be used to test for multicollinearity.  
b) They are differences between observed and estimated values of *Y*.  
c) Their sum will always equal zero even if there are outliers.  
d) They may be used to detect heteroscedasticity.
143. If you re-run a regression, omitting a predictor  $X_6$ , which would be unlikely?  
a) The new *t*'s will decline if  $X_6$  was a relevant predictor.  
b) The new standard error will increase if  $X_6$  was a relevant predictor.  
c) The remaining estimated  $\beta$ 's will change if  $X_6$  was collinear with other predictors.  
d) The numerator degrees of freedom for the *F* test will increase.
144. A fitted multiple regression equation is  $Y = 12 + 3X_1 - 5X_2 + 7X_3 + 2X_4$ . When  $X_1$  increases 2 units and  $X_2$  increases 2 units as well, while  $X_3$  and  $X_4$  remain unchanged, what change would you expect in your estimate of *Y*?  
a) Decrease by 2.  
b) Decrease by 4.  
c) Increase by 4.  
d) None of these.
145. A test is conducted in 22 cities to see if giving away free transit system maps will increase the number of bus riders. In a regression analysis, the dependent variable *Y* is the increase in bus riders (in thousands of persons) from the start of the test until its conclusion. The independent variables are  $X_1$  = the number (in thousands) of free maps distributed, and a binary variable  $X_2 = 1$  if city has free downtown parking, 0 otherwise. The estimated regression equation is. In city 3, the observed *Y* value is 7.3 with  $X_1 = 140$  and  $X_2 = 0$ . The residual for city 3 (in thousands) is:  
a) 6.15  
b) 1.15  
c) 4.83  
d) 1.57185
146. The Measurement of economical relationship with respect to statistics is called  
a) Mathematical statistics  
b) General theory of economics  
c) Econometrics  
d) Managerial economics

147. The theory which postulates exact relationship between various economic magnitudes is called  
 a) Statistics  
 b) Mathematical statistics  
 c) Mathematical economics  
 d) None of these
148. Difference between mathematical and Econometric model is  
 a) The constant  
 b) The error term  
 c) The slope coefficient  
 d) All of the above
149. Data collected for the same sample, at repeated time points is called  
 a) Crossectional data  
 b) Panel data  
 c) Time series data  
 d) Engineering data
150. The variables that are formed on the basis of estimation is called  
 a) Dependent variables  
 b) Endogenous variables  
 c) Independent variables  
 d) Explanatory variables  
 e) Both c) and d)
151. The variables which are formed on the basis of explanatory variables are called  
 a) Independent variables  
 b) Response variables  
 c) Dependent variables  
 d) Both b) and c)
152. A scatter diagram is a graph of  
 a) Two qualitative variable  
 b) One qualitative and other quantitative  
 c) Two quantitative variables  
 d) None of the above
153. A variable that is presumed to cause a change in another variable is called  
 a:  
 a) Categorical variable  
 b) Dependent variable  
 c) Independent variable  
 d) Intervening variable  
 e) None of the above
154. An independent variable in one study may be a dependent variable in another study.  
 a) True  
 b) False
155. A positive correlation is present when  
 a) Two variables move in opposite directions.  
 b) Two variables move in the same direction.  
 c) One variable goes up and one goes down.  
 d) Several variables never change.
156. When interpreting a correlation coefficient expressing the relationship between two variables, it is very important to avoid.  
 a) Checking the strength of relationship  
 b) Jumping to the conclusion of causality  
 c) Checking the direction of the relationship  
 d) Expressing a relationship with a correlation coefficient
157. If older people are less likely to attend church than younger people, the relationship between age and church attendance is:  
 a) Positive  
 b) Negative  
 c) Curvilinear  
 d) Symmetric
158. A correlation coefficient of .10 between SAT score and college GPA indicates:  
 a) A weak association  
 b) A strong association  
 c) A statistically significant association  
 d) Both b) and c)
159. A nonspurious relationship implies that:  
 a) There is a noncausal association between the variables  
 b) There is a positive association between the variables

- c) There is a negative association between the variables  
 d) There is a causal link between the variables
160. Which correlation is the strongest?  
 a) +0.10      b) -0.95  
 c) +0.90      d) -1.00
161. The correlation between intelligence test scores and grades is  
 a) Positive  
 b) Negative  
 c) Perfect  
 d) They are not correlated
162. Pearson's correlation coefficient can be calculated for any regression line  
 a) True  
 b) False
163. It is hypothesized that among the elderly there is a relationship between marital status and happiness in this hypothesis, happiness is a \_\_\_\_\_ variable and marital status is a \_\_\_\_\_ variable while age is a \_\_\_\_\_ variable.  
 a) Independent, dependent, control  
 b) independent, control, dependent  
 c) Dependent, independent, control  
 d) Dependent, control, independent
164. Which technique of causal modeling involves narrowing the dependent variable?  
 a) Sub classification  
 b) Elaboration  
 c) Cross tabulation  
 d) Specification  
 e) None of these
165. How do you calculate the coefficient of alienation?  
 a) Multiply the standard deviation by the mean  
 b) Square the correlation coefficient  
 c) Subtract the coefficient of determination from one  
 d) Add the KR-20 coefficient to the square root of the sample size minus one  
 e) None of these
166. If the correlation between two variables is 0.75, then the value of coefficient of alienation is  
 a) 0.70  
 b) 0.56  
 c) 0.44  
 d) 0.50
167. The number of police officers and the number of crimes are positively related. This relationship is  
 a) A causal relationship  
 b) A direct relationship  
 c) A probabilistic causal relation  
 d) A spurious relationship
168. Partial correlation analysis involves  
 a) Examining the relationship between two or more variables controlling for additional variables statistically  
 b) Including only one group in a correlational analysis  
 c) Matching participants on potential confounding variables  
 d) Limiting the sample to individuals at a constant level of an extraneous variable
169. The positive correlation between teachers' salaries and the price of liquor is  
 a) Spurious  
 b) Due to a third-variable  
 c) Nonspurious  
 d) Both a) and b)
170. Which of the following is considered a special case of the general linear model?  
 a) A variable  
 b) Partial correlation  
 c) Analysis of covariance  
 d) Both b) and c)
171. GLM refers to which of the following?  
 a) General Logit Model  
 b) General Limited Model  
 c) General Lab Model  
 d) General Linear Model
172. If your dependent (Y) variable is a (0,1) type dummy variable that

- indicates which of two outcomes describes each observation:
- You cannot use ordinary least squares because the estimation algorithm will fail, much as it does in the presence of perfect multicollinearity
  - You can use OLS to get a general idea about which explanatory variables are probably important determinants of the Y variable, but it is not the ideal method
  - There are virtually no instances when we would want to have a (0,1) dummy dependent variable, so the issue does not come up
  - It is not possible to estimate a regression-type model if the dependent variable is not a continuous variable
173. In modeling with an observation with a standardized residual greater than 2.5 in absolute value is called
- Heteroscedastic term
  - Autocorrelated term
  - An Outlier
  - None of the above
174. When the point (15, 2) is included, the slope of regression line ( $y = a+bx$ ) is  $b = 0.54$ . The correlation is  $r = 0.82$ . When the point is removed, the new slope is  $-1.04$  and the new correlation coefficient is 0.95. What name is given to a point whose removal has this kind of effect on statistical calculations?
- Outlier
  - Statistically significant point
  - Point of discontinuity
  - Unusual point
  - Influential point
175. The assumption not relevant to a two variable model is
- No Autocorrelation
  - No Multicollinearity
  - No Heteroscedasticity
  - All of the above
176. Normality assumption of error term is relevant for
- Estimation of slope
  - Estimation of constant
  - Inference making
  - All of the above
177. Interval estimates are wider due to
- Smaller the standard error
  - Bigger the standard error
  - Lower confidence level
  - Higher significance level
178. In Modeling an influential observation is measured by
- Standardized residual
  - Studentized residual
  - Cook's distance
  - None of the above
179. If there are replicates in Y observations at some of the values of X, then to obtain correct estimate of  $\sigma^2$ , we use
- Kalmogrove test
  - Anderson darling test
  - Lack of fit test
  - Runs test
  - None of the above
180. If Response variable has replicates of regressors. The ANOVA residual sum of squares can be portioned into a pure error plus a remainder is called
- Within variation
  - Between the variation
  - Lack of fit component
  - None of the above
181. If the assumption of CLRM  $E(\epsilon_i) = \sigma^2$  is violated then the problem arise is
- Errors in variables
  - Auto correlation
  - Heteroscedasticity
  - Multicollinearity
  - None of these
182. Which of the following is not the test heteroscedasticity
- Spearman rank correlation test
  - Goldfeld and Quandt test

- c) Durbin Watson Test  
 d) Park Test  
 e) None of these
- 183 For Goldfeld and Quandt Test the central observations which are omitted are  
 a)  $\frac{1}{2}$  of the total observations  
 b)  $\frac{1}{5}$  of the total observations  
 c)  $\frac{1}{4}$  of the total observations  
 d)  $\frac{1}{3}$  of the total observations  
 e) None of these
- 184  $\frac{1}{\sigma^2}$  transformation is used for  
 (Subject GRE, 2010)  
 a) Heteroscedasticity  
 b) Autocorrelation  
 c) Multicollinearity  
 d) None of the above
- 185 Which of the following reasons make OLS a less-than-perfect choice of an estimation method when the dependent variable is a dummy variable?  
 a) There will be major-league heteroscedasticity problems  
 b) If we interpret the fitted value of Y as the probability of the 1 outcome, then for some sets of explanatory variable values, this fitted probability will be either negative or greater than one  
 c) The conditional distribution of the Y variable, given a particular set of X values, is a two-valued discrete distribution, not a continuous approximately normal distribution  
 d) All of the above.
- 186 Indirect Least square (ILS) is used when the equation is  
 a) Under identified  
 b) Exactly identified  
 c) Over identified  
 d) None of these
- 187 Indirect Least square (ILS) is used for  
 a) System of equations  
 b) Single equation  
 c) Both a) and b)  
 d) None of these
- 188 Instrumental variable method is a  
 a) Single equation method  
 b) System of equation method  
 c) Both a) and b)  
 d) None of these
- 189 Instrumental variable method is used when the equation is  
 a) Under identified  
 b) Exactly identified  
 c) Over identified  
 d) None of these
- 190 Which of the following method is useful for the solution of simultaneous equation bias  
 a) ILS  
 b) WLS  
 c) Instrumental variable method  
 d) None of these
- 191 Instrumental variable method must satisfy the condition  
 a) IV is correlated with endogenous variable  
 b) IV is correlated with error term  
 c) Both a) and b)  
 d) None of these
- 192 IV estimates are consistent but asymptotically  
 a) Efficient  
 b) Not efficient  
 c) Both a) and b)  
 d) None of these
- 193 For large sample 2SLS estimates is  
 a) Biased  
 b) Unbiased  
 c) Efficient  
 d) Both b) and c)  
 e) None of these
- 194 When the Multicollinearity is occur in the data then the estimates are  
 a) Biased  
 b) Unbiased  
 c) Efficient  
 d) Both b) and c)
- 195 The existence of linear relationship among explanatory variable in regression analysis is called  
 a) Correlation  
 b) Multicollinearity

- c) Factor analysis  
d) None of these
- 196 The FARRAR and GLAUBER test is used for testing  
a) Heteroscedasticity  
b) Multicollinearity  
c) Autocorrelation  
d) Error in variables
- 197 The FARRAR and GLAUBER test is a combination of  
a) 5 tests  
b) 3 tests  
c) 2 tests  
d) None of these
- 198 For testing the pattern of multicollinearity which of the following test is used  
a) F-test  
b) T-test  
c) Chi-test  
d) None of these
- 199 Which of the following test is used for testing the multicollinearity  
a) FARRAR and GLAUBER test  
b) VIF  
c) Rule of thumb  
d) Condition number index  
e) All of the above
- 200 Variance inflation factor (VIF)  $> 10$  indicate that the Multicollinearity is  
a) None  
b) High degree  
c) Moderate degree  
d) None of these
- 201 The remedial for Multicollinearity is  
a) Ridge regression  
b) Dropping or adding variable  
c) Principle component  
d) Transformation  
e) All of the above
- 202 Ridge regression is the remedial measure of (Subject GRE, 2010)  
a) Autocorrelation  
b) Multicollinearity  
c) Heteroscedasticity  
d) None of the above
- 203 Multicollinearity may be present in your model if  
a) Some or all of your t-ratios are individually small (cannot reject individual slopes being zero), but the F-test value is large (rejects all slopes simultaneously zero)  
b) The t-ratios and the F-test value are both small  
c) The maximized value of the log-likelihood is large  
d) The R-squared value for the fitted model is large
- 204 Multicollinearity exists if there is a high correlation between your dependent variable and any one (or more) of your explanatory variables.  
a) True  
b) False  
c) Uncertain
- 205 Multicollinearity  
a) Compromises the goodness-of-fit of your regression model  
b) Can make it difficult to distinguish the individual effects on the dependent variable of one regressor from that of another regressor.  
c) Causes the values of estimated coefficients to be insensitive to the presence or absence of other variables in the model  
d) Causes t-ratios to be high and F-tests to be low
- 206 The best way to detect multicollinearity in a regression model is to  
a) Look at plots of each regressor against all other regressors  
b) High pairwise correlations  
c) Look for low t-ratios  
d) Look for high R-squared values  
e) conduct auxiliary regressions and look for high R-squareds
- 207 Auxiliary R-squared regressions consist of  
a) Y regressed on each of the X-variables in turn

- b) Each of the X-variables regressed on each other X variable, one at a time  
 c) Each of the X-variables regressed on all other X variables at once  
 d) Each of the X-variables regressed on all possible subsets of X-variables
- 208 A circumstance where you might not be too concerned about multicollinearity would be  
 a) When the collinearity in your variables will persist over all possible forecasting scenarios  
 b) When only one pair of collinear variables will begin to vary independently under your forecasting scenarios  
 c) When you care only about slope coefficients, not goodness-of-fit  
 d) When you only want to make *ceteris paribus* predictions about the effect of one regressor on Y, holding the others constant
- 209 Pairwise correlations will always successfully reveal whether multicollinearity is present in your estimating sample.  
 a) True      b) False  
 c) Uncertain
- 210 Good ways to deal with multicollinearity include  
 a) Getting additional data, where the collinearities between your regressors may not be as strong  
 b) Using outside information about the relationship between coefficients on collinear variables to reduce the number of slope parameters to be estimated  
 c) Using different types of data, such as cross-sectional, to identify some of the coefficients and then imposing these coefficient values in the dataset afflicted by multicollinearity  
 d) Dropping one of the multicollinear variables from your regression model  
 e) All of the above except d)
- 211 Ideally, all potential regressors in an OLS specification should be orthogonal (vary arbitrarily with respect to each other, be uncorrelated) because  
 a) The coefficients on the included variables would be insensitive to the presence or absence of other variables  
 b) The sampling variation in any set of regressors would successfully span the domain of the regression function, allowing individual coefficients to be robustly estimated  
 c) Adding more regressors from this set of potential regressors would improve the R-squared value but the central tendency of individual slope estimates would be unaffected  
 d) All of the above
- 212 When the explanatory variable and error term are correlated, then which of the following data problem is arise  
 a) Multicollinearity  
 b) Errors in variables  
 c) Simultaneous equation bias  
 d) Both b) and c)  
 e) None of these
- 213 In the presence of error of measurement the OLS estimates are  
 a) Biased  
 b) Unbiased  
 c) Inconsistent  
 d) Both a) and c)  
 e) None of these
- 214 In the presence of error of measurement, the OLS estimates asymptotically are  
 a) Biased  
 b) Unbiased  
 c) Inconsistent  
 d) Both a) and c)  
 e) None of these
- 215 For the solution of errors in variables which of the following method is used for obtaining consistent estimates  
 a) Inverse least square

- b) Weighted regression  
 c) IV method  
 d) All of the above
216. If the regression model includes not only current but also lagged values of the explanatory variable is called  
 a) Auto regressive model  
 b) Distributed lagged model  
 c) None of these  
 d) Both a) and b)
217. The problem of Almon Lag scheme is  
 a) To specify lag length  
 b) To specify the degree of polynomial  
 c) Both a) and b)  
 d) None of these
218. Which of the following test is used for testing heteroscedasticity  
 a) Park test  
 b) Glejser test  
 c) Goldfeld and Quandt test  
 d) All of the above
219. Which of the following method is used for the solution of heteroscedasticity  
 a) WLS  
 b) Log transformation  
 c) GLS  
 d) Both a) and b)
220. For the solution of heteroscedasticity, when variance is known then which of the following method is used?  
 a) Generalized least square  
 b) Restricted least squares  
 c) Weighted least square  
 d) Inverse least square
221. In the presence of heteroscedasticity the OLS estimates are  
 a) Biased  
 b) Unbiased  
 c) Efficient  
 d) Not efficient  
 e) Both b) & d)
222. Heteroscedasticity in your data is a problem because:  
 a) Ordinary OLS assumes that the data are homoscedastic and calculates the point estimates of regression parameters accordingly  
 b) Ordinary OLS assumes that the data are homoscedastic and calculates the standard error estimates of the parameters accordingly  
 c) It is contagious  
 d) It biases the parameter point estimates
223. The point at which homoscedasticity assumption is necessary in derivation of the variance of the slope estimate for simple regression is:  
 a) When we express the slope as a linear function of the observed values of the random variable Y, with coefficients constructed from the observed X values.  
 b) When we express the variance of a linear combination of random variables as the same linear combination of the variances of the individual random variables.  
 c) When we ignore the covariance terms in the formula for a linear combination of random variables  
 d) When we factor sigma-squared out of the summation in the numerator and then cancel one sum-of-squared-X-deviation from the numerator and denominator.
224. If we had some way to estimate the observation-specific error variances sigma-sub-i-squared, we could use these in the correct formulas for OLS parameter variances under heteroscedasticity to repair the standard error and t-ratios so they were not biased.  
 a) True  
 b) False  
 c) Uncertain
225. Contingent market scenarios sometimes used to:  
 a) Create independent variation in prices that does not exist in the actual data  
 b) Enlarge the number of degrees of freedom for estimation

- c) Minimize out-of-sample forecasting problems in fitted models  
 d) All of the above
226. Weighted least squares (WLS)  
 a) Is a special case of ordinary least squares (OLS)  
 b) Assigns lesser influence to small values of  $Y$  and greater influence to large values of  $Y$   
 c) Assigns lesser influence to observations where the data are more noisy, and more influence to observations where the data are less noisy  
 d) Requires additional variables in your OLS statement
227. Generalized least squares (GLS)  
 a) Encompasses WLS as a special case  
 b) Encompasses OLS as a special case  
 c) Renders unnecessary the assumption of no perfect multicollinearity  
 d) Both a) and b).
228. If your dataset has heteroscedasticity, but you completely ignore the problem and use a plain OLS command, you will  
 a) Get biased parameter estimates  
 b) Get parameter standard error estimates that could be either too large or too small  
 c) Get t-test statistics that make you overly optimistic about your parameters being statistically different from zero  
 d) Get t-test statistics that make it look like your parameters are not different from zero when in fact they are
229. If you use OLS under heteroscedasticity, but are careful to employ the correct formulas for the parameter variances, you will  
 a) Have the best estimates you can come up with, because OLS is the Best Linear Unbiased Estimator
- b) Have point estimates that are unbiased, but standard errors with unknown bias  
 c) Have point estimates that are unbiased, but parameter standard errors that are not as small as they could be under WLS, which is a more efficient estimator in the presence of heteroscedasticity  
 d) Be committing the crime of the century
230. A useful graphical method for detecting the presence of heteroscedasticity is  
 a) Plot  $Y$  against each  $X$  variable in turn.  
 b) Plot the residuals from a preliminary regression against the  $X$  variables, each in turn  
 c) Plot the squared residuals from a preliminary regression against the  $X$  variables, each in turn.  
 d) Plot the logarithm of the squared residuals from a preliminary regression against the  $X$  variables, each in turn.  
 e) Both c) and d)
231. It is always possible to prove that there is no heteroscedasticity in your model.  
 a) True  
 b) False  
 c) Uncertain
232. Heteroscedasticity is most commonly an affliction of  
 a) Time-series data, where observations pertain to the same entity over time.  
 b) Cross-sectional data, where observations pertain to different individual entities at the same point in time  
 c) Pooled time-series cross-section data, when you have data on several different cross-sectional entities at several different points in time, but not necessarily the same cross-sectional units being followed over time.

- d) Panel data, when you have repeated observations over time on the same set of cross-sectional entities  
e) b), c) and d)
233. The relationship among the successive values of the same variable is called  
a) Correlation  
b) Autoregressive  
c) Autocorrelation  
d) Markov process
234. When the assumption of CLRM,  $\text{Cov}(u_i, u_j) \neq 0$  violates the problem now arise is  
a) Multicollinearity  
b) Autocorrelation  
c) Simultaneous equation bias  
d) Heteroscedasticity
235. Which of the following is the source of autocorrelation  
a) Omission of explanatory variable  
b) Mis-specification of the mathematical model  
c) Both a) and b)  
d) None of these
236. The autocorrelation which arises due to omitting of explanatory variable is called  
a) Serial autocorrelation  
b) Quasi autocorrelation  
c) Spatial autocorrelation  
d) None of these
237. Which of the following test is used to detect autocorrelation  
a) Von-neumann ratio test  
b) Durbin Watson test  
c) Wallis test  
d) All of the above
238. The value of Durbin Watson d statistic lies in the interval  
a) (0, 1)  
b) (0, 4)  
c) (2, 4)  
d) (2, 4)
239. Durbin Watson test is used for  
a) Large sample  
b) Small sample
- c) First order auto-regressive scheme  
d) Both b) and c)
240. In the presence of autocorrelation the OLS estimates are  
a) linear  
b) unbiased  
c) BLUE  
d) Both a) and b)  
e) None of the above
241. Breusch Godfrey is a test of autocorrelation, use for  
a) First order autocorrelation  
b) Second order autocorrelation  
c) Higher order autocorrelation  
d) None of the above
242. Runs test is also called  
a) Wallis test  
b) Geary test  
c) BG test  
d) None of these
243. If we are dealing with Quarter time series data, which of the following test is used for testing autocorrelation  
a) Von-neumann ratio test  
b) Wallis test  
c) Runs test  
d) None of these
244. The limits of standardized residual for outlier identification is the interval  
a) (0, 1)  
b) (-2.5, 0)  
c) (-2.5, 2.5)  
d) (0, 2.5)  
e) None of these
245. When processes ought to behave randomly over time, then which of the following is used to identify nonrandom behavior  
a) Histogram  
b) Run chart  
c) Scatter diagram  
d) None of these
246. If error variances are small at low values of X, large for medium values of X and small again for large values of X,

- a)  $\log(\text{ehat-squared}) = b_1 + b_2 \cdot X$  will pick up the form of the heteroscedasticity  
 b)  $\text{ehat} = b_1 + b_2 \cdot X$  will reveal the form of the heteroscedasticity  
 c)  $\log(\text{ehat-squared}) = b_1 + b_2 \cdot X + b_3 \cdot X \cdot X$  will pick up the form of the heteroscedasticity  
 d)  $\log(\text{ehat-squared}) = b_1 + b_2 \cdot \log(X)$  will pick up the form of the heteroscedasticity
- 247 When you are working with aggregate time-series data, one of the first data pathologies you might suspect would be  
 a) Heteroscedasticity  
 b) Multicollinearity  
 c) Serially correlated errors  
 d) Endogeneity of regressors
- 248 The ordinary Durbin-Watson test statistic  
 a) Is approximately equal to  $2(1-\rho)$ , where  $\rho$  is the correlation between the sample regression function error terms, and ranges from 0 to 4 since  $\rho$  ranges from -1 to +1  
 b) Is more quickly calculated than the Exact Durbin-Watson test statistic  
 c) Sometimes leads to ambiguous conclusions concerning the presence or absence of serial correlation in regression errors  
 d) Is derived under the assumption that the null hypothesis--zero error correlation is true  
 e) All of the above
- 249 When you referee a research proposal, the objective is to make yourself look good and the other person look stupid.  
 a) True  
 b) False  
 c) Uncertain
- 250 When a referee completely misinterprets the whole point of your research proposal, you should  
 a) Consider him/her a complete turkey and refuse to submit anything else for review  
 b) Cast aspersions on the footwear of his/her mother  
 c) Assume the editor/funding agency will appreciate your creativity and ignore the referee comments, giving them the zero weight they deserve  
 d) Acknowledge that misinterpretation can be a direct consequence of a poorly articulated argument
- 251 It is possible to fix up the OLS parameter variance calculations if you know the true values of the  $\rho$  parameter. Once fixed,  
 a) You will have a best linear unbiased estimator  
 b) The parameter standard errors will be larger than they could be under an alternative estimator, which is another in the family of Generalized Least Squares (GLS) estimators  
 c) The corrected OLS estimator will produce the same point estimate and standard error as the preferred GLS estimator  
 d) All of the above
- 252 If we can come up with a way of estimating the unknown population error correlation parameter  $\rho$ , we can achieve the goal of purging our AR (1) error term of its unwelcome component by  
 a) Transforming all of our data by multiplying through all variables by  $\rho$   
 b) Transforming all of our data by subtracting  $\rho$  times the value of the same variable from the previous period  
 c) Transforming all of our data by dividing through all variables by  $\rho$ , so that observations with high values of  $\rho$  are given low weight and observations with low values of  $\rho$  are given higher weight

(since we want to put the most emphasis on observations that are less contaminated by serial correlation)

- d) Dropping observations with Y values that is very similar to the value of Y in the adjacent period

253 If your dependent variable sometimes takes on a value of zero, what should you do if you want to use a log-linear or log-log specification?

- a) No problem go ahead
- b) You cannot do it, since the log of zero is negative infinity
- c) Just add 1 to all values of Y and then take its log and use in your regression
- d) Add 0.001 to just the zero values of Y and then take its log and use in your regression
- e) Contemplate using a generalization of regression designed specifically for data with non-negative values that can be zero

254 What is errors-in-variables attenuation?

- a) You need to stop and think before using X variables that are measured with error
- b) As a rule of thumb slope coefficients are generally pushed closer to zero (whether they are positive or negative) if your X variable is measured with (random) errors, so that the apparent effect of X on Y (whether positive or negative) appears less than it actually is
- c) You need to attenuate (reduce) your sample to eliminate observations on X that are measured with error
- d) If you have included the wrong variables in your regression model, the amount of regression output produced is less than if you have used the right variables

255 What is usually the reason for including in your regression an interaction term  $XZ (=X \cdot Z)$  between two continuous explanatory variables X and Z?

- a) You suspect that the effect of X on Y depends on the level of Z and that the effect of Z on Y depends on X. This is equivalent to a suspicion that the derivative of Y with respect to X is not constant (as in a straight linear-in-variables model) but gets larger or smaller with the value of Z. Likewise for the derivative of Y with respect to Z depending on X.
- b) Your R-squared from a linear model is abysmally small and you have to do something
- c) You suspect that the effect of X on Y depends on the level of X
- d) You suspect that the effect of X on Y depends on the level of Y

256 If it is expedient for you to have a fitted model wherein on elasticity of Y with respect to X applies for all values of X you might consider, your first choice of functional form would probably be

- a) A linear model
- b) A log-linear model
- c) A log-log model
- d) A linear-logarithmic model

257 You probably want to avoid log-log specifications if

- a) It is possible for Y or X to take on zero or negative values
- b) The elasticity of Y with respect to X is one
- c) You have values of Y which are large and values of X which are small
- d) You cannot remember the definition of a logarithm

- 258 Reciprocal models (where  $Y$  is proposed to depend not on  $X$ , but on  $1/X$ ) are popular when
- It is convenient to have a single elasticity estimate for the entire range of values for  $X$
  - The fitted value of  $Y$ , as a function of  $X$ , approaches a particular value asymptotically as  $X$  increases
  - The fitted value of  $Y$  approaches zero as  $X$  approaches zero
  - The derivative of  $Y$  with respect to  $X$  is constant throughout the data
- 259 The term kitchen-sink regression is used:
- Widely, to describe the applied econometric work of researchers at plumbing hardware suppliers such as Moen, Price-P fister, American Standard, etc.
  - Glowingly, to show approval of researchers who have been meticulous about collecting and using virtually every explanatory variable that could even remotely be relevant to determining the level of the dependent variable.
  - Disparagingly, to describe a model that includes, indiscriminately, every explanatory variable that the researcher has available, without recourse to economic theory to differentiate between sensible variables and silly variables.
  - To describe dinner parties where the guests all eventually gravitate to the kitchen, to converse with the hosts while the dishes are being washed.
- 260 What character traits may be the most important in an applied econometrician?
- Brilliant mathematical skills
  - Persistence, motivation in the presence of delayed or infrequent gratification, and a high tolerance for ambiguity
  - Empathy and sensitivity, leading to an intuitive understanding of what motivates the relationships in the data being modeled
  - a gift for schmoozing with economic theorists
- 261 In comparing Maximum Likelihood Estimation (MLE) of unknown parameters with Ordinary Least Squares (OLS) estimation of unknown parameters, which of the following statements is false?
- OLS chooses the best-fitting line by minimizing the sum of squared vertical deviations of each  $Y$  value from the line that is chosen
  - MLE chooses the best-fitting line by maximizing the logarithm of the joint probability of observing the  $n$  independent observations on  $Y$  in the sample
  - If we assume that the errors in a typical regression model are normally distributed, then the best-fitting line by MLE will have exactly the same intercept and slope as the best-fitting line by OLS
  - MLE can be considered as a special case of OLS methods
- 262 In the early days of empirical research in economics, researchers who needed to explain discrete outcomes tended to use:
- A so-called Linear Probability Model (LPM), which was just OLS applied to the  $(0,1)$  outcomes, with some fix-ups for heteroscedasticity and unacceptable negative or larger-than-unity fitted probabilities
  - Probit models, since these models were most compatible with the convenient normality assumption for the errors in regular OLS models
  - Logit models, since these models allowed the necessary cumulative probability density values to be calculated simply as a ratio of exponentiated quantities, rather than as a non-closed form integral
  - An abacus

263. Once a sensible specification for a probit or logit model has been estimated,

- a) We can interpret the estimated slope coefficients as the change in the probability of the 1 outcome for a one-unit change in that  $X$  variable
- b) We can interpret the estimated slope coefficients as the change in the propensity to choose the 1 outcome. The fitted probability of choosing the 1 outcome is the cumulative standard normal (or standard logistic) probability density evaluated at the fitted value of this propensity for each individual. Thus the effect of a one-unit change in  $X$  on the choice probability depends upon all of the characteristics ( $X$ -variable values) for a given individual as well as on the slope coefficient on the one  $X$  variable in question
- c) The derivative of the probability of choosing the 1 outcome, with respect to a particular explanatory variable,  $X$ , cannot be calculated
- d) We can be sure we have no multicollinearity problems, no heteroscedasticity and no serial or spatial correlation in our data

264. Which phonetic pronunciations of the names of the two main discrete-choice models are correct?

- a) Probe-it and low-jit
- b) Log-it and prob-it
- c) Pro-bite and low-jite
- d) Hi-jak and lo-jak

			ANSWERS		
1	a	2	c	3	d
5	a	6	c	7	c
9	c	10	a	11	e
13	b	14	e	15	d
17	a	18	a	19	c
21	c	22	d	23	a
25	b	26	d	27	c
29	c	30	d	31	a
33		34	a	35	d
37	d	38	a	39	c
41	b	42	d	43	d
45	e	46	b	47	b
49	a	50	e	51	e
53	c	54	b	55	e
57	a	58	d	59	c
61	d	62	c	63	d
65	c	66	c	67	b
69	a	70	a	71	b
73	c	74	b	75	b
77	c	78	a	79	a
81	c	82	a	83	a
85	a	86	a	87	c
89	a	90	a	91	a
93	a	94	b	95	b
97	c	98	c	99	a
101	b	102	c	103	b
105	a	106	d	107	b
109	b	110	c	111	a
113	b	114	d	115	a
117	d	118	a	119	a
121	c	122	c	123	a
125	c	126	a	127	b
129	c	130	c	131	d
133	c	134	a	135	b
137	a	138	c	139	b
141	d	142	a	143	d
145	b	146	c	147	b
149	b	150	e	151	d
					152 c



# TIME SERIES AND FORECASTING

1. A collection of observations made sequentially in time, is called
  - Stochastic process
  - Random Process
  - Time Series
  - None of these
2. A computer analysis reveals that the best-fitting trend model is  $Y_t = 4.12e^{0.087t}$ . The trend was fitted using year-end common stock prices for Melodic Kortholt Outlet for the last six years. The  $R^2$  is 0.8571. Which conclusion is not correct?
  - The absolute annual growth (in dollars per share) is increasing.
  - Few investments could match the astounding growth rate.
  - At the end of year 3 the stock price would be nearly \$80.
  - The exponential model is inappropriate for financial data
3. If we fit a linear trend to 10 observations on time series data that are growing exponentially, then it is most likely that
  - The fitted trend will be too high at  $t = 1$  and  $t = 10$ .
  - The fitted trend will be too low in the middle.
  - The forecasts (if extrapolated) will be too low.
  - None of these
4. Suppose the estimated quadratic model  $Y_t = 500 + 20t - t^2$  is the best-fitting trend of sales of XYZ Inc. using data for the past twenty years ( $t = 1, 2, \dots, 20$ ). Which statement is incorrect?
  - Sales are increasing by about 20 units per year.
  - The turning point would be in period 10.
  - Latest year sales are no better than in year zero.
  - The trend was higher in year 10 than in year 20.
5. Which is a time series?
  - The M1 components of the U.S. money supply ( $n=20$  quarters).
  - The unemployment rates for the U.S. states ( $n = 50$  states).
  - The Gross Domestic Product for the E.U. members ( $n = 15$  nations).
  - The inflation rate for housing in U.S. metropolitan areas ( $n = 46$  cities).
6. Which is not an additive component of a time series?
  - Trend.
  - Seasonal.
  - Irregular.
  - Periodicity
7. The fitted annual sales trend is  $Y_t = 187.3e^{0.047t}$ . On average, sales are
  - Rising by an increasing absolute amount each year.
  - Rising by a declining absolute amount each year.
  - Falling by a declining absolute amount each year.
  - Falling by an increasing absolute amount each year
8. The fitted annual sales trend is  $Y_t = 187.3e^{0.047t}$ . The sales forecast for year 5 would be
 

a) 236.9	b) 178.7
c) 168.2	d) 148.1

9. Concerning a multiplicative seasonal index for monthly data, which statement is incorrect?  
 a) An index value of 1.000 indicates no seasonal deviation from trend  
 b) The estimated indexes are adjusted so they always sum to 12  
 c) To make forecasts, we divide the projected trend by each month's index  
 d) They are obtained by the process called decomposition of a time series
10. Forecasts  
 a) Become more accurate with longer time horizons  
 b) Are rarely perfect  
 c) Are more accurate for individual items than for groups of items  
 d) All of the above  
 e) None of the above
11. One purpose of short-range forecasts is to determine  
 a) Production planning  
 b) Inventory budgets  
 c) Research and development plans  
 d) Facility location  
 e) Job assignments
12. Forecasts are usually classified by time horizon into three categories  
 a) Short-range, medium-range, and long-range  
 b) Finance/accounting, marketing, and operations  
 c) Strategic, tactical, and operational  
 d) Exponential smoothing, regression, and time series  
 e) Departmental, organizational, and industrial
13. A forecast with a time horizon of about 3 months to 3 years is typically called  
 a)  
 a) Long-range forecast  
 b) Medium-range forecast  
 c) Short-range forecast  
 d) Weather forecast  
 e) Strategic forecast
14. Forecasts used for new product planning, capital expenditures, facility location or expansion, and R&D typically utilize a  
 a) Short-range time horizon  
 b) Medium-range time horizon  
 c) Long-range time horizon  
 d) Naïve method because there is no data history  
 e) All of the above
15. The three major types of forecasts used by business organizations are  
 a) Strategic, tactical, and operational  
 b) Economic, technological, and demand  
 c) Exponential smoothing, Delphi, and regression  
 d) Causal, time-series, and seasonal  
 e) Departmental, organizational, and territorial
16. Which of the following is not a step in the forecasting process?  
 a) Determine the use of the forecast  
 b) Eliminate any assumptions  
 c) Determine the time horizon  
 d) Select a forecasting model(s)  
 e) Validate and implement the results
17. The two general approaches to forecasting are  
 a) Qualitative and quantitative  
 b) Mathematical and statistical  
 c) Judgmental and qualitative  
 d) Historical and associative  
 e) Judgmental and associative
18. Which of the following uses three types of participants: decision makers, staff personnel, and respondents?  
 a) Executive opinions  
 b) Sales force composites  
 c) The Delphi method  
 d) Consumer surveys  
 e) Time series analysis
19. Which of the following is not a type of qualitative forecasting?  
 a) Executive opinions  
 b) Sales force composites  
 c) Consumer surveys  
 d) The Delphi method  
 e) Moving average

**09 Computation of statistics**

- 20 The forecasting model that pools the opinions of a group of experts or managers is known as the  
a) Sales force composition model  
b) Multiple regression  
c) Jury of executive opinion model  
d) Consumer market survey model  
e) Management coefficients model
- 21 Which of the following techniques uses variables such as price and promotional expenditures which are related to product demand, to predict demand?  
a) Associative models  
b) Exponential smoothing  
c) Weighted moving average  
d) Simple moving average  
e) Time series
- 22 Which of the following statements about time-series forecasting is true?  
a) It is based on the assumption that future demand will be the same as past demand  
b) It makes extensive use of the data collected in the qualitative approach  
c) The analysis of past demand helps predict future demand  
d) Because it accounts for trends, cycles, and seasonal patterns, it is more powerful than causal forecasting
- 23 Time-series data may exhibit which of the following behaviors?  
a) Trend  
b) Random variations  
c) Seasonality  
d) Cycles  
e) All of the above
- 24 Gradual, long-term movement in time-series data is called  
a) Seasonal variation  
b) Cycles  
c) Trends  
d) Exponential variation  
e) Random variation

25 Which of the following is not primary in time series?

- a) Trend  
b) Seasonality  
c) Cyclical variations  
d) Random variations

26 The fundamental difference between cycles and seasonality is the

- a) Duration of the repeating pattern  
b) Magnitude of the variation  
c) Ability to attribute the pattern to a cause  
d) All of the above

27 In time series, which of the following cannot be predicted?

- a) Large increases in demand  
b) Technological trends  
c) Seasonal fluctuations  
d) Random fluctuations  
e) Large decreases in demand

28 What is the approximate forecast for May using a four-month moving average?

Nov	Dec	Jan	Feb	Mar	April
39	36	40	42	48	49
a) 38					
b) 42					
c) 43					
d) 44					
e) 47					

29 Which time series model below assumes that demand in the next period will be equal to the most recent period's demand?

- a) Naive approach  
b) Moving average approach  
c) Weighted moving average approach  
d) Exponential smoothing approach  
e) None of the above

30 Which of the following is not a characteristic of simple moving averages?

- a) It smoothes random variations in the data  
b) It has minimal data storage requirements

- c) It weights each historical value equally  
 d) It smoothes real variations in the data
- 31 A six-month moving average forecast is better than a three-month moving average forecast if demand  
 a) Is rather stable  
 b) Has been changing due to recent promotional efforts  
 c) Follows a downward trend  
 d) Follows an upward trend
- 32 Increasing the number of periods in a moving average will accomplish greater smoothing, but at the expense of  
 a) Manager understanding  
 b) Accuracy  
 c) Stability  
 d) Responsiveness to changes
- 33 Which of the following statements comparing the weighted moving average technique and exponential smoothing is true?  
 a) Exponential smoothing is more easily used in combination with the Delphi method.  
 b) More emphasis can be placed on recent values using the weighted moving average.  
 c) Exponential smoothing is considerably more difficult to implement on a computer.  
 d) Exponential smoothing typically requires less record-keeping of past data.
- 34 Which time series model uses past forecasts and past demand data to generate a new forecast?  
 a) Naïve  
 b) Moving average  
 c) Weighted moving average  
 d) Exponential smoothing
- 35 Which is not a characteristic of exponential smoothing?  
 a) Smoothes random variations in the data  
 b) Easily altered weighting scheme
- c) Weights each historical value equally  
 d) Has minimal data storage requirements
- 36 Which of the following smoothing constants would make an exponential smoothing forecast equivalent to a naive forecast?  
 a) 0  
 b) 1 divided by the number of periods  
 c) 0.5  
 d) 1.0
- 37 Given an actual demand of 103, a previous forecast value of 99, and an alpha of .4, the exponential smoothing forecast for the next period would be  
 a) 94.6  
 b) 97.4  
 c) 100.6  
 d) 101.6  
 e) 103.0
- 38 A forecast based on the previous forecast plus a percentage of the forecast error is a  
 a) Qualitative forecast  
 b) Naïve forecast  
 c) Moving average forecast  
 d) Weighted moving average forecast  
 e) Exponentially smoothed forecast
- 39 Given an actual demand of 61, a previous forecast of 58, and  $\alpha = 0.30$ , what would the forecast for the next period be using simple exponential smoothing?  
 a) 45.5  
 b) 57.1  
 c) 58.9  
 d) 61.0  
 e) 65.5
- 40 Which of the following values of alpha would cause exponential smoothing to respond the most quickly to forecast errors?  
 a) 0.00  
 b) 0.10  
 c) 0.20  
 d) 0.40

41. The primary purpose of the mean absolute deviation (MAD) in forecasting is to  
 a) Estimate the trend line  
 b) Eliminate forecast errors  
 c) Measure forecast accuracy  
 d) Seasonally adjust the forecast
42. Given forecast errors of -1, 4, 8, and -3, what is the mean absolute deviation?  
 a) 2  
 b) 3  
 c) 4  
 d) 8  
 e) 16
43. For a given product demand, the time-series trend equation is  $25.3 + 2.1X$ . What is your forecast of demand for period 7?  
 a) 23.2  
 b) 25.3  
 c) 27.4  
 d) 40
44. In trend-adjusted exponential smoothing, the Forecast Including Trend (FIT) consists of,  
 a) An exponentially smoothed forecast and an estimated trend value  
 b) An exponentially smoothed forecast and a smoothed trend factor  
 c) The old forecast adjusted by a trend factor  
 d) The old forecast and a smoothed trend factor
45. Which of the following is true regarding the two smoothing constants of the Forecast Including Trend (FIT) model?  
 a) Their values are determined independently.  
 b) They are called alpha and beta.  
 c) Alpha is always smaller than beta.  
 d) All of the above are true.
46. The tracking signal is  
 a) The standard error of the estimate
- b) The running sum of forecast errors (RSFE)  
 c) The mean absolute deviation (MAD)  
 d) The ratio RSFE/MAD
47. A stochastic process which has same mean, variance and covariance over time is called  
 a) Random walk model  
 b) Weakly stationary  
 c) Trend Stationary  
 d) Difference Stationary  
 e) None of these
48. A graph of autocorrelation is called  
 a) Histogram  
 b) Correlogram  
 c) Historogram  
 d) Ogive
49. When the time series contains unit root problem then autocorrelation can be tested by  
 a) Durbin waston test  
 b) Runs test  
 c) Dickey-Fuller test  
 d) None of these
50. Alexis Popov, CFA, wants to estimate how sales have grown from one quarter to the next on average. The most direct way for Popov to estimate this would be:  
 a) An AR(1) model.  
 b) An AR(1) model with a seasonal lag.  
 c) A linear trend model.  
 d) None of the above
51. Which of the following statements regarding seasonality is false?  
 a) Not correcting for seasonality when, in fact, seasonality exists in the time series results in a violation of an assumption of linear regression.  
 b) A time series that is first differenced can be adjusted for seasonality by incorporating the first-differenced value for the previous year's corresponding

- c) The presence of seasonality makes it impossible to forecast using a time-series model.  
d) None of the above
52. Barry Phillips, CFA, is analyzing quarterly data. He has estimated an AR(1) relationship ( $x_t = b_0 + b_1 x_{t-1} + e_t$ ) and wants to test for seasonality. To do this he would want to see if which of the following statistics is significantly different from zero?  
a) Correlation ( $e_t, e_{t-4}$ ).  
b) Correlation ( $e_t, e_{t-1}$ ).  
c) Correlation ( $e_t, e_{t-5}$ ).  
d) None of the above
53. Which of the following statements regarding seasonality is false?  
a) The presence of seasonality makes it impossible to forecast using a time-series model.  
b) Not correcting for seasonality when, in fact, seasonality exists in the time series results in a violation of an assumption of linear regression.  
c) A time series that is first differenced can be adjusted for seasonality by incorporating the first-differenced value for the previous year's corresponding period.  
d) None of the above
54. Which of the following is a seasonally adjusted model?  
a)  $Sales_t = b_0 + b_1 Sales_{t-1} + b_2 Sales_{t-2} + e_t$ .  
b)  $(Sales_t - Sales_{t-1}) = b_0 + b_1 (Sales_{t-1} - Sales_{t-2}) + b_2 (Sales_{t-4} - Sales_{t-5}) + e_t$ .  
c)  $Sales_t = b_1 Sales_{t-1} + e_t$ .  
d) None of the above
55. Which of the following is least likely a consequence of a model containing ARCH(1) errors? The:  
a) Variance of the errors can be predicted.  
b) Regression parameters will be incorrect.
- c) Model's specification can be corrected by adding an additional lag variable.  
d) None of the above
56. Suppose you estimate the following model of residuals from an autoregressive model:  
 $e_t^2 = 0.25 + 0.6e_{t-1}^2 + \mu_t$ , where  $\mu_t = \hat{\epsilon}_t$ . If the residual at time t is 0.9, the forecasted variance for time t+1 is:  
a) 0.790  
b) 0.736  
c) 0.850  
d) None of the above
57. One choice a researcher can use to test for nonstationarity is to use a  
a) Breusch-Pagan test, which uses a modified t-statistic.  
b) Dickey-Fuller test, which uses a modified t-statistic.  
c) Dickey-Fuller test, which uses a modified  $\chi^2$  statistic.  
d) None of the above
58. Alexis Popov, CFA, is analyzing monthly data. Popov has estimated the model  $x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + e_t$ . The researcher finds that the residuals have a significant ARCH process. The best solution to this is to:  
a) Re-estimate the model using only an AR (1) specification.  
b) Re-estimate the model with generalized least squares.  
c) Re-estimate the model using a seasonal lag.  
d) None of the above
59. The primary concern when deciding upon a time series sample period is which of the following factors?  
a) The length of the sample time period.  
b) Current underlying economic and market conditions.  
c) The total number of observations.  
d) None of the above
60. Which of the following statements regarding the instability of time-series

41. The primary purpose of the mean absolute deviation (MAD) in forecasting is to  
 a) Estimate the trend line  
 b) Eliminate forecast errors  
 c) Measure forecast accuracy  
 d) Seasonally adjust the forecast
42. Given forecast errors of -1, 4, 8, and -3, what is the mean absolute deviation?  
 a) 2  
 b) 3  
 c) 4  
 d) 8  
 e) 16
43. For a given product demand, the time-series trend equation is  $25.3 + 2.1 X$ . What is your forecast of demand for period 7?  
 a) 23.2  
 b) 25.3  
 c) 27.4  
 d) 40
44. In trend-adjusted exponential smoothing, the Forecast Including Trend (FIT) consists of,  
 a) An exponentially smoothed forecast and an estimated trend value  
 b) An exponentially smoothed forecast and a smoothed trend factor  
 c) The old forecast adjusted by a trend factor  
 d) The old forecast and a smoothed trend factor
45. Which of the following is true regarding the two smoothing constants of the Forecast Including Trend (FIT) model?  
 a) Their values are determined independently.  
 b) They are called alpha and beta, Producer's Risk and Consumer's Risk.  
 c) Alpha is always smaller than beta.  
 d) All of the above are true
46. The tracking signal is  
 a) The standard error of the estimate
- b) The running sum of forecast errors (RSFE)  
 c) The mean absolute deviation (MAD)  
 d) The ratio RSFE/MAD
47. A stochastic process which has constant mean, variance and covariance over time is called  
 a) Random walk model  
 b) Weakly stationary  
 c) Trend Stationary  
 d) Difference Stationary  
 e) None of these
48. A graph of autocorrelation is called  
 a) Histogram  
 b) Correlogram  
 c) Historogram  
 d) Ogive
49. When the time series contains root problem then autocorrelation can be tested by  
 a) Durbin Watson test  
 b) Runs test  
 c) Dickey-Fuller test  
 d) None of these
50. Alexis Popov, CFA, wants to estimate how sales have grown from one quarter to the next on average. The most direct way for Popov to estimate this would be:  
 a) An AR(1) model.  
 b) An AR(1) model with a seasonal lag.  
 c) A linear trend model.  
 d) None of the above
51. Which of the following statements regarding seasonality is false?  
 a) Not correcting for seasonality when, in fact, seasonality exists in the time series results in a violation of an assumption of linear regression.  
 b) A time series that is first differenced can be adjusted for seasonality by incorporating the first-differenced value for the previous year's corresponding period

- c) The presence of seasonality makes it impossible to forecast using a time-series model.
- d) None of the above
52. Barry Phillips, CFA, is analyzing quarterly data. He has estimated an AR(1) relationship ( $x_t = b_0 + b_1 x_{t-1} + e_t$ ) and wants to test for seasonality. To do this he would want to see if which of the following statistics is significantly different from zero?
- Correlation ( $e_t, e_{t-4}$ ).
  - Correlation ( $e_t, e_{t-1}$ ).
  - Correlation ( $e_t, e_{t-5}$ ).
  - None of the above
53. Which of the following statements regarding seasonality is false?
- The presence of seasonality makes it impossible to forecast using a time-series model.
  - Not correcting for seasonality when, in fact, seasonality exists in the time series results in a violation of an assumption of linear regression.
  - A time series that is first differenced can be adjusted for seasonality by incorporating the first-differenced value for the previous year's corresponding period.
  - None of the above
54. Which of the following is a seasonally adjusted model?
- $Sales_t = b_0 + b_1 Sales_{t-1} + b_2 Sales_{t-2} + e_t$ .
  - $(Sales_t - Sales_{t-1}) = b_0 + b_1 (Sales_{t-1} - Sales_{t-2}) + b_2 (Sales_{t-4} - Sales_{t-5}) + e_t$ .
  - $Sales_t = b_1 Sales_{t-1} + e_t$ .
  - None of the above
55. Which of the following is least likely a consequence of a model containing ARCH(1) errors? The:
- Variance of the errors can be predicted.
  - Regression parameters will be incorrect.
56. Suppose you estimate the following model of residuals from an autoregressive model:  
 $e_t = 0.25 + 0.6e_{t-1} + \mu_t$ , where  $\mu_t \sim N(0, \sigma^2)$ . If the residual at time  $t$  is 0.9, the forecasted variance for time  $t+1$  is
- 0.790
  - 0.736
  - 0.850
  - None of the above
57. One choice a researcher can use to test for nonstationarity is to use a
- Breusch-Pagan test, which uses a modified t-statistic.
  - Dickey-Fuller test, which uses a modified t-statistic.
  - Dickey-Fuller test, which uses a modified  $\chi^2$  statistic.
  - None of the above
58. Alexis Popov, CFA, is analyzing monthly data. Popov has estimated the model  $x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + e_t$ . The researcher finds that the residuals have a significant ARCH process. The best solution to this is to:
- Re-estimate the model using only an AR (1) specification.
  - Re-estimate the model with generalized least squares.
  - Re-estimate the model using a seasonal lag.
  - None of the above
59. The primary concern when deciding upon a time series sample period is which of the following factors?
- The length of the sample time period.
  - Current underlying economic and market conditions.
  - The total number of observations
  - None of the above
60. Which of the following statements regarding the instability of time-series

- models is most accurate? Models estimated with
- Shorter time series are usually more stable than those with longer time series.
  - A greater number of independent variables are usually more stable than those with a smaller number.
  - Longer time series are usually more stable than those with shorter time series.
  - None of the above
61. The main reason why financial and time series intrinsically exhibit some form of nonstationarity is that
- Most financial and economic relationships are dynamic and the estimated regression coefficients can vary greatly between periods.
  - Serial correlation, a contributing factor to nonstationarity, is always present to a certain degree in most financial and time series.
  - Most financial and time series have a natural tendency to revert toward their means.
  - None of the above
62. A time series that has a unit root can be transformed into a time series without a unit root through
- Calculating moving average of the residuals.
  - Mean reversion.
  - First differencing.
  - None of the above
63. An AR (1) autoregressive time series model
- Cannot be used to test for a unit root.
  - Can be used to test for a unit root, which exists if the slope coefficient is less than one.
  - Can be used to test for a unit root, which exists if the slope coefficient equals one.
  - None of the above
64. Which of the following statements regarding unit roots in a time series is false?
- Even if a time series has a unit root, the predictions from the estimated model are valid.
  - A time series that is a random walk has a unit root.
  - A time series with a unit root is not covariance stationary.
  - None of the above
65. Given an AR (1) process represented by  $x_{t+1} = b_0 + b_1 x_t + e_t$ , the process would not be a random walk if:
- $b_1 = 1$ .
  - $E(e_t) = 0$ .
  - The long run mean is  $b_0 + b_1$ .
  - None of the above
66. Which of the following statements regarding time series analysis is FALSE?
- We cannot use an AR (1) model on a time series that consists of a random walk.
  - If a time series is a random walk, first differencing will result in covariance stationarity.
  - An autoregressive model with two lags is equivalent to a moving-average model with two lags.
  - None of the above
67. A time series  $x$  that is a random walk with a drift is best described as:
- $x_t = b_0 + b_1 x_{t-1}$ .
  - $x_t = b_0 + b_1 x_{t-1} + e_t$ .
  - $x_t = x_{t-1} + e_t$ .
  - None of the above
68. Which of the following statements regarding an out-of-sample forecast is least accurate?
- There is more error associated with out-of-sample forecasts, as compared to in-sample forecasts.
  - Forecasting is not possible for autoregressive models with more than two lags.

- c) Out-of-sample forecasts are of more importance than in-sample forecasts to the analyst using an estimated time-series model.  
 d) None of the above

69 The procedure for determining the structure of an autoregressive model is:

- a) Estimate an autoregressive model (for example, an AR (1) model), calculate the autocorrelations for the model's residuals, test whether the autocorrelations are different from zero, and add an AR lag for each significant autocorrelation.  
 b) Estimate an autoregressive model (e.g., an AR (1) model), calculate the autocorrelations for the model's residuals, test whether the autocorrelations are different from zero, and revise the model if there are significant autocorrelations.  
 c) Test autocorrelations of the residuals for a simple trend model, and specify the number of significant lags.  
 d) None of the above

70 The model  $x_t = b_0 + b_1 x_{t-1} + b_2 x_{t-2} + b_3 x_{t-3} + b_4 x_{t-4} + e_t$  is

- a) An autoregressive conditional heteroskedastic model, ARCH.  
 b) An autoregressive model, AR (4).  
 c) A moving average model, MA(4).  
 d) None of the above

71 To qualify as a covariance stationary process, which of the following does not have to be true?

- a) Covariance ( $x_t, x_{t-1}$ ) = Covariance ( $x_t, x_{t-2}$ )  
 b)  $E[x_t] = E[x_{t+1}]$   
 c) Covariance ( $x_t, x_{t-2}$ ) = Covariance ( $x_t, x_{t+2}$ )  
 d) None of the above

72 Which of the following is NOT a requirement for a series to be covariance stationary? The:

- a) Expected value of the time series is constant over time.  
 b) Covariance of the time series with itself (lead or lag) must be constant.  
 c) Time series must have a positive trend.  
 d) None of the above

73 Which of the following statements regarding covariance stationarity is true?

- a) A time series that is covariance stationary may have residuals whose mean changes over time.  
 b) A time series may be both covariance stationary and have heteroskedastic residuals.  
 c) The estimation results of a time series that is not covariance stationary are meaningless.  
 d) None of the above

74 Modeling the trend in a time series of a variable that grows at a constant rate with continuous compounding is best done with:

- a) A log-linear transformation of the time series.  
 b) A moving average model.  
 c) Simple linear regression.  
 d) None of the above

## ANSWERS

- 1 c 11 e 21 a 31 a 41 c 51 c 61 a 71 a  
 2 d 12 a 22 c 32 b 42 c 52 a 62 c 72 c  
 3 c 13 b 23 e 33 d 43 d 53 a 63 c 73 c  
 4 a 14 c 24 c 34 d 44 b 54 b 64 a 74 a  
 5 a 15 b 25 b 35 c 45 a 55 c 65 c 75  
 6 d 16 b 26 a 36 d 46 d 56 b 66 c 76  
 7 c 17 a 27 d 37 c 47 b 57 b 67 b 77  
 8 d 18 c 28 d 38 e 48 b 58 b 68 b 78  
 9 c 19 e 29 a 39 c 49 c 59 b 69 b 79  
 10 b 20 c 30 b 40 d 50 c 60 a 70 b 80

# SURVEY SAMPLING AND RESEARCH METHODOLOGY

1. Mr. Javed is writing her daily observations of a student and writes, without interpretation, that the student is not completing the class work and is constantly speaking out of turn. Which of the following objectives does she appear to be using?
  - Prediction
  - Description
  - Explanation
  - Exploration
  - None of the above
2. Which of the following can best be described as a categorical variable?
  - Age
  - Annual income
  - Grade point average
  - Religion
3. The strongest evidence for causality comes from which of the following research methods?
  - Experimental
  - Causal-comparative
  - Correlational
  - Ethnography
4. A researcher studies the relation between early reading and later school achievement. She decides that a potentially extraneous variable in the relationship is IQ. In developing her groups for her study, she pairs each child who was an early reader with a child of the same IQ level who was not an early reader. The control technique she used was:
  - Holding the extraneous variable constant
  - Statistical control
  - Matching
  - Random assignment
5. The directors of a graduate program in educational research wish to see what types of jobs their graduates take after they finish their program. They randomly sample students from the program and have them fill out questionnaires with items asking about the types of jobs they have had. They also are asked to describe the roles they play in their current positions. This project is best described as having what kind of objective:
  - Descriptive
  - Predictive
  - Explanatory
6. When research is done to test hypotheses and theories about how and why phenomena operate as they do, then the primary purpose of such research is:
  - Descriptive
  - Predictive
  - Explanatory
7. The variable the researcher matches to eliminate it as an alternative explanation is called a \_\_\_\_\_ variable.
  - Matching
  - Independent
  - Dependent
  - Partial
8. When a researcher starts with the dependent variable and moves backwards, it is called \_\_\_\_\_
  - Predictive research
  - Retrospective research
  - Exploratory research
  - Descriptive research

9. The method of working multiple hypotheses refers to a technique for identifying rival explanations.
- True
  - False
10. Which one of the following is not a step in non experimental research?
- Determine research problem and hypotheses
  - Analyze data
  - Interpret results
  - All are steps
11. If a research finding is statistically significant, then
- The observed result is probably not due to chance
  - The observed result cannot possibly be due to chance
  - The observed result is probably a chance result
  - The null hypothesis of "no relationship" is probably true
12. Which of the following is/are necessary conditions for causation?
- The relationship condition
  - The temporal antecedence condition
  - The lack of alternative explanation condition
  - All of the above
13. Which of the following independent variables cannot be manipulated in a research study?
- Gender
  - Ethnicity
  - Intelligence and other traits
  - None of the above can be manipulated in a research study
14. \_\_\_\_\_ is a form of explanatory research in which the researcher develops a theoretical model and empirically tests the model to determine how well the model fits the data.
- Causal modeling
  - Predictive research
  - Descriptive research
  - Exploratory research
15. Mrs. Smith is writing her daily observations of a student and writes, without interpretation, that the student is not completing the class work and is constantly speaking out of turn. Which of the following objectives does she appear to be using?
- Prediction
  - Description
  - Explanation
  - Exploration
16. Which of the following is a form of research typically conducted by teachers, counselors, and other professionals to answer questions they have and to specifically help them solve local problems?
- Action research
  - Basic research
  - Predictive research
  - Oriental research
17. How much confidence should you place in a single research study?
- You should completely trust a single research study.
  - You should trust research findings after different researchers have found the same findings
  - Neither a) nor b)
  - both a) and b)
18. The development of a solid foundation of reliable knowledge typically is built from which type of research?
- Basic research
  - Action research
  - Evaluation research
  - Oriental research
19. Which form of reasoning is the process of drawing a specific conclusion from a set of premises?
- Rationalism
  - Deductive reasoning
  - Inductive reasoning
  - Probabilistic
20. The idea that when selecting between two different theories with equal explanatory value, one should select

*Take Note*  
96 Compendium of Statistics

- the theory that is the most simple, concise, and succinct is known as
- Criterion of falsifiability
  - Critical theory
  - Guide of simplicity
  - Rule of parsimony
21. Research that is done to examine the findings of someone else using the "same variables but different people" is which of the following?
- Exploration
  - Hypothesis
  - ✓ Replication
  - Empiricism
22. \_\_\_\_\_ is the idea that knowledge comes from experience.
- Rationalism
  - Deductive reasoning
  - Logic
  - Empiricism
23. According to research five key objectives of science are:
- Prediction, summary, conclusion, explanation, description
  - Influence, prediction, questions, exploration, answers
  - Exploration, description, explanation, prediction, influence
  - Questions, answers, prediction, explanation, summary
24. A researcher designs an experiment to test how variables interact to influence how well children learn spelling words. In this case, the main purpose of the study was:
- ✓ Explanation
  - Description
  - Influence
  - Prediction
25. There is a set of churches in the U.S. where part of the service involves snake handling. The researcher wants to find out why the people attending these churches do this and how they feel and think about it. In this case, the primary purpose of the study is:
- Exploration
  - Description
26. Which of the following is not a characteristic of a good theory or explanation?
- It is parsimonious
  - It is testable
  - It is general enough to apply more than one place, situation, or person
  - All of the above are characteristics of good theories
27. Which of the following is not a basic assumption of science?
- Science cannot provide answers to all questions
  - It is possible to distinguish between more and less plausible claims
  - Researchers should follow certain agreed upon norms and practices
  - ✓ Science is best at solving value conflicts, such as whether abortion is immoral
28. What general type of research is focused on collecting information to help a researcher advance an ideological or political position?
- Evaluation research
  - Basic research
  - Action research
  - ✓ Orientational research
29. Which "scientific method" follows these steps: 1) observation/data, 2) patterns, 3) theory?
- ✓ Inductive
  - Deductive
  - Imductive
  - Top down
30. René Descartes is associated with which of the following approaches to knowledge generation?
- Empiricism
  - Rationalism
  - Expert opinion
  - None of the above

- 31 Which scientific method is a top-down or confirmatory approach?
- Deductive method
  - Inductive method
  - Hypothesis method
  - Pattern method
- 32 Which scientific method is a bottom-up or generative approach to research?
- Deductive method
  - Inductive method
  - Hypothesis method
  - Pattern method
- 33 Which scientific method focuses on testing hypotheses developed from theories?
- Deductive method
  - Inductive method
  - Hypothesis method
  - Pattern method
- 34 Which scientific method often focuses on generating new hypotheses and theories?
- Deductive method
  - Inductive method
  - Hypothesis method
  - Pattern method
- 35 Which of the following statements is true of a theory?
- It most simply means "explanation"
  - It answers the "how" and "why" questions
  - It can be a well developed explanatory system
  - All of the above are correct
- 36 In research, something that does not "vary" is called a
- Variable
  - Method
  - Constant
  - Control group
- 37 Non-experimental research in which the primary independent variable of interest is categorical is sometimes called
- Causal-comparative research
  - Correlational research
- 38 Which approach is the strongest for establishing that a relationship is causal?
- Causal-comparative
  - Correlational
  - Experimental
  - Historical
- 39 Which approach is the strongest for establishing that a relationship is causal?
- Causal-comparative
  - Correlational
  - One cannot say without additional information (i.e., it could be either depending on how well the researcher established the three necessary conditions for cause and effect)
- 40 \_\_\_\_\_ is the most commonly used technique for controlling for extraneous variables in non-experimental research
- Matching
  - Holding extraneous variables constant
  - Statistical control
  - Static control
- 41 It is best to use the method of working multiple hypotheses when
- You are finished with your research
  - You are planning your research study
  - You are hoping to publish your already obtained research results
  - None of the above
- 42 Matching can be done when your independent variable is categorical or quantitative.
- True
  - False
- 43 Which research paradigm is based on the pragmatic view of reality?
- Quantitative research
  - Qualitative research
  - Mixed research
  - None of the above

88 Compendium of Statistics

- 44 Which research paradigm is least concerned about generalizing its findings?
- Quantitative research
  - Qualitative research
  - Mixed research
  - None of the above
- 45 Which of the following best describes quantitative research?
- The collection of nonnumerical data
  - An attempt to confirm the researcher's hypotheses
  - Research that is exploratory
  - Research that attempts to generate a new theory
- 46 A condition or characteristic that can take on different values or categories is called
- A constant
  - A variable
  - A cause-and-effect relationship
  - A descriptive relationship
- 47 A variable that is presumed to cause a change in another variable is called
- Categorical variable
  - Dependent variable
  - Independent variable
  - Intervening variable
- 48 All of the following are common characteristics of experimental research except:
- It relies primarily on the collection of numerical data
  - It can produce important knowledge about cause and effect
  - It uses the deductive scientific method
  - It rarely is conducted in a controlled setting or environment
- 49 Qualitative research is often exploratory and has all of the following characteristics except
- It is typically used when a great deal is already known about the topic of interest
  - It relies on the collection of nonnumerical data such as words and pictures
- 50 It is used to generate the hypothesis and develop theory about phenomena in the world
- 50 Which type of research provides the strongest evidence about the existence of cause-and-effect relationships?
- Nonexperimental Research
  - Experimental Research
- 51 What is the key defining characteristic of experimental research?
- Extraneous variables are never present
  - A positive correlation usually exists
  - A negative correlation usually exists
  - Manipulation of the independent variable
- 52 In \_\_\_\_\_, random assignment to groups is never possible and the researcher cannot manipulate the independent variable.
- Basic research
  - Quantitative research
  - Experimental research
  - Causal-comparative and correlational research
- 53 What is the defining characteristic of experimental research?
- Resistance to manipulation
  - Manipulation of the independent variable
  - The use of open-ended questions
  - Focuses only on local problems
- 54 Research in which the researcher uses the qualitative paradigm for one phase and the quantitative paradigm for another phase is known as
- Action research
  - Basic research
  - Quantitative research
  - Mixed method research e. mixed model research
- 55 Research in which the researcher uses both qualitative and quantitative research within a stage or across two

- of the stages in the research process is known as
- Action research
  - Basic research
  - Quantitative research
  - Mixed method research
  - Mixed model research
56. Research that is done to understand an event from the past is known as?
- Experimental research
  - Historical research
  - Replication
  - Archival research
57. \_\_\_\_\_ research occurs when the researcher manipulates the independent variable.
- Causal-comparative research
  - Experimental research
  - Ethnography
  - Correlational research
58. Which of the following includes examples of quantitative variables?
- Age, temperature, income, height
  - Grade point average, anxiety level, reading performance
  - Gender, religion, ethnic group
  - Both a) and b)
59. What is the opposite of a variable?
- A constant
  - An extraneous variable
  - A dependent variable
  - A data set
60. Which of the following is the type of non-experimental research in which the primary independent variable of interest is categorical?
- Causal-comparative research
  - Experimental research
  - Qualitative research
  - Mixed research
61. Which of the following can best be described as a categorical variable?
- Age
  - Annual income
  - Grade point average
  - Religion
62. When interpreting a correlation coefficient expressing the relationship between two variables, it is very important to avoid.
- Checking the strength of relationship
  - Jumping to the conclusion of causality
  - Checking the direction of the relationship
  - Expressing a relationship with a correlation coefficient
63. A researcher studies achievement by children in poorly funded elementary schools. She develops a model that posits parent involvement as an important variable. She believes that parent involvement has an impact on children by increasing their motivation to do school work. Thus, in her model, greater parent involvement leads to higher student motivation, which in turn creates higher student achievement. Student motivation is what kind of variable in this study?
- Manipulated variable
  - Extraneous variable
  - Confounding variable
  - Mediating or intervening variable
64. The strongest evidence for causality comes from which of the following research methods?
- Experimental
  - Causal-comparative
  - Correlational
  - Ethnography
65. Which correlation is the strongest? (Fpsc, Lecturer 2008)
- +0.10
  - 0.95
  - +0.90
  - 1.00
66. The correlation between intelligence test scores and grades is: (Fpsc, Lecturer 2008)
- Positive
  - Negative
  - Perfect
  - They are not correlated

67. A good qualitative problem statement
- Defines the independent and dependent variables
  - Conveys a sense of emerging design
  - Specifies a research hypothesis to be tested
  - Specifies the relationship between variables that the researcher expects to find
68. The "tool" function of theory is to:
- Summarize existing knowledge
  - Summarize existing hypotheses
  - Suggest new relationships and make new predictions
  - Suggest new theories
69. The statement of purpose in a research study should:
- Identify the design of the study
  - Identify the intent or objective of the study
  - Specify the type of people to be used in the study
  - Describe the study
70. Why is the statement "What are the effects of extracurricular activities on cognitive development of school age children" not a good statement of a quantitative research question?
- Because there is no connection between extracurricular activities and cognitive development
  - Because there are not enough school age children engaged in extracurricular activities to conduct the study
  - Because the study would be too difficult to do given all the different extracurricular activities
  - Because the statement was not specific enough to provide an understanding of the variables being investigated
71. A qualitative research question:
- Asks a question about some process, or phenomenon to be explored
  - Is generally an open-ended question
72. Which of the following orders is recommended in the flowchart of the development of a research idea?
- Research topic, research problem, research question, hypothesis
  - Research topic, research purpose, research problem, research question, hypothesis
  - Research topic, research problem, research purpose, research question, hypothesis
  - Research topic, research hypothesis, research problem, research question, research purpose
73. It is essential that you evaluate the quality of internet resources because information obtained via the internet ranges from very poor to very good.
- True
  - False
74. One step that is not included in planning a research study is
- Identifying a researchable problem
  - A review of current research
  - Statement of the research question
  - Conducting a meta-analysis of the research
  - Developing a research plan
75. Sources of researchable problems can include:
- Researchers' own experiences as educators
  - Practical issues that require solutions
  - Theory and past research
  - All of the above
76. A key characteristic of past research that guides researchers in new research questions is that:
- Extensive research conclusively and definitively answers research questions
  - Studies typically generate more research questions than they answer

- 77 Which of the following is a function of theory?
- Integrating and summarizing current knowledge
  - Making predictions
  - Explaining phenomena
  - All of the above are important functions of theory
- 78 A review of the literature prior to formulating research questions allows the researcher to do which of the following?
- To become familiar with prior research on the phenomenon of interest
  - To identify potential methodological problems in the research area
  - To develop a list of pertinent problems relative to the phenomenon of interest
  - All of the above
- 79 Sometimes a comprehensive review of the literature prior to data collection is not recommended by grounded theorists.
- True
  - False
- 80 What kind of ideas can't be empirically researched?
- Effectiveness of different methods of instruction
  - Description of educational practices
  - Issues of values and morality such as the correctness of having prayer in schools
  - Factors helpful in predicting future drug use
- Which of the following is not database containing information to be used during the literature review?
- ERIC
  - PsychINFO
  - SocioFILE
  - All of the above are potentially useful data bases
- 82 Computer database searches can be done
- With a computer with CD-ROM drive
  - At the library
  - Online
  - All of the above
- 83 The feasibility of a research study should be considered in light of
- Cost and time required to conduct the study
  - Skills required of the researcher
  - Potential ethical concerns
  - All of the above
- 84 A formal statement of the research question or "purpose of research study" generally.
- Is made prior to the literature review
  - Is made after the literature review
  - Will help guide the research process
  - All of the above
  - Both b) and c)
- 85 Is the following qualitative research purpose statement "well stated" or "poorly stated"? "The focus of the present study was to explore distressing and nurturing encounters of patients with caregivers and to ascertain the meanings that are engendered by such encounters. The study was conducted on one of the surgical units and the obstetrical/gynecological unit of a 374-bed community hospital."
- It is a well stated
  - It is poorly stated
- 86 Which of the following quantitative research questions is superior?
- "What is the effect of participation in various extracurricular activities on academic performance?"
  - "What effect does playing high school football have on students' overall grade point average during the football season?"

- Answers*
- 102 Compendium of Statistics
87. A statement of the quantitative research question should exactly address the purpose(s) the researcher will be specifying in selecting appropriate participants, research methods, measures, and materials.
- a) Extend the statement of purpose by specifying exactly the question(s) the researcher will address
- b) Help the research in selecting appropriate participants, research methods, measures, and materials
- c) Specify the variables of interest
- d) All of the above
88. The research participants are described in detail in which section of the research plan?
- a) Introduction
- b) Method
- c) Data analysis
- d) Discussion
89. Research hypotheses are
- a) Formulated prior to a review of the literature
- b) Statements of predicted relationships between variables
- c) Stated such that they can be confirmed or refuted
- d) Both b) and c)
90. Hypotheses in qualitative research studies usually..
- a) Are very specific and stated prior to beginning the study
- b) Are often generated as the data are collected, interpreted, and analyzed
- c) Are never used
- d) Are always stated after the research study has been completed
91. A research plan
- a) Should be detailed
- b) Should be given to others for review and comments
- c) Sets out the rationale for a research study
- d) All of the above
92. The Method section of the research plan typically specifies
- a) The research participants
- b) The results of prior studies address the phenomena of interest
- c) The apparatus, instruments, materials for the research study
- d) The planned research procedures
- e) a), c) and d)
93. The Introduction section of the research plan
- a) Gives an overview of prior relevant studies
- b) Contains a statement of the purpose of the study
- c) Concludes with a statement of the research questions and, for quantitative research, it includes the research hypothesis
- d) All of the above
94. Which of the following is not a source of research ideas?
- a) Everyday life
- b) Practical issues
- c) Past research
- d) Theory
- e) All of the above ARE sources of research ideas
95. Ethics is the set of principles and guidelines that help us to uphold the things we value.
- a) True
- b) False
96. Which of the following is necessary in obtaining informed consent?
- a) A description of the statistical analyses that will be carried out
- b) A description of the purpose of the research
- c) A description of the reliability and validity of test instruments
- d) A list of publications that the researcher has had in the last ten years
97. Which of the following need(s) to be obtained when doing research with children?
- a) Informed consent from the parent or guardian
- b) Assent from the child if he or she is capable

98. Which of the following is true about the use of deception in research?
- It should never be used
  - It can be used anytime
  - If there is deception in a study, the participants may need to be debriefed
  - The use of deception must be outweighed by other benefits of the study
  - Both c) and d) are true
99. Which of the following generally cannot be done in qualitative studies conducted in the field?
- Getting informed consent
  - Keeping participants from physical harm
  - Maintaining consent forms
  - Having full anonymity rather than just confidentiality
100. Which of the following approaches says that ethical issues should be judged on the basis of some universal code?
- Deontological
  - Ethical skepticism
  - Utilitarianism
101. Which of the following is not an ethical guideline for conducting research with humans? Getting informed consent of the participant
- Telling participants they must continue until the study has been completed.
  - Keeping participants' identity anonymous.
  - Telling participants they are free to withdraw at any time.
102. Which of the three ethics approaches says research ethics should be a matter of the individual's conscience?
- Deontological approach
  - Ethical skepticism
  - Utilitarianism
  - Ontological skepticism
103. \_\_\_\_\_ means that the participant's identity, although known to the researcher, is not revealed to anyone outside of the researcher and his or her staff.
- Anonymity
  - Confidentiality
104. Which of the following is not true?
- Misrepresenting and creating fraudulent data is dishonest
  - Misrepresenting data is very easy to detect
  - Misrepresenting data can be difficult to detect
  - Breaking confidentiality is not a problem
105. Ideally, the research participant's identity is not known to the researcher. This is called:
- Anonymity
  - Confidentiality
  - Deception
  - Desensitizing
106. Which of the following approaches taken by people to resolve ethical issues is the primary approach used by the federal government and most professional organizations?
- Deontological approach
  - Ethical skepticism
  - Utilitarianism
  - None of the above
107. What is it called when the participants are not revealed to any one but researcher and staff?
- Confidentiality
  - Anonymity
  - Ethics Discretion
108. Research participants must give what before they can participate in a study?
- Guidelines
  - A commitment
  - Informed consent
  - Private information
109. There are three basic approaches that people tend to adopt when considering ethical issues in research.

Which one of the following is not one of the approaches?

- a) Ethical skepticism
- b) Deontology
- c) Ontology
- d) Utilitarianism

110. Identify the term that refers to a post study interview in which all aspects of the study are revealed, reasons for the use of deception are given, and the participants' questions are answered?

- a) Desensitizing
- b) Debriefing
- c) Dehoaxing
- d) Deploying

111. A set of principles to guide and assist researchers in deciding which goals are most important and in reconciling conflicting values when conducting research is called.

- a) Research ethics
- b) Deontological approach
- c) Utilitarianism
- d) None of the above

112. IRB is an acronym for which of the following?

- a) Internal Review Board
- b) Institutional Rating Board
- c) Institutional Review Board
- d) Internal Request Board

113. When it is necessary to engage in a good amount of deception to conduct a scientifically valid study, what procedure should a researcher consider following?

- a) Debriefing
- b) Dehoaxing
- c) Desensitizing
- d) All of the above should be considered

114. The act of publishing the same data and results in more than one journal or publication refers to which of the following professional issues:

- a) Partial publication
- b) Duplicate publication
- c) Deception

115. Full publication concerning "authorship" in educational research, intellectual ownership, predominantly a function of

- a) Effort expended
- b) Creative contribution
- c) Professional position
- d) Level of higher education

116. Which term refers to publishing several articles from the data collected in one large study?

- a) Duplicate publication
- b) Partial publication
- c) Triplicate publication
- d) None of these

117. Which of the following is a right of each participant according to the AERA?

- a) Deception
- b) Utilitarianism
- c) Freedom to withdraw
- d) Participants have no rights

118. If a correlation coefficient is .96, we would probably be able to say that the relationship is:

- a) Weak
- b) Strong
- c) Statistically significant
- d) b is true and c is probably true

119. What happens in a completely spurious relationship once the researcher controls for a confounding third-variable?

- a) The relationship between the original variables will get stronger
- b) The relationship between the original variables will remain unchanged
- c) The correlation coefficient will get closer to 1.0
- d) The relationship between the original variables will get weaker or, if the original relationship is fully spurious, it will disappear (i.e., the original relationship will become zero as measured by a correlation coefficient)

- 120 Which of the following is NOT a form of descriptive research?
- Yield study
  - Mean study
  - Cross-sectional study
- 121 Researching a relationship between two variables is NOT sufficient grounds for concluding that the relationship is a causal relationship.
- True
  - False
- 122 This type of research tests hypotheses and theories in order to explain how and why a phenomenon operates as it does.
- Descriptive research
  - Predictive research
  - Explanatory research
  - None of the above
- 123 Which of the following is not an assumption underlying testing and measurement?
- Various approaches to measuring aspects of the same thing can be useful
  - Error is rarely present in the measurement process
  - Present-day behavior predicts future behavior
  - Testing and assessment benefit society
- 124 Systematic error is associated with:
- Reliability
  - Validity
- 125 Which of the following is a type of criterion-related validity evidence?
- Concurrent evidence
  - Predictive evidence
  - Internal consistency
  - Both a) and b) are correct
- 126 If a test measures a single construct then:
- The items should correlate with the total score
  - The items should not correlate with the total score
- 127 The test should not correlate with other measures of the same construct.
- There must be a reliable alternative form.
- 128 Professor X develops a test of emotional intelligence. Which of the following represent convergent and discriminant evidence?
- The test correlates highly with another test of emotional intelligence and is uncorrelated with self-efficacy
  - The test correlates with highly with another test of emotional intelligence and is highly correlated with self-efficacy
  - The test does not correlate with another test of emotional intelligence, but does correlate with self-efficacy
  - The test does not correlate with other tests of emotional intelligence nor with self-efficacy
- 129 \_\_\_\_\_ tests focus on information acquired through the informal learning that goes on in life.
- Personality
  - Achievement
  - Aptitude
  - Intelligence
- 130 Let's say that a test accurately indicates participants' scores on a future criterion (e.g., the PSAT is used to indicate high-school GPA scores). This test would clearly have which of the following?
- Face validity
  - Concurrent validity
  - Predictive validity
  - Content validity
- 131 Which of the following is most clearly an example of a psychological trait?
- Anxiety enduring for months or years
  - Anxiety over just seeing a spider

- c) Shyness when meeting a stranger  
d) for the first time  
d) Depression caused by the loss of a ball game
131. All of the following are examples of intelligence tests except:
- Wechsler Scales
  - Stanford-Binet
  - Minnesota Multiphasic Personality Inventory(MMPI)
  - Slosson
132. Reliability is most simply known as which of the following?
- Consistency or stability
  - Appropriateness of interpretations on the basis of test scores
  - Ways in which people are the same
  - A rank order of participants on some characteristic
133. Which of the following is not a type of reliability?
- Test-retest
  - Split-half
  - Content
  - Internal consistency
134. Which of the following statements accurately describes test-retest reliability?
- Measure of consistency of test scores over time
  - Measure of consistency of scores obtained from two equivalent halves of the same test
  - Measure of consistency with which a test measures a single construct or concept
  - Measure of degree of agreement between two or more scorers, judges, or raters
135. Which of the following types of reliability refers to the consistency of test scores over time?
- Equivalent forms reliability
  - Split-half reliability
  - Test-retest reliability
  - Inter-scorer reliability
136. Identify the following term that most closely refers to a judgment of the extent to which scores from a test can be used to infer, or predict, the examinees' performance in some activity
- Content reliability
  - Face validity
  - Criterion-related validity
  - Inference validity
137. Which is the process of gathering evidence supporting inferences based on test scores?
- Validation
  - Validity
  - Reliability
  - Prediction
138. When evaluating tests and assessments, "reliability" refers to asking ourselves which of the following questions?
- Does it measure what it is supposed to measure?
  - Are there ways to avoid subjective judgments when measuring something?
  - Does it give consistent results?
  - Does it measure multiple constructs?
139. Validity of a test designed to measure a construct such as self-esteem is best described by which of the following?
- Scores from the test correlate highly with most intelligence tests
  - Scores from the test correlate highly with most tests of different constructs
  - Scores from the test are not correlated with anything
  - Scores from the test have a relatively strong and positive correlation with other tests of the same construct (i.e., with other measures of self-esteem) but much lower correlations with tests of different constructs

40 Which type of reliability refers to the consistency of a group of individuals' scores on two equivalent forms of a test designed to measure the same characteristic?

- a) Split-half
- b) Test-retest
- c) Split-forms
- d) Equivalent forms

41. \_\_\_\_\_ refers to how well the particular sample of behaviors used to measure a characteristic reflects the entire domain of behaviors that constitutes that characteristic.

- a) Construct validity evidence
- b) Criterion-related validity evidence
- c) Content validity evidence
- d) Face validity evidence

42 According to theory, how many points should a rating scale have?

- a) Five
- b) Four
- c) Ten
- d) Somewhere from 4 to 11 points

43. What is the problem with this set of response categories to the question "What is your current age?"

- 1-5
- 5-10
- 10-20
- 20-30
- 30-40

- a) The categories are not mutually exclusive
- b) The categories are not exhaustive
- c) Both a) and b) are problems
- d) There is no problem with the above set of response categories

44. You should mix methods in a way that provides complementary strengths and non overlapping weaknesses. This is known as the fundamental principle of mixed research.

- a) True
- b) False

145. Questionnaires can address events and characteristics taking place when?

- a) In the past (retrospective questions)
- b) In the present (current time questions)
- c) In the future (prospective questions)
- d) All of the above

146. Which of the following are principles of questionnaire construction?

- a) Consider using multiple methods when measuring abstract constructs
- b) Use multiple items to measure abstract constructs
- c) Avoid double-barreled questions
- d) All of the above
- e) Only b and c

147. Which of these is not a method of data collection?

- a) Questionnaires
- b) Interviews
- c) Experiments
- d) Observations

148. Secondary/existing data may include which of the following?

- a) Official documents
- b) Personal documents
- c) Archived research data
- d) All of the above

149. An item that directs participants to different follow-up questions depending on their response is called a.

- a) Response set
- b) Probe
- c) Semantic differential
- d) Contingency question

150. Which of the following terms best describes data that were originally collected at an earlier time by a different person for a different purpose?

- a) Primary data
- b) Secondary data
- c) Experimental data

- Researchers use both open-ended data and closed-ended questions to collect statements. Which of the following statements is true?
- Open-ended questions directly provide quantitative data based on the researcher's predetermined response categories
  - Closed-ended questions provide quantitative data in the participant's own words
  - Open-ended questions provide qualitative data in the participant's own words
  - Closed-ended questions directly provide qualitative data in the participants own words
- 152 Open ended questions provide primarily \_\_\_\_\_ data
- Confirmatory data
  - Qualitative data
  - Predictive data
  - None of the above
- 153 Which of the following is true concerning observation?
- It takes less time than self-report approaches
  - It costs less money than self-report approaches
  - It is often not possible to determine exactly why the people behave as they do
  - All of the above
- 154 Qualitative observation is usually done for exploratory purposes; it is also called
- Observation
  - Structured
  - Naturalistic
  - Complete
  - Probed
- 155 When constructing a questionnaire it is important to do each of the following except.
- Use "leading" or "loaded" questions
  - Use natural language
- c) Understand your research participants  
d) Pilot your test questionnaire
- 156 Another name for a Likert Scale is a
- Interview protocol
  - Event sampling
  - Summated rating scale
  - Ranking
- 157 Which of the following is not one of the six major methods of data collection that are used by educational researchers?
- Observation
  - Interviews
  - Questionnaires
  - Checklists
- 158 The type of interview in which the specific topics are decided in advance but the sequence and wording can be modified during the interview is called
- The interview guide approach
  - The informal conversational interview
  - A closed quantitative interview
  - The standardized open-ended interview
- 159 Which one of the following is not a major method of data collection:
- Questionnaires
  - Interviews
  - Secondary data
  - Focus groups
  - All of the above are methods of data collection
- 160 A question during an interview such as "Why do you feel that way?" known as a:
- Probe
  - Filter question
  - Response
  - Pilot
- 161 A census taker often collects data through which of the following?
- Standardized tests
  - Interviews
  - Secondary data
  - Observations

- 162 When desired information is available for all objects in the population, we have what is called a  
 a) Population  
 b) Sample  
 c) Census  
 d) Stem-and-leaf display
- 163 The researcher has secretly placed him or herself (as a member) in the group that is being studied. This researcher may be which of the following?  
 a) A complete participant  
 b) An observer-as-participant  
 c) A participant-as-observer  
 d) None of the above
- 164 Which of the following is not a major method of data collection?  
 a) Questionnaires  
 b) Focus groups  
 c) Correlational method  
 d) Secondary data
- 165 Which type of interview allows the questions to emerge from the immediate context or course of things?  
 a) Interview guide approach  
 b) Informal conversational interview  
 c) Closed quantitative interview  
 d) Standardized open-ended interview
- 166 When conducting an interview, asking "Anything else? What do you mean? Why do you feel that way?" etc, are all forms of:  
 a) Contingency questions  
 b) Probes  
 c) Protocols  
 d) Response categories
- 167 When constructing a questionnaire, there are 15 principles to which you should adhere. Which of the following is not one of those principles?  
 a) Do not use "leading" or "loaded" questions  
 b) Avoid double-barreled questions  
 c) Avoid double negatives  
 d) Avoid using multiple items to measure a single construct
- 168 Census is not applicable when the population is  
 a) Finite  
 b) Infinite  
 c) Homogeneous  
 d) Heterogeneous  
 e) None of the above
- 169 Sampling is used when  
 a) Complete enumeration is not possible  
 b) Population is infinite  
 c) Results required in short time  
 d) All of the above  
 e) None of the above
- 170 A subset of a population is known as  
 a) A sample  
 b) A continuous variable  
 c) A random variable  
 d) A discrete variable
- 171 Quasi-Random sampling is also called  
 a) Snowball sampling  
 b) Simple random sampling  
 c) Stratified sampling  
 d) Systematic sampling  
 e) None of the above
- 172 Which of the following sampling is the representative of the whole population?  
 a) Stratified sampling  
 b) Simple random sampling  
 c) Systematic sampling  
 d) All of the above  
 e) None of the above
- 173 A survey is to be carried out in the central area of a town and the residential outskirts of it. The shopping habits and expenditure of people living there are to be studied. The best sampling method would be  
 a) A stratified random sample of individuals  
 b) A stratified random sample of households  
 c) A simple random sample of individuals  
 d) A simple random sample of households

- e) A systematic sample of individuals using a list of registered electors
174. In a village of 300 population, 60% constitute Hindus, 20% Muslims, 10% Sikhs and 10% Christians. We want to take a sample of 10% of the population to study the eating habits of this population. The best method would be:
- Simple random sampling
  - Stratified random sampling
  - Systematic random sampling
  - Non-random sampling
  - Cluster sampling
175. Accuracy refers to the:
- Degree of agreement of measurements with an accepted reference value.
  - Number of digits on an LCD readout.
  - Ability to obtain the same value with more than one measuring device.
  - Ability to get the same answer more than once using the same measurement system.
176. When each member of a population has an equally likely chance of being selected, this is called:
- A nonrandom sampling method
  - A quota sample
  - A snowball sample
  - An Equal probability selection method
177. A sample of size 35 is to be drawn from a large population. The sampling technique is such that every possible sample of size 35 that could be drawn from the population is equally likely. What name is given to this type of sample?
- Systematic sample
  - Cluster sample
  - Voluntary response sample
  - Random sample
  - Simple random sample
178. Which of the following techniques yields a simple random sample?
- Choosing volunteers from an introductory psychology class to participate
  - Listing the individuals by ethnic group and choosing a proportion from within each ethnic group at random.
  - Numbering all the elements of a sampling frame and then using a random number table to pick cases from the table.
  - Randomly selecting schools, and then sampling everyone within the school.
179. Which of the following is not true about stratified random sampling?
- It involves a random selection process from identified subgroups
  - Proportions of groups in the sample must always match their population proportions
  - Disproportional stratified random sampling is especially helpful for getting large enough subgroup samples when subgroup comparisons are to be done
  - Proportional stratified random sampling yields a representative sample
180. Which of the following statements are true?
- The larger the sample size, the greater the sampling error
  - The more categories or breakdowns you want to make in your data analysis, the larger the sample needed
  - The fewer categories or breakdowns you want to make in your data analysis, the larger the sample needed
  - As sample size decreases, so does the size of the confidence interval

181. Which of the following formulae is used to determine how many people to include in the original sampling?
- Desired sample size/Desired sample size + 1
  - Proportion likely respond/desired sample size to
  - Proportion likely respond/population size to
  - Desired sample size/Proportion likely to respond
182. Which of the following sampling techniques is an equal probability selection method (i.e., EPSEM) in which every individual in the population has an equal chance of being selected?
- Simple random sampling
  - Systematic sampling
  - Proportional stratified sampling
  - Cluster sampling using the PPS technique
  - All of the above are EPSEM
183. Which of the following is not a form of nonrandom sampling?
- Snowball sampling
  - Convenience sampling
  - Quota sampling
  - Purposive sampling
  - They are all forms of nonrandom sampling
184. Which of the following will give a more "accurate" representation of the population from which a sample has been taken?
- A large sample based on the convenience sampling technique
  - A small sample based on simple random sampling
  - A large sample based on simple random sampling
  - A small cluster sample
185. Sampling in qualitative research is similar to which type of sampling in quantitative research?
- Simple random sampling
  - Systematic sampling
  - Quota sampling
  - Purposive sampling
186. Which of the following would generally require the largest sample size?
- Cluster sampling
  - Simple random sampling
  - Systematic sampling
  - Proportional stratified sampling
187. For a survey, a village is divided into 5 lanes, each lane is sampled randomly. It is an example of
- Simple random sampling
  - Standard random sampling
  - Systematic random sampling
  - Cluster random sampling
  - Quasi random sampling
188. From its 32 regions, the Federal Aviation Administration randomly selects 5 regions, and then randomly audits 20 departing commercial flights in each region for compliance with legal fuel and weight requirements. This most nearly resembles
- Simple random sampling
  - Systematic random sampling
  - Two Stage cluster sampling
  - Judgment sampling
  - None of these
189. Which statement is false?
- Random dialing phone surveys have low response and are poorly targeted.
  - Selection bias means that many respondents dislike the interviewer.
  - Simple random sampling requires a list of the population.
  - Web surveys are economical but suffer from non-response bias.
190. Professor Hardtack chose a sample of 7 students from his statistics class of 35 students by picking every student who was wearing red that day. Which kind of sample is this?
- Simple random sample
  - Judgment sample
  - Systematic sample
  - Convenience sample
  - None of the above

191. Which of the following situations is a cluster sample?
- Survey five friends concerning their opinions of the local hockey team.
  - Take a random sample of five voting precincts in a large metropolitan area and do an exit poll at each voting site.
  - Measure the length of time each fifth person entering a restaurant has to wait to be seated.
  - From a list of all students in your school, randomly select 20 to answer a survey about Internet use.
  - Identify four different ethnic groups at your school. From each group, choose enough respondents so that the final sample contains roughly the same proportion of each group as the school population.
192. How often does the Census Bureau take a complete population count?
- Every year
  - Every five years
  - Every ten years
  - Twice a year
193. People who are available, volunteer, or can be easily recruited are used in the sampling method called
- Simple random sampling
  - Cluster sampling
  - Systematic sampling
  - Convenience sampling
194. Which of the following types of sampling involves the researcher determining the appropriate sample sizes for the groups identified as important, and then taking convenience samples from those groups?
- Proportional stratified sampling
  - Quota sampling
  - One-stage cluster sampling
  - Two-stage cluster sampling
195. A type of sampling used in qualitative research that involves selecting cases that disconfirm the researcher's expectations and generalizations referred to as
- Extreme case sampling
  - Typical-case sampling
  - Critical-case sampling
  - Negative-case sampling
196. In which of the following nonrandom sampling techniques does the researcher ask the research participants to identify other potential research participants?
- Snowball
  - Convenience
  - Purposive
  - Quota
197. Which of the following is the most efficient random sampling technique?
- Simple random sampling
  - Proportional stratified sampling
  - Cluster random sampling
  - Systematic sampling
198. If we to the 500 people attending a school in New York City, divided them by gender, and then to a random sample of the males and a random sampling of the females, the variable on which we would divide the population is called the
- Independent variable
  - Dependent variable
  - Stratification variable
  - Sampling variable
199. A number calculated with complete population data and quantifies a characteristic of the population is called which of the following?
- A datum
  - A statistic
  - A parameter
  - A population
200. The type of sampling in which each member of the population selected for the sample is returned to the population before the next member is selected is called
- Sampling without replacement
  - Sampling with replacement

- c) Simple random sampling  
d) Systematic sampling
- 201 If sampling without replacement occurs, do the picks follow the Binomial Distribution?  
a) Yes, because each pick is independent from the others  
b) No, because the probability of success on each pick changes  
c) Yes, if we are counting the number of successes.  
d) No, because we may not have any successes.
- 202 Which of the following would usually require the smallest sample size because of its efficiency?  
a) One stage cluster sampling  
b) Simple random sampling  
c) Two stage cluster sampling  
d) Quota sampling
- 203 A technique used when selecting clusters of different sizes is called.  
a) Cluster sampling  
b) One-stage sampling  
c) Two-stage sampling  
d) Probability proportional to size or PPS
- 204 The process of drawing a sample from a population is known as.  
a) Sampling  
b) Census  
c) Survey research  
d) None of the above
- 205 It is recommended to use the whole population rather than a sample, when the population size is of what size?  
a) 500 or less  
b) 100 or less  
c) 1000 or less  
d) you should always use a sample
- 206 Which of the following is not an example of a nonrandom sampling technique?  
a) Purposive  
b) Quota  
c) Convenience  
d) Cluster
- 207 Which of the following sampling methods is the best way to select a group of people for a study if you are interested in making statements about the larger population?  
a) Convenience sampling  
b) Quota sampling  
c) Purposive sampling  
d) Random sampling
- 208 \_\_\_\_\_ is a set of elements taken from a larger population according to certain rules.  
a) Sample  
b) Population  
c) Statistic  
d) Element
- 209 Determining the sample interval (represented by  $k$ ) randomly selecting a number between 1 and  $k$  and including each  $k$ th element in your sample are the steps for which form of sampling?  
a) Simple Random Sampling  
b) Stratified Random Sampling  
c) Systematic Sampling  
d) Cluster sampling
- 210 In a study carried out in the hospital ward, every 10th admitted patient was included in the sample. Which sampling procedure is this?  
a) Random sampling  
b) Stratified sampling  
c) Quota sampling  
d) Convenient sampling  
e) Systematic sampling
- 211 The nonrandom sampling type that involves selecting a convenience sample from a population with a specific set of characteristics for your research study is called  
a) Convenience sampling  
b) Quota sampling  
c) Purposive sampling  
d) Snowball sampling
- 212 When a extraneous variable systematically varies with the independent variable and influences the dependent variable, it is called

- a) Another dependent variable
- b) A confounding variable
- c) A moderating variable
- d) An unreliable variable

213. Which of the following statements is true?

- a) A statistical relationship is sufficient evidence to infer causality
- b) Temporal order of the cause and effect is not important in inferring causality
- c) A statistical relation of X and Y is insufficient evidence for inferring causality
- d) Temporal order of cause and effect variables and statistical relation are all that are needed to infer causality

214. A school district examines a program that uses mentors to help very poor readers improve their reading performance. The children in the program are at the 4<sup>th</sup> percentile at pretest. At posttest they are around the 20<sup>th</sup> percentile. While it is possible that the program made the difference, another reason for the change in scores could be:

- a) History
- b) Regression artifact
- c) Multiple-treatment interference
- d) Differential selection

215. A group of researchers do a study where children from particular classrooms are assigned to treatment or control conditions. After the study, the researcher finds out that the students in the control group are higher achievers than those in the experimental group. He found no treatment effect. The failure to find an effect may be due to:

- a) A treatment effect
- b) A testing effect
- c) A differential selection effect
- d) A maturation effect

216. Which type of validity refers to the degree to which you can infer that the relationship between two variables is causal?

- a) Internal validity
- b) Population validity
- c) Ecological validity
- d) Statistical conclusion validity

217. Which type of validity refers to the ability to infer that the independent and dependent variables are related and that the measured strength of the relationship is accurate?

- a) Internal validity
- b) Population validity
- c) Ecological validity
- d) Statistical conclusion validity

218. An extraneous variable that systematically varies with the independent variable and also influences the dependent variable is known as a.

- a) Confounding variable
- b) Third variable
- c) Second variable
- d) Both a) and b) are correct

219. The use of multiple observers to allow cross-checking of observations to make sure that the investigators agree with what to place is known as.

- a) Interpretive validity
- b) Researcher bias
- c) Multiple operationalism
- d) Investigator triangulation

220. \_\_\_\_\_ is the lowest inference descriptor of all because it uses the participant's own words.

- a) Participant feedback
- b) A verbatim
- c) Data triangulation
- d) Investigator triangulation

221. \_\_\_\_\_ refers to physical or mental changes that may occur within individuals over time, such as aging, learning, boredom, hunger, and fatigue.

- a) Instrumentation
- b) History
- c) Maturation
- d) Testing

222. What type of validity refers to the extent to which the results of a study can be generalized across time?
- Ecological validity
  - External validity
  - Internal validity
  - Temporal validity
223. Which of the following best describes interpretive validity?
- Factual accuracy of an account as reported by the researcher
  - Accurately portraying the meanings given by the participants to what is being studied
  - Degree to which a theoretical explanation fits the data
  - Ability to generalize the study results across settings
224. Which of the following terms is a strategy where the researcher actively engages in critical self-reflection about his or her potential biases and predispositions?
- Experimenter effect
  - Reactivity
  - Investigator triangulation
  - Reflexivity
225. Which of the following is not considered one of the criteria for inferring causality?
- Evidence that the independent and dependent variables are related
  - Evidence that the relationship between the variables being investigated is not due to a confounding extraneous variable
  - Evidence that changes in variable A occur before changes in variable B
  - The temporal ordering of the variables being investigated does not matter because a relationship is all that is really needed
226. The use of multiple data sources to help understand a phenomenon is one strategy that is used to promote qualitative research validity. Which of the following terms describes this strategy?
- a) Data matching  
 b) Pattern matching  
 c) Data triangulation  
 d) Data feedback
227. What may happen when different comparison groups experience a different history event?
- History effect
  - Selection-history effect
  - Selection effect
  - Group effect
228. What is another term that refers to a confounding extraneous variable?
- Last variable
  - First variable
  - Third variable
  - Fourth variable
229. Which of the following refers to any systematic change that occurs over time in the way in which the dependent variable is assessed?
- Instrumentation
  - Maturation
  - Testing
  - Selection
230. Which of the following terms describes the ability to generalize from the sample of individuals on which a study was conducted to the larger target population of individuals and across different subpopulations within the larger target population?
- External validity
  - Population validity
  - Ecological validity
  - Temporal validity
231. Which of the following is not a strategy used to promote qualitative research validity?
- Peer review
  - Theory triangulation
  - Extended fieldwork
  - Random assignment
232. The use of several measures of a construct is called:
- Multiple operationalism
  - Multiple construct measurement
  - Operationalism
  - Methods triangulation

233. A physical or mental change that occurs in participants over time that affects their performance on the dependent variable is called.
- Instrumentation
  - Maturation
  - Regression
  - None of above
234. Attrition generally occurs in research where.
- You do demographic research
  - The study fails
  - Some participants do not complete the study
  - The study is very brief
235. Differential attrition occurs when the people dropping out from one group are different from the others in their group or from the people in the comparison group.
- True
  - False
236. Internal validity refers to which of the following?
- The ability to infer that a causal relationship exists between 2 variables
  - The extent to which study results can be generalized to and across populations of persons, settings, and times
  - The use of effective measurement instruments in the study
  - The ability to generalize the study results to individuals not included in the study
237. Which strategy used to promote qualitative research validity uses multiple research methods to study a phenomenon?
- Data triangulation
  - Methods triangulation
  - Theory triangulation
  - Member checking
238. Which type of validity refers to the factual accuracy of an account as reported by the researcher?
- Ecological validity
  - Temporal validity
  - Descriptive validity
  - None of the above
239. Which of the following is not one of the key threats to internal validity?
- Maturation
  - Instrumentation
  - Temporal change
  - History
240. This type of validity refers to the ability to generalize the results of a study across settings.
- Temporal validity
  - Internal validity
  - Ecological validity
  - External validity
241. Which is not a direct threat to the internal validity of a research design?
- History
  - Testing
  - Sampling error
  - Differential selection
242. Alteration in performance due to being aware that one is participating in a study is known as.
- Operationalism
  - Reactivity
  - Temporal validity
  - Mortality
243. The idea that the more times a research finding is shown with different sets of people, the more confidence we can place in the finding and in generalizing beyond the original participants is known as.
- Naturalistic generalization
  - Methods generalization
  - Data triangulation
  - Replication logic

- 244 In Survey the validity of a data is measured  
 a) Content validity  
 b) Construct validity  
 c) Both a & b  
 d) Non of the above
- 245 The reliability of a survey data is tested by which of the following  
 a) Retest method  
 b) Split halves method  
 c) Internal consistency method  
 d) All of the above
- 246 Internal consistency is measured by  
 a) Efficiency  
 b) Rank correlation  
 c) Cronbach's Alpha  
 d) Non of the above
- 247 The variance of the mean  $\bar{y}$  from a simple random sampling is  
 a) 
$$\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$$
  
 b) 
$$\frac{S^2}{n} \frac{(N-1)}{n}$$
  
 c) 
$$\frac{S^2}{n} \frac{(N-n)}{N}$$
  
 d) Non of the above
- 248 If  $\hat{Y}$  is an estimate of the total population then variance ( $\hat{Y}$ ) is  
 a) 
$$\frac{NS^2}{n} \frac{(N-n)}{N}$$
  
 b) 
$$\frac{N^2 S^2}{n} \frac{(N-n)}{N}$$
  
 c) 
$$\frac{N^2 S^2}{n} \frac{(N-1)}{N}$$
  
 d) Non of the above
- 249 Ratio Estimators are  
 a) Unbiased  
 b) Biased  
 c) Both a & b  
 d) Non of the above
- 250 The Design Effect (Deff) is a \_\_\_\_\_ technique  
 a) Parameter Estimation  
 b) Sampling Plan  
 c) Parameter Testing  
 d) Non of the above
- 251 In stratified sampling, subpopulations are  
 a) Overlapping  
 b) Non overlapping  
 c) Both a & b  
 d) Non of the above
- 252 Analysis of covariance is:  
 a) A statistical technique that can be used to help equate groups on specific variables  
 b) A statistical technique that can be used to control sequencing effects  
 c) A statistical technique that substitutes for random assignment to groups  
 d) Adjusts scores on the independent variable to control for extraneous variables
- 253 To determine whether noise affects the ability to solve math problems, a researcher has one group solve math problems in a quiet room and another group solve math problems in a noisy room. The group solving problems in the noisy room completes 15 problems in one hour and the group solving problems in the quiet room completes 22 problems in one hour. In this experiment, the independent variable is \_\_\_\_\_ and the dependent variable is \_\_\_\_\_.  
 a) The number of problems solved; the difficulty of the problems  
 b) The number of problems solved; the noise level in the room  
 c) The noise level in the room; the number of problems solved  
 d) The noise level in the room; the difficulty of the problems

254. The posttest-only design with nonequivalent groups is likely to control for which of the following threats to internal validity:
- History
  - Differential selection
  - Additive and interactive effects
  - Differential attrition
255. When all participants receive all treatment conditions, the study is susceptible to:
- Order effects
  - Carryover effects
  - Analysis of covariance
  - a) and b)
256. A researcher is interested in the effects of a preschool program on later school performance. Because she is concerned that socio-economic-status (SES) is a potential extraneous variable in her study, she picks children to study who are only from low SES homes. The control technique she used in this study was:
- Matching
  - Random assignment
  - Holding the extraneous variable constant
  - Statistically controlling the extraneous variable
257. Which of the following terms best describes an interaction effect?
- The effect of one independent variable (on a DV) depends on the level of another independent variable
  - Eliminating any differential influence of extraneous variables
  - Sequencing effect that occurs from the order in which the treatment conditions are administered
  - The effect of one independent variable on the dependent variable
258. Which of the following terms refers to a statistical method that can be used to statistically equate groups on a pretest or some other variable?
- Experimental control
  - Differential influence
- c) Matching  
d) Analysis of covariance
259. Which of the following is not a way to manipulate an independent variable?
- Presence technique
  - Amount technique
  - Type technique
  - Random technique
260. Which of the following designs permits a comparison of pretest scores to determine the initial equivalence of groups on the pretest before the treatment variable is introduced into the research setting?
- One-group pretest-posttest design
  - Pretest-posttest control group design
  - Posttest-only design with nonequivalent groups
  - Both b and c
261. Counterbalancing is.
- Usually based on random selection of participants
  - Only used when one pretest variable needs to be controlled
  - Chosen to control for such things as order and carryover effects\*
  - All of the above
262. The group that receives the experimental treatment condition is the.
- Experimental group
  - Control group
  - Participant group
  - Independent group
263. Which of the following control techniques available to the researcher controls for both known and unknown variables?
- Building the extraneous variable into the design
  - Matching
  - Random assignment
  - Analysis of covariance
264. The group that does not receive the experimental treatment condition is the.
- Experimental group

- b) Control group
- c) Treatment group
- d) Independent group

265 There are a number of ways in which confounding extraneous variables can be controlled. Which control technique is considered to be the best?

- a) Random assignment
- b) Matching
- c) Counterbalancing
- d) None of the above

266 Which of the following could be used for randomly assigning participants to groups in an experimental study?

- a) Split-half (e.g., first half versus second half of a school directory)
- b) Even versus odd numbers
- c) Use a list of random numbers or a computer randomization program
- d) Let the researcher decide which group will be the best

267 Which term is not a related to counterbalancing?

- a) Carryover effect
- b) Order effect
- c) Sequencing effects
- d) Matching

268 Which of the following designs does an excellent job of controlling for rival hypotheses that threaten the internal validity of an experiment?

- a) Posttest-only design with nonequivalent groups
- b) Posttest-only control-group design
- c) Pretest-posttest control-group design
- d) Both b) and c) are excellent designs

Manipulating the independent variable by varying the type on the independent variable that is presented to the different comparison groups is known as.

- a) Amount technique
- b) Absence technique
- c) Type technique
- d) Presence technique

270 Which of the following terms is a sequencing effect that occurs from the order in which the treatment conditions are administered?

- a) Carry-over effect
- b) Order effect
- c) Sequencing effects
- d) None of the above

271 When manipulating the independent variable in an educational experiment, which of the following describes this method?

- a) An independent variable is manipulated using the presence or absence technique
- b) The researcher varies the amount of the independent variable that is administered
- c) The researcher varies the type of the independent variable
- d) All of the above are possible

272 Which method of controlling confounding extraneous variables takes precedence over all other methods?

- a) Matching individual participants
- b) Holding extraneous variables
- c) Building the extraneous variable into the research design
- d) Counterbalancing
- e) Randomly assign research participants to the groups

273 In an experimental research study, the primary goal is to isolate and identify the effect produced by the.

- a) Dependent variable
- b) Extraneous variable
- c) Independent variable
- d) Confounding variable

274 The design in which one group of research participants is administered a treatment and is then compared, on the dependent variable, with another group of research participants who did not receive the experimental treatment is

- a) One-group posttest-only design
- b) One-group pretest-posttest design

- 275 A sequencing effect that occurs when performance in one treatment condition is influenced by participation in a prior treatment condition is known as:  
a) Counterbalancing effect  
b) Carryover effect  
c) Treatment effect  
d) Order effect

- 276 In a changing-criterion design, changes in criterion are best done:  
a) As soon as the prior criterion is met  
b) When the previous criterion is met and the behavior has stabilized  
c) Regardless of the previous criteria  
d) After a fixed number of trials

- 277 The non-equivalent comparison group design is a quasi-experimental design in which, for reasons of practicality, we cannot insure that the control and experimental groups are equivalent to each other when the experiment begins. The major interpretational difficulty imposed by this design is:  
a) Measuring whether the two groups are different from each other on the posttest  
b) Deciding how much each group has gained  
c) Determining when enough data points are collected  
d) Being sure that any differences between groups at the end of the experiment are due to the independent variable's influence and not due to preexisting group differences

- 278 A treatment effect is demonstrated in the regression discontinuity design by:  
a) A discontinuity in the regression line  
b) A significant difference in the pretest and posttest scores

- 279 Which of the following is a primary threat to the interrupted time-series design?  
a) History effect  
b) Selection-history  
c) Selection-maturation  
d) All of the above

- 280 A design consisting of an experimental and a control group but participants are not randomly assigned to the groups is which of the following?  
a) Interrupted time-series design  
b) Nonequivalent comparison-group design

- c) Single case design  
d) A-B-A-B design

- 281 A form of the nonequivalent comparison-group design is recommended when:

- a) It is not possible to control for a basic history effect  
b) It is not possible to randomly assign participants to groups  
c) It is not possible to identify two groups  
d) All of the above

- 282 A threat to internal validity in the nonequivalent comparison-group design is the \_\_\_\_\_ effect.

- a) Selection-maturation effect  
b) Selection-history effect  
c) Selection-regression  
d) All of the above are threats

- 283 A threat to internal validity in the nonequivalent comparison-group design is the \_\_\_\_\_ effect.

- a) Maturation effect  
b) Selection-history effect  
c) Failure to revert to baseline  
d) All of the above

284. How many variables should be changed at a time when conducting a single-case design?
- 4
  - 3
  - 2
  - 1
285. In quasi-experimental research designs, causal interpretations can be made \_\_\_\_\_.
- Only when rival explanations have been shown to be plausible
  - Only when rival explanations have been shown to be implausible
  - Only when the participants have been randomly selected
  - Only when there is a single participant in the experiment
286. In single-case research, "baseline" refers to.
- The beginning point of the treatment condition
  - The end point of the treatment condition
  - The rate of response established prior to the experimental intervention
  - The time during which a treatment condition is administered
287. Which type of design can be used when the goal is to create a step-by-step increase (or decrease) in the amount, accuracy, or frequency of some behavior over a period of time?
- Nonequivalent comparison-group design
  - A-B-A-B
  - Changing-criterion design
  - A-B design
288. Which of the following occurs in a comparison group design when one of the two groups of participants grows or naturally develops faster than the other group?
- Main effect
  - Sequencing effect
  - Order effect
  - Selection-maturation effect
289. An observation of a dependent variable response prior to any attempt to change this response is known as the.
- Flat line
  - Baseline
  - Variance
  - Reverse
290. The most frequently used quasi-experimental design is the \_\_\_\_\_ design.
- Nonequivalent comparison-group
  - Interrupted time-series
  - Changing-criterion
  - Regression discontinuity
291. A baseline.
- Is used as the standard against which change induced by the treatment is assessed
  - Is the occurrence of a response in its freely occurring or natural state
  - Is first obtained prior to the administration of a treatment
  - All of the above are true
292. A single-case experimental design in which the response to a treatment is compared to baseline occurring before and after the treatment is called what?
- A-B-A design
  - Single-case design
  - Multiple-baseline
  - Changing-criterion
293. In a single-case design, you hope that the behavior of the participants prior to the administration of a treatment condition is \_\_\_\_\_.
- Not highly variable
  - Highly variable
  - Moving at a steep rate of change
  - None of the above
294. The \_\_\_\_\_ design rules out history by demonstrating that the dependent variable response reverts back to the baseline when the treatment is withdrawn.
- Changing-criterion design
  - A-B

matching experimental and control participants on important variables

295. Which design would use analysis of covariance during data analysis?  
 a) Nonequivalent comparison-group design  
 b) Interrupted time-series design  
 c) Changing criterion design  
 d) A-B-A-B design
296. The interrupted time-series design can also be viewed as a  
 a) A-B design  
 b) A-B-A design  
 c) A-B-A-B design  
 d) Control-group design
297. Why is it important to change one variable at a time in single case designs?  
 a) Changing one variable allows isolation of the cause of the change  
 b) Changing more than one variable at a time confounds those independent variables  
 c) Both a and b are true  
 d) None of the above
298. What is the difference between A-B-A design and A-B-A-B design?  
 a) Both designs end on the treatment condition  
 b) Neither design ends on a treatment condition  
 c) Baseline conditions are only established in the A-B-A-B design  
 d) A-B-A-B allows the reintroduction of the treatment condition during the last phase
299. Which of the following is not a phase in the A-B-A design?  
 a) Baseline measurement  
 b) Introduction of treatment  
 c) Introduction of a second treatment  
 d) Removal of treatment
300. Researchers can attempt to eliminate the threat of bias from the selection-maturation effect in the nonequivalent comparison-group design by
- a)\* True  
 b) False
301. Group comparison designs are always superior to single-case designs.  
 a)\* True  
 b) False
302. The number of police officers and the number of crimes are positively related. This relationship is:  
 a) A causal relationship  
 b)\* A direct relationship  
 c) A probabilistic causal relation  
 d) A spurious relationship
303. A research studies the relation between early reading and later school achievement. She decides that a potentially extraneous variable in the relationship is IQ. In developing her groups for her study, she pairs each child who was an early reader with a child of the same IQ level who was not an early reader. The control technique she used was:  
 a) Holding the extraneous variable constant  
 b) Statistical control  
 c)\* Matching  
 d) Random assignment

#### ANSWERS

1	b	2	c	3	a	4	c
5	a	6	c	7	a	8	b
9	a	10	d	11	a	12	d
13	d	14	a	15	b	16	a
17	b	18	a	19	b	20	d
21	c	22	d	23	c	24	a
25	a	26	d	27	d	28	d
29	a	30	b	31	a	32	b
33	a	34	b	35	d	36	c
37	a	38	c	39	c	40	c
41	b	42	a	43	c	44	b
45	b	46	b	47	c	48	d
49	a	50	b	51	d	52	d
53	b	54	d	55	e	56	b
57	b	58	d	59	a	60	a
61	d	62	b	63	d	64	a
65	d	66	a	67	b	68	c

69	b	70	d	71	c	72	a	189	b	190	d	191	b	192	c
73	a	74	d	75	d	76	b	193	d	194	b	195	d	196	a
77	d	78	d	79	a	80	c	197	b	198	c	199	c	200	b
81	d	82	d	83	d	84	e	201	b	202	b	203	d	204	a
85	a	86	b	87	d	88	b	205	b	206	d	207	d	208	a
89	d	90	b	91	d	92	e	209	c	210	e	211	c	212	b
93	d	94	e	95	a	96	b	213	c	214	b	215	c	216	a
97	d	98	e	99	d	100	a	217	d	218	d	219	d	220	b
101	a	102	b	103	b	104	b	221	c	222	d	223	b	224	d
105	a	106	c	107	a	108	c	225	d	226	c	227	b	228	c
109	c	110	b	111	a	112	c	229	a	230	b	231	d	232	a
113	d	114	b	115	b	116	b	233	b	234	c	235	a	236	a
117	c	118	d	119	d	120	c	237	b	238	c	239	c	240	c
121	a	122	c	123	b	124	b	241	c	242	b	243	d	244	c
125	d	126	a	127	a	128	c	245	d	246	c	247	c	248	b
129	c	130	a	131	c	132	a	249	b	250	b	251	b	252	a
133	c	134	a	135	c	136	c	253	c	254	a	255	d	256	c
137	a	138	c	139	d	140	d	257	a	258	d	259	d	260	b
141	c	142	d	143	c	144	a	261	c	262	a	263	c	264	b
145	d	146	d	147	c	148	d	265	a	266	c	267	d	268	d
149	d	150	b	151	c	152	b	269	c	270	b	271	d	272	e
153	c	154	c	155	a	156	c	273	c	274	c	275	b	276	b
157	d	158	a	159	e	160	a	277	d	278	a	279	a	280	b
161	a	162	c	163	a	164	c	281	b	282	d	283	b	284	d
165	b	166	b	167	d	168	b	285	b	286	c	287	c	288	d
169	d	170	a	171	d	172	b	289	b	290	a	291	d	292	a
173	b	174	b	175	a	176	d	293	a	294	c	295	a	296	a
177	e	178	c	179	b	180	b	297	c	298	d	299	c	300	a
181	d	182	e	183	e	184	c	301	b	302	d	303	c		
185	d	186	a	187	d	188	c								

# STATISTICAL INFERENCE

1. Inferential statistics are used to:
  - Summarize data from single-subject designs.
  - Infer the characteristics of a population from the characteristics of a sample.
  - Help you decide whether you should replicate your study.
  - Describe your data with a few numbers.
  
2. Which of the following are the main branches of inference?
  - Estimation theory
  - Test of hypothesis
  - Non Parametric tests
  - Sequential Analysis
  - All of the above
  
3. Which of the following statements are true?
  - A point estimate of a population parameter  $\theta$  is a single number that can be regarded as a sensible value of  $\theta$ .
  - A point estimate of a population parameter  $\theta$  is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic is called the point estimator of  $\theta$ .
  - The sample mean  $\bar{X}$  is a point estimator of the population mean  $\mu$ .
  - The sample variance  $S^2$  is a point estimator of the population variance  $\sigma^2$ .
  - All of the above statements are true.
  
4. Which of the following statements is not true?
  - The symbol  $\hat{\theta}$  is customarily used to denote the estimator of parameter  $\theta$  and the point estimate resulting from a given sample.
  - The equality  $\hat{\mu} = \bar{X}$  is read as "the point estimator of  $\bar{X}$  is  $\hat{\mu}$ ".
  - The difference between  $\hat{\theta}$  and the parameter  $\theta$  is referred to as error of estimation.
  - None of the above statements is true.
  
5. Which of the following statements is not always true?
  - A point estimator  $\hat{\theta}$  is said to be an unbiased estimator of parameter  $\theta$  if  $E(\hat{\theta}) = \theta$  for every possible value of  $\theta$ .
  - If the estimator  $\hat{\theta}$  is not unbiased of parameter  $\theta$ , the difference  $E(\hat{\theta}) - \theta$  is called the bias of  $\hat{\theta}$ .
  - A point estimator  $\hat{\theta}$  is unbiased if its probability sampling distribution is always "centered" at the true value of the parameter  $\theta$ , where "centered" here means that the median of the distribution of  $\hat{\theta}$  is  $\theta$ .
  - All of the above statements are true.

Which of the following statements are true if  $X_1, X_2, \dots, X_n$  is a random sample from a distribution with mean  $\mu$  and variance  $\sigma^2$ ?

- a)  $S^2 = \frac{\sum(X - \bar{X})^2}{n+2}$  is an unbiased estimator of  $\sigma^2$
- b)  $S^2 = \frac{\sum(X - \bar{X})^2}{n+1}$  is an unbiased estimator of  $\sigma^2$
- c)  $S^2 = \frac{\sum(X - \bar{X})^2}{n}$  is an unbiased estimator of  $\sigma^2$
- d)  $S^2 = \frac{\sum(X - \bar{X})^2}{n-1}$  is an unbiased estimator of  $\sigma^2$
- e) All of the above statements are true provided that the sample size  $n > 30$ .

Hypothesis testing and estimation are both types of descriptive statistics.

- a) True
- b) False
- c) Increase the accuracy by increasing the sample size, this property is called
  - a) Sufficiency
  - b) Consistency
  - c) Efficiency
  - d) Precision

Let  $T_1$  and  $T_2$  are two unbiased estimators of a parameter  $\theta$  based on a sample of size  $n$  then we will prefer  $T_1$  if

- a)  $V(T_1) > V(T_2)$
- b)  $E(T_1) < E(T_2)$
- c)  $V(T_1) < V(T_2)$
- d)  $Cov(T_1, T_2) > 0$
- e) None of the above

16 Which of the following statements are not true?

- a) Maximizing likelihood function gives the parameter values for which the observed sample is most likely to have been generated that is, the parameter values that "agree most likely" with the observed data.
- b) Different principles of estimation may yield different estimators of the unknown parameters.
- c) The maximum likelihood estimator of the population standard deviation  $\sigma$  is the sample standard deviation  $S$ .
- d) None of the above statements are true.
- e) All of the above statements are true.

11 Which of the following statements are not true?

- a) Maximizing the likelihood function gives the parameter values for which the observed sample is most likely to have been generated that is, the parameter values that "agree most likely" with the observed data.
- b) Different principles of estimation may yield different estimators of the unknown parameters.
- c) The maximum likelihood estimator of the population standard deviation  $\sigma$  is the sample standard deviation  $S$ .
- d) None of the above statements are true.

12 Which of the following statements are true?

- a) Maximizing the likelihood estimation is the most widely used estimation technique among statisticians.
- b) Under very general conditions on the joint distribution of the sample when the sample size  $n$  is large, the maximum likelihood estimator of any parameter  $\theta$  is approximately unbiased, that is,

$$E(\hat{\theta}) \approx \theta$$

- 6) Under very general conditions on the joint distribution of the sample, when the sample size  $n$  is large, the maximum likelihood estimator of any parameter  $\theta$  has variance, is nearly as small as small as can be achieved by any estimator.
- 7) In recent years, statisticians have proposed an estimator, called an M-estimator, which is based on a generalization of maximum likelihood estimation.
- 8) All of the above are true.
- 9) statements.
- 10) Based on  $n$  i.i.d. observations  $X_1, X_2, \dots, X_n$  from  $N(\mu, \sigma^2)$  the MLE of  $\sigma^2$  is
- Always unbiased and sufficient for  $\sigma^2$
  - Unbiased and sufficient when  $\mu$  is known
  - Never unbiased but always sufficient
  - Unbiased but not sufficient when  $\mu$  is known
  - None of the above
- 11) Let  $X_1, X_2, \dots, X_n$  from the pdf  $f(x, \theta) = \theta x^{\theta-1}, 0 < x < 1 \& \theta > 0$  then
- $\sum_{i=1}^n x_i$  is sufficient statistic for  $\theta$
  - $\prod_{i=1}^n x_i$  is sufficient statistic for  $\theta$
  - $\frac{\sum_{i=1}^n x_i}{n}$  is sufficient statistic for  $\theta$
  - None of the above
- 12) A random sample is selected from a population of measurements. The mean of the sample is not equal to the mean of the population. This is due to
- Type I error
  - Type II error
  - Sampling error
  - Abnormal population
  - Inexact measurements
- 13) The amount of variability in expected sample means across a series of samples is estimated by the
- Standard deviation
  - Population variance
  - Standard error of the mean
  - Inter quartile range
- 14) A population of size 1,000,000 has mean 42 and standard deviation 6. Sixty random samples of size 15 are selected. According to the central limit theorem, the distribution of the 60 sample means has a mean of approximately:
- 42
  - $\frac{42}{6}$
  - $\frac{42}{15}$
  - $\frac{42}{\sqrt{15}}$
  - None of the above
- 15) A population of size 1,000,000 has mean 42 and standard deviation 6. Sixty random samples of size 15 are selected. According to the central limit theorem, the distribution of the 60 sample means has a standard deviation of approximately:
- 42
  - $\frac{42}{6}$
  - $\frac{42}{15}$
  - $\frac{42}{\sqrt{15}}$
  - None of the above
- 16) The probability is 0.2 that a term selected at random from a normal distribution with mean 600 and

standard deviation 15 will be above what number?

- a) 0.84
- b) 803.80
- c) 812.8
- d) 887.4
- e) 818.8

10. A \_\_\_\_\_ from a sample is used to estimate a population \_\_\_\_\_. The two words that best fill these blanks are:

- a) Item, value.
- b) Value, statistic.
- c) Statistic, parameter.
- d) Parameter, value.
- e) Parameter, statistic.

11. An assumption underlying parametric statistics is that:

- a) Sampling was done from a normally distributed population.
- b) Your data were measured on a nominal or an ordinal.
- c) Your data need not meet any strict requirements.
- d) Both a) and b)
- e) None of these

12. If your independent variable has no effect on the dependent variable, the distributions representing the different groups in your experiment:

- a) Represent two distinct populations.
- b) Are independent samples drawn from the same population.
- c) Are probably positively skewed.
- d) Are probably negatively skewed.

13. An estimator

$\hat{\theta}$ . (unbiased) for  $\theta$  is said to be a consistent estimator if

- a)  $Var(\hat{\theta}) = 0$
- b)  $Var(\hat{\theta}_n) = 0$
- c)  $\lim_{n \rightarrow \infty} Var(\hat{\theta}_n) = 0$

24. The part of likelihood function involving parameter is called

- a) Joint probability
- b) Pivotal quantity
- c) Kernel
- d) Support
- e) All of the above
- f) None of the above

25. The first derivative of likelihood or log likelihood function is called

- a) Maximum likelihood estimation
- b) Support
- c) Score
- d) Information Matrix

26. The diagonal elements of information matrix are called

- a) Loadings
- b) Standard deviations of the estimates
- c) Variances of the estimates
- d) Standard error of the estimates
- e) None of these

27. The difference between reality and our perception of reality is called

- a) Bias
- b) Systematic error
- c) Error
- d) None of these

28. The probability of rejecting the null hypothesis, when null hypothesis is false, is called

- a) Producer risk
- b) Consumer risk
- c) Power of the test
- d) Type II error
- e) All of the above
- f) None of these

29. The difference between sample statistics and population parameter is called

- a) Systematic error
- b) Bias
- c) MSE
- d) Sampling error
- e) All of these
- f) None of these

30. The formula for determining the sample size  $n$  is
- $Z^2 \sigma^2 / E^2$
  - $Z^2 \sigma^2 / M^2$
  - $[Z^2 \sigma^2 / M^2]^{1/2}$
  - $[Z \sigma / E]^2$
  - None of these
31. Mandla's statistic is a test for
- Variance
  - Average
  - Multivariate normality
  - None of these
32. Method of analyzing a sample data to make a decision is
- Critical value method
  - Confidence interval method
  - P-value
  - All of the above
  - None of these
33. Whenever data is analyzed without planning before data collection then
- One tail test is used
  - Two tail test is used
  - Both a) and b)
  - None of these
34. All hypothesis tests start with
- A null hypothesis
  - Alternative hypothesis
  - Test statistic
  - Type I-Error
35. If the experimental units are relatively heterogeneous and large positive correlation with in pairs, then which of the following test is used for testing the equality of two population means.
- T-test
  - Z-test
  - Paired T-test
  - All of the above
  - None of the above
36. In testing of hypothesis, if  $H_0: \mu = \mu_0$ , then this type of test is called
- One tail test
  - Two tailed test
  - Both a) and c)
  - None of these
37. A range of values that contain the true value with high probability is called
- Confidence coefficient
  - Estimator
  - Confidence interval
  - None of these
38. The main objective of non central distributions is
- To minimize variance
  - Testing of hypothesis
  - Both a) and b)
  - None of these
39. A statement about the population parameter (unknown), which may or may not be true, is called
- Test statistics
  - Statistical hypothesis
  - Null hypothesis
  - Alternative hypothesis
40. If the statement about the population parameter is completely specified is called
- Null hypothesis
  - Simple hypothesis
  - Composite hypothesis
  - None of these
41. A functional part of statistical hypothesis which leads us to accept or reject the null hypothesis is called
- Null hypothesis
  - Simple hypothesis
  - Composite hypothesis
  - Test statistics
42. The paired t test is
- Impractical for large samples
  - Equivalent to a chi-squared test
  - Suitable for very small samples
  - Used for independent samples
  - Requires the assumption that differences between paired observations follow a Normal Distribution

43.  $S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$  is \_\_\_\_\_

- estimator of population Variance  $\sigma^2$
- Biased
  - Unbiased
  - Both a) and b)
  - None of these

44. The hypothesis that says that your sample means were drawn from the same population is the
- Alternative hypothesis.
  - Central limit hypothesis.
  - Null hypothesis.
  - Post hoc hypothesis.

45. If the probability that the difference between sample means could have resulted by sampling the same population is sufficiently small, then we say that the difference between means is:
- Not statistically significant.
  - Statistically significant.
  - Valid.
  - None of the above

46. Which of the following statements are true?
- A confidence interval is always calculated by first selecting a confidence level, which is a measure of the degree of reliability of the interval.
  - A confidence level of 95% implies that 95% of all samples would give an interval that includes the parameter being estimated, and only 5% of all samples would yield an erroneous interval.
  - Information about the precision of an interval estimate is conveyed by the width of the interval.
  - The higher the confidence level, the more strongly we believe that the value of the parameter being estimated lies within the interval.
  - All of the above statements are true.

47. Which of the following statements are true?

- The interval  $\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$  is random, while its width is not random.
- The interval  $\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$  is not random, while its width is random.
- The interval  $\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$  is random, while its width is not random.
- The interval  $\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$  is not random, while its width is random.
- None of the above statements are true.

48. Which of the following statements are not true?

- A correct interpretation of a  $100(1-\alpha)\%$  confidence interval for the mean  $\mu$  relies on the long-run frequency interpretation of probability.
- It is correct to write a statement such as  $P[\mu \text{ lies in the interval } (70, 80)] = .95$
- The probability is .95 that the random interval  $\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$  includes or covers the true value of  $\mu$

- The interval  $\bar{X} \pm 1.645 \frac{\sigma}{\sqrt{n}}$  is a 90% confidence interval for the mean  $\mu$ .
- None of the above statements are true.

49. A 99% confidence interval for the mean  $\mu$  of a normal population when the standard deviation  $\sigma$  is known is found to be 98.6 to 118.4. If the

- confidence level is reduced to .95, the confidence interval for  $\mu$
- Becomes wider
  - Becomes narrower
  - Remains unchanged
  - None of the above answers are correct.

50. If the width of a confidence interval for  $\mu$  is too wide when the population standard deviation  $\sigma$  is known, which one of the following is the best action to reduce the interval width?
- Increase the confidence level
  - Reduce the population standard deviation  $\sigma$
  - Increase the population mean  $\mu$
  - Increase the sample size  $n$
  - None of the above answers are correct.

51. Which of the following statements are not true?

- a) Provided that the sample size  $n$  is large, the standardized variable

$$Z = \frac{(\bar{X} - \mu)}{\sigma} / \sqrt{n}$$

approximately normally distributed,

while the variable

$$Z = \frac{(\bar{X} - \mu)}{S} / \sqrt{n}$$

- b) The formula  $\bar{X} \pm Z_{\alpha/2} \frac{S}{\sqrt{n}}$  is a

large-sample confidence interval for  $\mu$  with confidence level

approximately  $100(1 - \alpha)\%$ .

- c) Generally speaking,  $n > 40$  will be sufficient to justify the use of the

$$\text{formula } \bar{X} \pm Z_{\alpha/2} \frac{S}{\sqrt{n}}$$

as a large-sample confidence interval for  $\mu$ .

- d) None of the above statements are true.
- e) All of the above statements are true.

52. If one wants to develop a 90% confidence interval for the mean  $\mu$  of a normal population, when the standard deviation  $\sigma$  is known, the confidence level is
- 0.10
  - 0.45
  - 0.90
  - 1.645

53. Which of the following expressions is true about a large-sample upper confidence bound for the population mean  $\mu$ ?

a)  $\mu < \bar{X} - Z_{\alpha/2} S \sqrt{n}$

b)  $\mu < \bar{X} + Z_{\alpha/2} S \sqrt{n}$

c)  $\mu < \bar{X} - Z_{\alpha} S \sqrt{n}$

d)  $\mu < \bar{X} + Z_{\alpha} S \sqrt{n}$

- e) None of the above statements are true.

54. Suppose that an investigator believes that virtually all values in the population are between 38 and 70. The appropriate sample size for estimating the true population mean  $\mu$  within 2 units with 95% confidence level is approximately

- a) 61

- b) 62

- c) 15

- d) 16

- e) None of the above answers are correct.

55. You need to construct a large sample 94% confidence interval for a population mean. What is the upper critical value of  $z$  to be used in constructing this interval?

- a) 0.9699

- b) 1.96

- c) 1.555  
d) -1.88  
e) 1.88

56. A 99% confidence interval for the population mean  $\mu$  is determined to be (65.32 to 73.54). If the confidence level is reduced to 90%, the 90% confidence interval for  $\mu$
- Becomes wider
  - Becomes narrower
  - Remains unchanged
  - None of the above answers are correct.

57. Which of the following statements are not true in developing a confidence interval for the population mean  $\mu$
- The width of the confidence interval becomes narrower when the sample mean increases.
  - The width of the confidence interval becomes wider when the sample mean increases.
  - The width of the confidence interval becomes narrower when the sample size  $n$  increases.
  - All of the above statements are true.
  - None of the above statements are true.

58. Which of the following statements are true when  $\bar{X}$  is the mean of a random sample of size  $n$  from a normal distribution with mean  $\mu$ ?

- a) The random variable

$$Z = \frac{(\bar{X} - \mu)}{\sqrt{s}} \text{ has}$$

approximately a standard normal distribution for large  $n$ .

- b) The random variable

$$T = \frac{(\bar{X} - \mu)}{\sqrt{s}} \text{ has a t-distribution with } n-1 \text{ degrees of freedom for small } n.$$

distribution with  $n-1$  degrees of freedom for small  $n$ .

- c) The normal distribution is governed by two parameters, the mean  $\mu$  and the standard deviation  $\sigma$ .
- d) A  $t$ -distribution is governed by only one parameter, called the number of degrees of freedom.
- e) All of the above answers are true.

59. Which of the following statements are not true?

- a) The notation  $t_{\alpha, v}$  is often used to denote the number on the measurement axis for which the area under the  $t$ -curve with  $v$  degrees of freedom to the left of  $t_{\alpha, v}$  is  $\alpha$  where  $t_{\alpha, v}$  is called a  $t$  critical value.
- b) The number of degrees of freedom for a  $t$ -variable is the number of freely determined deviations  $X_i - \bar{X}$  on which the estimated standard deviation in the denominator of
- $$T = \frac{(\bar{X} - \mu)}{\sqrt{s}}$$
- is based.

- c) A larger value of degrees of freedom  $v$  implies a  $t$ -distribution with smaller spread.
- d) All of the above statements are true.
- e) None of the above statements are true.

60. Serum cholesterol levels for two groups of Americans were recorded in 1989. The mean cholesterol levels of the two groups were compared. To determine whether the measurements were significantly different or not, the most appropriate statistical test would be:

- Chi-square test
- Correlation analysis
- F test (ANOVA)
- Student's  $t$  test
- Regression analysis

61. Probability of rejecting null hypothesis when it is true, called  
 a) Type I error  
 b) Producer Risk  
 c) Consumer Risk  
 d) Both a) and b)
62. Probability of not rejecting null hypothesis  $H_0$ , when  $H_1$  is true, called  
 a) Type I error  
 b) Type II error  
 c) Consumer risk  
 d) Both b) and c)
63. If you take steps to minimize a Type I error, then the probability of making a Type II error is  
 a) Increased.  
 b) Also decreased.  
 c) Unaffected.  
 d) Cut in half.
64. One-tailed tests should be used  
 a) Whenever you are unsure what kind of test to use.  
 b) In any situation in which you cannot predict the direction of an effect.  
 c) Only if there is some compelling a priori reason not to use a two-tailed test.  
 d) When nonparametric statistics are used.
65. Which of the following statements are not true?  
 a) A statistical hypothesis is a claim or assertion either about the value of a single parameter, about the values of several parameters, or about the form of an entire probability distribution.  
 b) In any hypothesis-testing problem, there are two contradictory hypotheses under consideration.  
 c) A test of hypothesis is a method for using sample data to decide whether the null hypothesis should be rejected.
66. Which of the following statements are true?  
 a) The null hypothesis, denoted by  $H_0$ , is the claim that is initially assumed to be true (the "prior belief" claim).  
 b) The alternative hypothesis, denoted by  $H_a$ , is the assertion that is opposing to the null hypothesis  $H_0$ .  
 c) The null hypothesis  $H_0$  will be rejected in favor of the alternative hypothesis only if sample evidence suggests that  $H_0$  is false.  
 d) If sample evidence does not strongly contradict the null hypothesis  $H_0$ , we will continue to believe in the truth of  $H_0$ .  
 e) All of the above statements are true.
67. Which of the following statements are not correctly stated?  
 a) The two possible conclusions from a hypothesis-testing analysis are rejecting the null Hypothesis  $H_0$  or accepting  $H_0$ .  
 b) In many situations, the alternative hypothesis  $H_a$  is referred to as the "research hypothesis" since it is the claim that the researcher would really like to validate.  
 c) In our treatment of hypothesis testing, the null hypothesis will always be stated as an equality claim.

- d) A test statistic is a rule, based on sample data, for deciding whether to reject the null hypothesis.
- e) All of the above statements are correctly stated.
68. Which of the following statements are not correct?
- It is possible that the null hypothesis may be rejected when it is true.
  - It is impossible that the null hypothesis may be rejected when it is true.
  - It is possible that the null hypothesis may not be rejected when it is false.
  - All of the above statements are correct.
  - None of the above statements are correct.
69. If  $\theta$  denotes the parameter of interest, and the simplified null hypothesis has the form  $H_0 : \theta = \theta_0$ , where  $\theta_0$  is a specified number called the "null value" of the parameter, then the alternative hypothesis will be
- $H_a : \theta > \theta_0$  (so the implicit null hypothesis is  $\theta \leq \theta_0$ )
  - $H_a : \theta > \theta_0$  (so the implicit null hypothesis is  $\theta \geq \theta_0$ )
  - $H_a : \theta = \theta_0$
  - The alternative hypothesis will look like any of the above three assertions.
  - The alternative hypothesis must be the assertion specified in c).
70. Which of the following statements are not true?
- A test statistic is a function of the sample data on which the decision to reject or not reject the null hypothesis is to be based.
  - A rejection region consists of the set of all test statistic values for which the null hypothesis will be rejected.
  - A rejection region consists of the set of all test statistic values for which the alternative hypothesis will be rejected.
  - A good hypothesis-testing procedure is one for which the probability of making either type I or type II error is small.
  - None of the above statements are true.
71. Which of the following statements are not true?
- A rejection region is called upper-tailed if it consists only of large values of the test statistic.
  - A rejection region is called upper-tailed if it consists only of small values of the test statistic.
  - A rejection region is called lower-tailed if it consists only of small values of the test statistic.
  - All of the above statements are true.
  - None of the above statements are true.
72. Which of the following statements are true?
- The probability of type I error  $\alpha$ , is computed using the probability distribution of the test statistic when the null hypothesis is true.
  - The probability of type II error  $\beta$ , requires knowing the distribution of the test statistic when the null hypothesis is false.
  - The probability of type I error  $\alpha$ , is computed by summing over probabilities of test statistic values in the rejection region.
  - The probability of type II error  $\beta$ , is computed by summing over probabilities of test statistic values in the complement of the rejection region.
  - All of the above statements are true.

- 134** Which of the following statements are not generally true?
- A type I error is usually more serious than a type II error.
  - A type II error is usually more serious than a type I error.
  - A test with significance level  $\alpha$  is one for which the type I error probability is controlled at the specified level.
  - When an experiment and a sample size are fixed, then decreasing the size of the rejection region to obtain a smaller value of  $\alpha$  (probability of type I error) results in a larger value of  $\beta$  (probability of type II error) for any particular parameter value consistent with the alternative hypothesis  $H_a$ .
  - None of the above statements are true.
- 73.** The most appropriate statistical test for an experiment with two independent groups and the dependent variable measured on an interval scale is:
- Chi-square.
  - The t test for independent samples.
  - The one-sample z test.
  - A two-factor ANOVA.
- 75.** Which of the following statements are not true?
- If the null hypothesis is  $H_0 : \mu = 50$ ,  $\bar{x} = 53$  and  $\sigma_x = \frac{\sigma}{\sqrt{n}} = 1.2$ , then the test statistic value is  $t = -2.5$ .
  - If the alternative hypothesis has the form  $H_a : \mu > \mu_0$ , then an  $\bar{x}$  value less than  $\mu_0$  certainly does not provide support for  $H_a$ .
  - If the alternative hypothesis has the form  $H_a : \mu > \mu_0$ , then
- $H_0 : \mu = 50$ ,  $\bar{x} = 53$  and  $\sigma_x = \frac{\sigma}{\sqrt{n}} = 1.2$ , then the test statistic value is  $t = -2.5$ .
- 76.** Which of the following statements are true?
- When the alternative hypothesis is  $H_a : \mu < \mu_0$ , the null hypothesis  $H_0$  should be rejected if  $\bar{x}$  is too far to the left of  $\mu_0$ .
  - When the alternative hypothesis is  $H_a : \mu > \mu_0$ , the null hypothesis  $H_0$  should be rejected if  $\bar{x}$  is too far to the right of  $\mu_0$ .
  - When the alternative hypothesis is  $H_a : \mu \neq \mu_0$ , the null hypothesis  $H_0$  should be rejected if  $\bar{x}$  is too far to either side of  $\mu_0$ .
  - All of the above statements are true.
  - None of the above statements are true.
- 77.** Which of the following statements are not true for any significance level?
- The one-tailed rejection region  $Z \geq Z_\alpha$  has type I error probability  $\alpha$
  - The one-tailed rejection region  $Z \geq -Z_\alpha$  has type I error probability  $\alpha$
  - The two-tailed rejection region  $Z \geq Z_{\frac{\alpha}{2}}$  or  $Z \leq -Z_{\frac{\alpha}{2}}$  has type I error probability  $\frac{\alpha}{2}$ , since area

- $\frac{\alpha}{2}$  is captured under each of the two tails of the z curve.
- d) All of the above statements are true.
- e) None of the above statements are true.
78. Which of the following statements are not true if a test procedure about the population mean  $\mu$  is performed when the population is normal with known standard deviation  $\sigma$ ?
- The rejection region for level  $\alpha$  test is  $Z \geq Z_{\alpha}$ , if the test is an upper-tailed test.
  - The rejection region for level  $\alpha$  test is  $Z \geq Z_{\frac{\alpha}{2}}$ , if the test is an upper-tailed test.
  - The rejection region for level  $\alpha$  test is  $Z \leq -Z_{\alpha}$ , if the test is a lower-tailed test.
  - The rejection region for level  $\alpha$  test is either  $Z \geq Z_{\frac{\alpha}{2}}$  or  $Z \leq -Z_{\frac{\alpha}{2}}$ , if the test is a two-tailed test.
  - None of the above statements are true.
79. Which of the following statements are true in testing  $H_0: \mu = 275$  versus  $H_a: \mu > 275$  based on a sample of size 15 from a normal population with unknown standard deviation  $\sigma$ ?
- The test procedure requires the use of standard normal distribution.
  - The test procedure requires the use of binomial distribution.
  - The test procedure requires the use of exponential distribution.
  - The test procedure requires the use of t-distribution with 15 degrees of freedom.

- e) None of the above statements are true.
80. Which of the following statements are true?
- Knowledge of the test statistic's distribution when the null hypothesis is true allows us to construct a rejection region for which the type I error probability is controlled at the desired level.
  - The rejection region for the t test differs from that for the z test only in that a t critical value replaces the z critical value.
  - If  $X_1, X_2, \dots, X_n$  ( $n$  is small) is a random sample from a normal distribution, the standardized variable  $T = \frac{(\bar{X} - \mu)}{S/\sqrt{n}}$  has a t distribution with  $n-1$  degrees of freedom.
  - All of the above statements are true.
  - None of the above statements are true.
81. Suppose that a two-tailed test procedure about the population mean  $\mu$  is performed when the population is normal, but the sample size  $n$  is small. The null hypothesis will be rejected at significance level  $\alpha$  if the value of the standardized test statistic  $t$  is such that
- $t \geq t_{\alpha/2, n-1}$
  - $t \leq t_{\alpha/2, n-1}$
  - Either  $t \geq t_{\alpha/2, n-1}$  or  $t \leq -t_{\alpha/2, n-1}$
  - $-t_{\alpha/2, n-1} \leq t \leq t_{\alpha/2, n-1}$
  - None of the above inequalities are correct
82. Let  $p$  denotes the proportion of individuals in a population who possess a specified property, and let  $X$  denote the number of individuals in

- the sample which possess the same property. Provided that the sample size  $n$  is small relative to the population size,  $X$  has approximately
- A normal distribution
  - A binomial distribution
  - An exponential distribution
  - A Poisson distribution
83. Let  $p$  denotes the proportion of individuals in a population who possess a specified property, and let  $X$  denote the number of individuals in the sample who possess the same property. Provided that the sample size  $n$  is large, the estimator  $\hat{p} = \frac{X}{n}$  has approximately
- A normal distribution
  - A binomial distribution
  - An exponential distribution
  - A Poisson distribution
84. Which of the following statements are not true?
- If the alternative hypothesis has the form  $H_a : p > p_0$ , then a  $\hat{p}$  value less than  $p_0$  certainly does not provide support for  $H_a$ .
  - If the alternative hypothesis has the form  $H_a : p > p_0$ , then a  $\hat{p}$  value exceeds  $p_0$  by only a small amount (corresponding to z value which is positive but small) does not suggest that  $H_0$  should be rejected in favor of  $H_a$ .
  - If the null hypothesis is  $H_0 : p = 0.50$ ,  $\hat{p} = 0.51$  and  $n = 100$ , then the test statistic value is  $z = 4$
  - All of the above statements are true.
  - None of the above statements are true.
85. Which of the following statements are true?
- When the alternative hypothesis is  $H_a : p < p_0$ , the null hypothesis  $H_0$  should be rejected if  $\hat{p}$  is far to the left of  $p_0$ .
  - When the alternative hypothesis is  $H_a : p > p_0$ , the null hypothesis  $H_0$  should be rejected if  $\hat{p}$  is far to the right of  $p_0$ .
  - When the alternative hypothesis is  $H_a : p \neq p_0$ , the null hypothesis  $H_0$  should be rejected if  $\hat{p}$  is far to either side of  $p_0$ .
  - All of the above statements are true.
  - None of the above statements are true.
86. Which of the following statements are not true?
- A p-value conveys much information about the strength of evidence against the null hypothesis  $H_0$  and allows an individual decision maker to draw a conclusion at any specified significance level  $\alpha$ .
  - The p-value (or observed significance level) is the largest level of significance at which the null hypothesis  $H_0$  would be rejected when a specified test procedure is used on a given data set.
  - If  $P - \text{value} \leq \alpha$ , we reject  $H_0$  at level  $\alpha$ .
  - If  $P - \text{value} > \alpha$ , we reject  $H_0$  at level  $\alpha$ .
  - All of the above statements are true.

87. Which of the following statements are true?
- It is customary to call the data significant when  $H_0$  is rejected.
  - It is customary to call the data not significant when  $H_0$  is not rejected.
  - The calculation of the p-value depends on whether the test is upper-tailed, lower-tailed, or two-tailed.
  - The p-value for a z test (one based on a test statistic whose distribution when  $H_0$  is true is at least approximately standard normal) is easily determined from the information in the standard normal probability table.
  - All of the above are true statements.

88. Suppose that when data from an experiment was analyzed, the p-value for testing  $H_0 : \mu = 50$  versus  $H_0 : \mu > 50$  was calculated as 0.0244. Which of the following statements are true?
- $H_0$  is rejected at .10 level.
  - $H_0$  is not rejected at .05 level.
  - $H_0$  is not rejected at 0.025 level.
  - $H_0$  is rejected at any level  $\alpha$ .
  - All of the above statements are true.

89. Which of the following statements are true if the value of the test statistic for a two-tailed z test is  $z = -1.56$ ?
- p-value = 0.4406
  - p-value = 0.0594
  - p-value = .1188
  - $0.0594 < \text{p-value} < 0.1188$
  - p-value = 0.0406

90. Which of the following statements are true about the p-value, where  $z$  is the calculated value of the test statistic and  $\varphi(z)$  is the corresponding

cumulative area under the standard normal curve?

- $P\text{-value} = 1 - \varphi(z)$  for an upper-tailed test.
- $P\text{-value} = \varphi(z)$  for a lower-tailed test.
- $P\text{-value} = 2[1 - \varphi(z)]$  for a lower-tailed test
- All of the above statements are true.
- None of the above statements are true.

91. Suppose that a t test of  $H_0 : \mu = 250$  versus  $H_a : \mu \neq 250$  is based on 12 degrees of freedom. If the calculated value of the test statistic is 2.8, then the p-value is
- 0.008
  - 0.992
  - 0.016
  - 0.492
  - 0.496

92. Which of the following statements are necessary to construct an appropriate test procedure?
- Specify a test statistic.
  - Decide on the general form of the rejection region.
  - Select the specified numerical critical value or values that will separate the rejection region from the acceptance region.
  - All of the above statements are necessary
  - Only A and B are necessary statements.

93. Which of the following statements are not true?
- The test statistic  $Z = \frac{(\bar{X} - \mu_0)}{\sigma / \sqrt{n}}$  has a standard normal distribution when  $H_0 : \mu = \mu_0$  is true.

- b) The reliability of hypothesis testing procedure in reaching a correct decision can be assessed by studying type I error probability.
- c) The process of reaching a decision by using the methodology of classical hypothesis testing involves selecting a level of significance  $\alpha$  and then rejecting or not rejecting the null hypothesis  $H_0$  at that level  $\alpha$ .
- d) All of the above statements are true.
- e) None of the above statements are true.

94. Which of the following statements are true?

- a) When the results of an experiment are to be communicated to a large audience, rejection of  $H_0$  at level  $\alpha$  will be much more convincing, if the observed value of the test statistic greatly exceeds the  $\alpha$  % critical value than if it barely exceeds that value.
- b) A large p-value would indicate statistical significance in that it would strongly suggest rejection of  $H_0$  in favor of  $H_a$ .
- c) In many experimental situations, only departures from  $H_0$  of small magnitude would be worthy of detection, whereas a large departure from  $H_0$  would have little practical significance.
- d) All of the above statements are true.
- e) None of the above statements are true.

95. Let  $X_1, X_2, \dots, X_m$  be a random sample from a normal population with mean  $\mu_1$  and known variance  $\sigma_1^2$ , and let  $Y_1, Y_2, \dots, Y_n$  be a random sample from a normal population with mean  $\mu_2$  and variance  $\sigma_2^2$  and that

the  $X$  and  $Y$  samples are independent of one another. Which of the following statements are true?

- a)  $\bar{X}$  is normally distributed with expected value  $\mu_1$  and variance  $\frac{\sigma_1^2}{m}$
- b)  $\bar{Y}$  is normally distributed with expected value  $\mu_2$  and variance  $\frac{\sigma_2^2}{n}$
- c)  $\bar{X} - \bar{Y}$  is normally distributed with expected value  $\mu_1 - \mu_2$  and variance  $\left( \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n} \right)$
- d)  $\bar{X} - \bar{Y}$  is an unbiased estimator of  $\mu_1 - \mu_2$ .
- e) All of the above statements are true.

96. Which of the following statements are true?

- a) When the alternative hypothesis is  $H_a : \mu_1 - \mu_2 < \Delta_0$  the null hypothesis  $H_0$  should be rejected if  $\bar{x} - \bar{y}$  is considerably smaller than the null value  $\Delta_0$ .
- b) When the alternative hypothesis is  $H_a : \mu_1 - \mu_2 > \Delta_0$  the null hypothesis should be rejected if  $\bar{x} - \bar{y}$  is considerably larger than the null value  $\Delta_0$ .
- c) When the alternative hypothesis is  $H_a : \mu_1 - \mu_2 \neq \Delta_0$  the null hypothesis should be rejected if  $\bar{x} - \bar{y}$  is too far to either side of the null value  $\Delta_0$ .
- d) All of the above statements are true.
- e) None of the above statements are true.

87 In testing  $H_0: \mu_1 - \mu_2 = 0$  versus  $H_1: \mu_1 - \mu_2 \neq 0$  the computed value of the test statistic is  $z = 1.98$ . The  $p$ -value for this two-tailed test is then

- a) 0.4761
- b) 0.0478
- c) 0.0239
- d) 0.2381
- e) 0.2619

88 In calculating 95% confidence interval for  $\mu_1 - \mu_2$ : the difference between the means of two normally distributed populations, summary statistics from two independent samples are:

$$m = 60, \bar{x} = 150, s_1^2 = 360, n = 45, \bar{y} = 160$$

Then, the lower limit of the confidence interval is:

- a) 29.994
- b) 11.587
- c) 10.006
- d) 10.797
- e) 28.413

89 In calculating 95% confidence interval for  $\mu_1 - \mu_2$ : the difference between the means of two normally distributed populations, summary statistics from two independent samples are:

$$m = 10, \bar{x} = 50, s_1^2 = 0.64, n = 10, \bar{y} = 40$$

Then, the upper limit of the confidence interval is:

- a) 10.953
- b) 9.047
- c) 9.216
- d) 10.784
- e) 10.0

90 Which of the following statements are true?

- a) In real problems, it is virtually always the case that the values of the population variances are unknown.
- b) The two-sample  $t$  test is applicable in situations in which population distributions are both normal when population variances have

- c) unknown values, and at least one of the two sample sizes are small.
- d) The pooled  $t$  test procedure is applicable if the two population distribution curves are assumed normal with equal spreads.
- e) All of the above statements are true.
- f) None of the above statements are true.

101. Which of the following statements are not true?

- a) Many statisticians recommend pooled  $t$  procedures over the two-sample  $t$  procedures.
- b) The pooled  $t$  test is not a likelihood ratio test, whereas the two-sample  $t$  test can be derived from the likelihood ratio principle.
- c) The significance level for the pooled  $t$  test is exact.
- d) The significance level for the two-sample  $t$  test is only approximate.
- e) All of the above statements are true

102. The degrees of freedom associated with the pooled  $t$  test, based on sample sizes 10 and 12 are

- a) 22
- b) 21
- c) 20
- d) 19

$$\text{and } s_2^2 \leq 1.86$$

103. Which of the following statements are not correct assumptions for developing pooled confidence intervals and for testing hypotheses about the difference between two population means  $\mu_1 - \mu_2$ ?

- a) Both populations are normally distributed
- b) The samples selected from the two populations are independent random samples.
- c) At least one of the two sample sizes is small.
- d) The two population variances are equal  $\sigma_1^2 = \sigma_2^2$ .
- e) The two population variances are not equal  $\sigma_1^2 \neq \sigma_2^2$

104. When variances  $s_1^2$  and  $s_2^2$  of two independent samples are combined and  $s^2$  is computed, the  $s^2$  is referred to as

- a) The pooled estimator of  $s^2$
- b) The combined estimator of  $s^2$
- c) The pooled estimator of the common variance  $\sigma^2$  of the two populations
- d) The adjusted estimator of  $s^2$
- e) None of the above answers are correct.

105. Two independent samples of sizes 15 and 17 are randomly selected from two normal populations with equal variances. Which of the following distributions should be used for developing confidence intervals and for testing hypotheses about the difference between the two population means  $\mu_1 - \mu_2$ ?

- a) The standard normal distribution
- b) The  $t$  distribution with 32 degrees of freedom
- c) The  $t$  distribution with 31 degrees of freedom
- d) The  $t$  distribution with 30 degrees of freedom
- e) Any continuous distribution since the sum of the two sample sizes exceeds 30

106. The number of degrees of freedom for a paired  $t$  test, where the data consists of 10 independent pairs, is equal to

- |       |       |
|-------|-------|
| a) 20 | b) 18 |
| c) 10 | d) 9  |
| e) 8  |       |

107. At the .05 significance level, the null hypothesis  $H_0: \mu_D \geq 0$  is rejected in a paired  $t$  test, where the data consists of 15 independent pairs, if

- a)  $t < 1.761$
- b)  $t > 1.761$
- c) either  $t < 1.761$  or  $t > 1.761$
- d)  $t < 1.701$
- d)  $t > 1.701$

108. Which of the following statements are true?

- a) Whenever there is positive dependence within pairs, the denominator for the paired  $t$  statistic should be smaller than that of the independent-samples test.
- b) When data is paired, the paired  $t$  confidence interval will usually be narrower than the (incorrect) two-sample  $t$  confidence interval.
- c) If there is great heterogeneity between experimental units and a large correlation within experimental units, a paired experiment is preferable to an independent-samples experiment.
- d) If the experimental units are relatively homogeneous and the correlation within pairs is not large, an independent-samples experiment should be used.
- e) All of the above statements are true.

109. In testing the difference between two population proportions  $\hat{p}_1 - \hat{p}_2$ , a weighted average  $\hat{p}$  of the sample proportions  $\hat{p}_1$  and  $\hat{p}_2$  should be used in computing the value of the test statistic when

- a) The two populations are normally distributed
- b) The two sample sizes are small
- c) The two samples are independent of each other
- d) The null hypothesis states that the two population proportions are equal
- e) The null hypothesis states that the two sample proportions are equal

110. If we want to compare the homogeneity of two variances, then which of the following test is used?

- a) Chi-square test
- b) T-test
- c) F-test
- d) DMRT

111. Then number of scores that can vary in a distribution with a known mean is the definition for:
- A sampling distribution
  - The standard error of the mean
  - Free variability
  - Degrees of freedom
112. Which of the following statements are not true about the  $F$  distribution with parameters  $v_1$  and  $v_2$ ?
- The parameter  $v_1$  is called the number of numerator degrees of freedom.
  - The parameter  $v_2$  is called the number of denominator degrees of freedom.
  - A random variable that has an  $F$  distribution can assume a negative value; depends on the values of  $v_1$  and  $v_2$ .
  - All of the above statements are true.
  - None of the above statements are true.
113. For an  $F$  distribution with parameters  $v_1$  and  $v_2$ , where  $v_1$  is the number of numerator degrees of freedom, and  $v_2$  is the number of denominator degrees of freedom, which of the following statements are true?
- $v_1$  must be larger than  $v_2$
  - $v_1$  must be smaller than  $v_2$
  - $v_1$  must equal  $v_2$
  - $v_1$  can be larger than, smaller than, or equal to  $v_2$
  - None of the above answers are true.
114. Let  $X_1, X_2, \dots, X_{20}$  be a random sample from a normal distribution with variance  $\sigma_1^2$ , let  $Y_1, Y_2, \dots, Y_{25}$  be another random sample (independent
- of the  $X_i$ 's) from a normal distribution with variance  $\sigma_2^2$ . Which of the following statements are not true in testing  $H_0: \sigma_1^2 = \sigma_2^2$ , where the test statistic value is  $f = \frac{S_1^2}{S_2^2}$  and the test is performed at 10% levels?
- The rejection region is  $f \geq F_{0.10, 19, 24}$  if  $H_0: \sigma_1^2 > \sigma_2^2$
  - The rejection region is  $f \leq F_{0.90, 19, 24}$  if  $H_0: \sigma_1^2 < \sigma_2^2$
  - The rejection region is either  $f \geq F_{0.10, 19, 24}$  or  $f \leq F_{0.90, 19, 24}$
  - All of the above statements are true.
  - None of the above statements are true.
115. For testing the homogeneity of variability for more than two groups one can use
- F-Test
  - Bartlett Test
  - T-Test
  - None of these
116. The test which is used for testing the equality of variances among 3 or more groups is
- Wallis Test
  - Bartlett Test
  - Chi-square Test
  - Levene Test
  - Both b) & d)
117. The test which is used to test the variance of more than two groups and normality assumption is not necessary is
- Bartlett Test
  - F-Test
  - Levene's test
  - Chi-square test
118. In testing of hypothesis, the size of a test we mean that
- Significance level
  - Type I-error

- c) Type II-error  
 d) Both a) and b) but not c)  
 e) None of these
119. In testing of hypothesis, power of a test we mean that  
 a)  $P(\text{rejecting null hypothesis, when it is false})$   
 b)  $1 - P(\text{type II-error})$   
 c)  $\Pr(\text{type II-error})$   
 d) Both a) and b)  
 e) None of these
120. The power of a statistical test refers to its  
 a) Ability to eliminate statistical errors.  
 b) Ability to analyze data that violate the assumptions of the test.  
 c) Ability to detect differences between means.  
 d) All of the above
121. The power of a statistical test is affected by  
 a) Sample size.  
 b) The alpha level chosen.  
 c) Effect size.  
 d) All of the above  
 e) Both b) and c) only
122. Which of the following will increase the power of a test?  
 a) Increase  $n$ .  
 b) Increase  $\alpha$ .  
 c) Reduce the amount of variability in the sample.  
 d) Consider an alternative hypothesis further from the null.  
 e) All of these will increase the power of the test.
123. A hypothesis test is set up so that  $P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true}) = 0.05$  and  $P(\text{failing to reject } H_0 \text{ when } H_0 \text{ is false}) = 0.26$ . What is the power of the test?  
 a) 0.26  
 b) 0.05  
 c) 0.95  
 d) 0.21  
 e) 0.74
124. When is it to use a confidence interval instead of computing a  $P$ -value in a hypothesis test?  
 a) In any significance test  
 b) In any hypothesis test with a two-sided alternative hypothesis  
 c) Only when the hypothesized value of the parameter is *not* in the confidence interval  
 d) Only when you are conducting a hypothesis test with a one-sided alternative  
 e) Only when doing a test for a single population mean or a single population proportion
125. Which of the following is not a required step for a significance test?  
 a) State null and alternative hypotheses in the context of the problem.  
 b) Identify the test to be used and justify the conditions for using it.  
 c) State the significance level for which you will decide to reject the null hypothesis.  
 d) Compute the value of the test statistic and the  $P$ -value.  
 e) State a correct conclusion in the context of the problem.
126. Which of the following best describes what we mean when say that  $t$ -procedures are robust?  
 a) The  $t$ -procedures work well with almost any distribution.  
 b) The numerical value of  $t$  is not affected by outliers.  
 c) The  $t$ -procedures will still work reasonably well even if the assumption of normality is violated.  
 d)  $T$ -procedures can be used as long as the sample size is at least 40.  
 e)  $T$ -procedures are as accurate as  $z$ -procedures.

17 The quantity  $\frac{(n-1)S^2}{\sigma^2}$  is distributed

- as  
 a) Chi-Square with  $n$  degrees of freedom  
 b) T- with  $n-1$  degrees of freedom  
 c) Chi-Square with  $n-1$  degrees of freedom  
 d) None of these

18 If population proportion  $P = 0.00015$ ,  
 Sample proportion  
 $\hat{p} = 0.001$ ,  $Z_{\alpha} = 1.645$  &  $Z_{\beta} = 2.326$   
 then sample size will be

- a) 12132  
 b) 12142  
 c) 12412  
 d) 12214

19 For large  $n$  if sample Proportion  
 $\hat{p} = 0.00015$  &  $Z_{\alpha} = 1.645$  then  
 critical value is

- a) 0.001                  b) 0.05  
 c) 0.005                  d) 0.0005

20 If  
 $P = 0.00015$ ,  $\hat{p} = 0.00375$  &  $n = 3200$   
 then p-value is

- a)  $10^{-92}$   
 b)  $10^{-62}$   
 c)  $10^{-32}$   
 d)  $10^{-16}$

21 When a 95% confidence interval is constructed for a parameter  $\theta$

- a)  $\theta$  lies in the interval with probability 0.95  
 b) 95% of the intervals constructed contain the estimate of  $\theta$  found from the data  
 c) The interval covers 95% of the possible values for  $\theta$   
 d) The calculation uses the middle 95% of the data.

132 A 95% confidence interval for a proportion  $p$  based on 250 observations is (0.35, 0.54). This shows that

- a)  $p$  cannot be less than 0.35  
 b)  $p$  cannot be more than 0.54  
 c) with 95% probability, the true value of  $p$  lies between 0.35 and 0.54  
 d)  $p$  is significantly greater than 0.4  
 e)  $p$  is significantly different from 0.5

133 A random sample size of 45 is obtained for the purpose of testing the hypothesis  $H_0 : P = 0.80$ . The sample proportion is determined to be  $\hat{p} = 0.75$ . What is the value of the standard error of  $\hat{p}$  for this test?

- a) 0.0042  
 b) 0.0596  
 c) 0.0036  
 d) 0.0645  
 e) 0.0055

134. Ten parts are randomly selected from a normally distributed population. The lengths of the parts are measured and the mean of the 10 values is 10.622. The sample standard deviation of the 10 values is .005. We are 95% confident that the mean of the population from which the sample is drawn is between:

- a) 10.600 and 10.644  
 b) 10.615 and 10.629  
 c) 10.618 and 10.626  
 d) 10.621 and 10.623

135. A sample of size 35 is selected from a population of 10,000. The resulting analysis shows that "The 95% confidence interval for the mean is (34.5, 45.6)." This indicates that:

- a) 95% of the sample values are between 34.5 and 45.6.  
 b) 95% of the population is between 34.5 and 45.6.  
 c) With 95% confidence, it is believed the population mean is between 34.5 and 45.6.

- d) There is a 95% probability that the sample mean is between 34.5 and 45.6.  
 e) 95% of the sample means are between 34.5 and 45.6.

136. A lot of 10,000 pieces of plastic tubing has been received. The purchase order states that the average inside diameter is to be .188 inch. Inside diameter is normally distributed. Twelve pieces are randomly selected. Their inside diameters are:

.182 .188 .186 .186 .187 .183 .182 .183 .185 .184 .183 .186 A hypothesis test is performed with null hypothesis

$$H_0: \mu = 0.188 \text{ and } \text{alternative}$$

hypothesis  $H_a: \mu < 0.188$ . The critical (or table) value would be ( $\alpha = 0.10$ ):

- a) 2.681  
 b) 2.718  
 c) 3.055  
 d) 3.106  
 e) 2.33

137. After conducting a hypothesis test, it is concluded that the null hypothesis can not be rejected at the .05 significance level. This means that:

- a) We can be 95% certain that the null hypothesis is true.  
 b) There is a 95% probability that the alternative hypothesis is false.  
 c) There is a 95% probability that the alternative hypothesis is true.  
 d) There is a 5% probability that the null hypothesis is true.  
 e) None of the above.

138. "A hypothesis test yields a  $p$ -value of 0.20." Which of the following best describes what is meant by this statement?

- a) The probability of getting a finding at least as extreme as obtained by chance alone if the null hypothesis is true is 0.20.  
 b) The probability of getting a finding as extreme as obtained by chance

alone from repeated random sampling is 0.20.  
 c) The probability is 0.20 that our finding is significant.  
 d) The probability of getting this finding is 0.20.  
 e) The finding we got will occur less than 20% of the time in repeated trials of this hypothesis test.

139. A random sample of 25 men and a separate random sample of 25 women are selected to answer questions about attitudes toward abortion. The answers were categorized as "pro-life" or "pro-choice." Which of the following is the proper null hypothesis for this situation?

- a) The variables "gender" and "attitude toward abortion" are related.  
 b) The proportion of "pro-life" men is the same as the proportion of "pro-life" women.  
 c) The proportion of "pro-life" men is related to the proportion of "pro-life" women.  
 d) The proportion of "pro-choice" men is the same as the proportion of "pro-life" women.  
 e) The variables "gender" and "attitude toward abortion" are independent.

140. A major polling organization wants to predict the outcome of an upcoming national election (in terms of the proportion of voters who will vote for each candidate). They intend to use a 95% confidence interval with margin of error of no more than 2.5%. What is the minimum sample size needed to accomplish this goal?

- a) 1536  
 b) 39  
 c) 1537  
 d) 40  
 e) 2653

141 Association between two nominal level variables is tested through

- a) t-test
- b) Chi-square test
- c) F-Test
- d) All of the above
- e) None of these

142 Association between two interval or ratio level variables is tested through

- a) t-Test
- b) Chi-square test
- c) F-test
- d) All of the above
- e) None of these

143 If independent is Nominal/Numerical and dependent is numerical, then we use

- a) Cox regression
- b) ANOVA
- c) Logistic regression
- d) MANOVA
- e) Both b) and d)

144 If independent is Numerical and no dependent variable, then we use

- a) Cox regression
- b) Cluster Analysis
- c) Logistic regression
- d) Factor analysis
- e) Both b) and d)

145 In testing of hypothesis if p-value is less than type I error, then we

- a) Reject  $H_a$
- b) Reject  $H_0$
- c) Accept  $H_a$
- d) Both b) and c) but not a)

146 Given  $H_0: \mu \geq 18$  and  $H_a: \mu < 18$ , we would commit Type I error if we

- a) Conclude that  $\mu \geq 18$  when the truth is that  $\mu < 18$ .
- b) Conclude that  $\mu < 18$  when the truth is that  $\mu \geq 18$
- c) Fail to reject  $\mu \geq 18$  when the truth is that  $\mu < 18$ .
- d) None of these are true

147 Which is not true of p-values?

- a) When they are small, we want to reject  $H_0$ .
- b) They must be specified before the sample is taken.
- c) They show the chance of Type I error if we reject  $H_0$ .
- d) None of these

148 Dulco Manufacturing claims that its alkaline batteries last at least forty hours on average in a certain type of portable CD player. But tests on a random sample of 18 batteries from a day's large production run showed a mean battery life of only 37.8 hours with a standard deviation of 5.4 hours. To test DulCo's hypothesis, the test statistic is

- a) -1.980
- b) -1.728
- c) -2.101
- d) -1.960

149 Last year, 10 percent of all teenagers owned an iPhone. This year, a sample of 260 randomly chosen teenagers showed that 39 owned an iPhone. The test statistic to find out whether the percent has risen is

- a) 2.687
- b) 2.758
- c) 0.0256
- d) 2.258

150 Last year, 10 percent of all teenagers purchased a new iPhone. This year, a sample of 260 randomly chosen teenagers showed that 39 had purchased a new iPhone. To test whether the percent has risen, the critical value at  $\alpha = 0.05$  is

- a) 1.686
- b) 1.655
- c) 1.645
- d) 1.960

151 Last year, 10 percent of all teenagers purchased a new iPhone. This year, a sample of 260 randomly chosen teenagers showed that 39 had purchased a new iPhone. To test

- What is the current has from the p-value is
- 0.0001
  - 0.0014
  - 0.0492
  - 0.1135
- 152 Assuming that other factors remain the same, which of the following statements is most nearly correct for a t-test of a mean?
- For a given  $\alpha$ , the critical value of Student's  $t$  is smaller if  $n$  is smaller.
  - If  $t_{\text{calc}} = 1.432$  with  $n = 22$ , we get a clear-cut rejection in a right-tailed test at  $\alpha = 0.05$ .
  - Rejecting  $H_0: \mu = 75$  in a two-tailed test implies rejection in a one-tailed test at the same  $\alpha$ .
  - A calculated p-value of 0.13 would lead us to reject the null hypothesis at  $\alpha = 0.10$ .
- 153 John rejected a null hypothesis in a right-tailed test for a mean at  $\alpha = .025$  because his critical  $t$  value was 2.000 and his calculated  $t$  value was 2.345. We can be sure that
- John did not commit Type I error.
  - John did not commit type II error.
  - John committed neither Type I nor Type II error.
  - None of the above can definitely be concluded.
- 154 In a right-tail test, a statistician came up with a z test statistic of 1.470. What is the p-value?
- 0.4292
  - 0.0708
  - 0.0874
  - 0.0301
- 155 In a right-tailed test of hypothesis for a population mean with 13 degrees of freedom, the value of the test statistic was 1.863. The p-value is
- Less than 0.025
  - Between 0.025 and 0.05
  - Between 0.05 and 0.10
  - Greater than 0.10
- 156 What minimum sample size is required to construct a confidence interval for the mean of a population with standard deviation 0.35? Assume a confidence level of 98% and a confidence interval width of .02
- 8
  - 33
  - 21
  - 67
- 157 A purchaser wants to determine whether or not there is any difference between the means of the corrugate paperboard cans supplied by two different vendors, A and B. A random sample of 100 cans is selected from the output of each vendor. The sample from A yielded a mean of 13.59 with a standard deviation of 5.94. The sample from B yielded a mean of 14.43 with a standard deviation of 5.61. Which of the following would be a suitable null hypothesis to test?
- $\mu_A = \mu_B$
  - $\mu_A > \mu_B$
  - $\mu_A < \mu_B$
  - $\mu_A \neq \mu_B$
- 158 A psychologist believes that positive rewards for proper behavior is more effective than punishment for bad behavior in promoting good behavior in children. A scale of "proper behavior" is developed.  $\mu_1$  = the "proper behavior" rating for children receiving positive rewards, and  $\mu_2$  = the "proper behavior" rating for children receiving punishment. If  $H_0: \mu_1 - \mu_2 = 0$ , which of the following is the proper statement of  $H_a$ ?
- $H_a: \mu_1 - \mu_2 > 0$
  - $H_a: \mu_1 - \mu_2 < 0$
  - $H_a: \mu_1 - \mu_2 \neq 0$

- d) Any of the above is an acceptable alternative to the given null.
- e) There isn't enough information given in the problem for us to make a decision.
159. If a sample size of 16 has an average of 2.53 and a standard deviation of 0.04, estimate the 95% confidence interval for the population mean,  $m$  (assume a normal distribution).
- a) (2.525, 2.535)  
 b) (2.52, 2.54)  
 c) (2.44, 2.62)  
 d) (2.51, 2.55)
160. The test used for testing significance in an analysis of variance table is:
- a) The z-test.  
 b) The t-test.  
 c) The F-test.  
 d) The chi-square test.
161. Let  $X_1, X_2, \dots, X_n$  be a random sample from the normal distribution with mean  $\theta$  and variance is 25. Then which of the following hypothesis is simple
- a)  $H_0: \theta \leq 17$   
 b)  $H_0: \theta \geq 17$   
 c)  $H_0: \theta = 17$   
 d) None of the above
162. When we want to compare two ordinal treatment groups then, which of the following test is used
- a) Unpaired T-test  
 b) Paired T-test  
 c) Mann-Whitney test  
 d) Chi-Square test  
 e) None of the above
163. When we want to compare three nominal treatment groups, then which of the following test is used
- a) F-test  
 b) Bartlett test  
 c) Kruskal-Wallis test  
 d) Chi-Square test  
 e) None of the above
164. A  $\chi^2$  goodness-of-fit test is performed on a random sample of 360 individuals to see if the number of birthdays each month is proportional to the number of days in the month.  $\chi^2$  is determined to be 23.5. The P-value for this test is
- a)  $0.001 < P < 0.005$   
 b)  $0.02 < P < 0.02$   
 c)  $0.025 < P < 0.05$   
 d)  $0.01 < P < 0.02$   
 e)  $0.05 < P < 0.10$
165. When we want to compare two survival time groups, then which of the following test is used
- a) F-test  
 b) Bartlett test  
 c) Kruskal-Wallis test  
 d) Gehan's test  
 e) None of the above
166. When we want to compare nominal multiple treatments of same individuals, then which of the following test is used
- a) F-test  
 b) Cochran Q-test  
 c) Friedman statistic  
 d) McNemar's test  
 e) None of the above
167. Three groups of subjects were followed over the course of five years to compare treatments for sideroblastic anemia. The most appropriate statistical analysis to determine the quantitative serologic differences resulting from these treatments would be a
- a) Regression analysis  
 b) F test (ANOVA)  
 c) Correlation analysis  
 d) Chi-square test  
 e) T test

168. When the standard for accepting the difference was at p-value of 0.05 and the calculated value was 0.01, the null hypothesis was rejected by the researcher. What do you think of results?

- a) Wrongly rejected
- b) Significant difference
- c) No difference
- d) Alternate hypothesis is wrong
- e) Sample size was small

169. Which of the following statements are false about the chi-squared distribution with  $v$  degrees of freedom?

- a) It is a discrete probability distribution with a single parameter  $v$ .
- b) It is positively skewed (long upper tail)
- c) It becomes more symmetric as  $v$  increases.
- d) All of the above statements are true.
- e) All of the above statements are false.

170. Which of the following statements are true about the percentiles of a chi-squared distribution with 20 degrees of freedom?

- a) The 5th percentile is 31.410
- b) The 95th percentile is 10.851
- c) The 10th percentile is 12.443
- d) The 90th percentile is 37.566
- e) All of the above statements are true.

171. The lower limit of a 95% confidence interval for the variance  $\sigma^2$  of a normal population using a sample of size  $n$  and variance value  $s^2$  is given by:

- a)  $(n-1)s^2 / \chi^2_{0.05, n-1}$
- b)  $(n-1)s^2 / \chi^2_{0.025, n-1}$

c)  $(n-1)s^2 / \chi^2_{0.95, n-1}$

d)  $(n-1)s^2 / \chi^2_{0.975, n-1}$

- e) None of the above answers are correct.

172. As the degrees of freedom increase (and especially when the degrees of freedom are more than 90), the graph of the chi-square distribution los more and more.

- a) Symmetrical
- b) Skewed to right
- c) Skewed to left
- d) Asymmetrical

173. The Goodness-of-Fit hypothesis test is typically a.

- a) Two-tailed test
- b) Wagging-tailed test
- c) Left-tailed test
- d) Right-tailed test

174. The standard statistical test for testing normality is

- a) F test
- b) The Kolmogorov-Smirnov test
- c) T-test
- d) Non of these

175. A higher value of Anderson-Darling test of normality test statistic indicate that

- a) Good fit
- b) Lack of fit
- c) Both a) and b)
- d) None of these

176. Critical value for Anderson-Darling test statistic is calculated by

- a) Chi-Square distribution
- b) Normal distribution
- c) Exponential distribution
- d) Binomial distribution

177. Which of the following statements are true?

- i. In order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table must be at

II. In order to use  $\chi^2$  procedures, you must have at least 2 degrees of freedom.

III. In a  $4 \times 2$  two-way table, the number of degrees of freedom is 3.

- a) I only
- b) I and III only
- c) I and II only
- d) III only
- e) I, II, and III

178. The critical value in a chi-square goodness-of-fit test depends on

- a) The number of categories.
- b) The normality of the population.
- c) The value of the test statistic.
- d) All of the above.

179. For a chi-square test, a  $4 \times 5$  contingency table will have how many degrees of freedom?

- a) 12
- b) 8
- c) 9
- d) 6

180. We sometimes combine two categories in a chi-square test if

- a) The sample size is less than 30.
- b) Their observed frequencies are below 5.
- c) Their expected frequencies are below 5.
- d) The p-value is less than  $\alpha$ .

181. Association between two nominal variables is measured by

- a) Linear regression
- b) Odds ratio
- c) Spearman rank correlation
- d) All of the above
- e) None of the above

182. If one finding is statistically significant at  $p < 0.01$  and a second at  $p < 0.05$ , it would be logical to say that:

- a) Finding 1 is more significant than finding 2.
- b) Finding 2 is more significant than finding 1.
- c) Finding 1 and finding 2 are equally significant

d) You can have greater confidence in rejecting the null hypothesis for finding 1 than finding 2.

183. A data transformation that changes the value of numbers, but not the scale of measurement are called:

- a) Nonlinear transformations.
- b) Geometric transformations.
- c) Linear transformations.
- d) Simple transformations.

184. A legitimate reason for transforming your data is:

- a) To help a non significant finding become significant.
- b) When your data do not meet assumptions of a parametric statistic and no nonparametric alternative is available.
- c) To reduce the effects of extraneous variables.
- d) All of the above

185. If for some reason you cannot use inferential statistics, you may have to:

- a) Establish reliability through replication.
- b) Redo your experiment so that you can use inferential statistics.
- c) Simply "eyeball" your results to determine reliability.
- d) Ignore reliability issues and interpret your data anyway.

186. Which of the following statements are not true?

- a) The chi-squared distribution is used to obtain a confidence interval for the variance  $\sigma^2$  of a normal population.
- b) Provided that  $np_i \geq 5$  for every  $i$  ( $i = 1, 2, \dots, k$ ), the  $\chi^2$  goodness-of-fit test statistic when all  $k$  category probabilities are completely specified has approximately a t distribution with  $k-1$  degrees of freedom.
- c) A multinomial experiment generalizes a binomial experiment by allowing each trial to result in one of  $k$  possible outcomes, where

k>2. In general, we refer to these outcomes as categories.

- d) All of the above statements are correct.
- e) None of the above statements are correct.

187. Which of the following statements are true regarding the critical value  $\chi_{v,\alpha}$ , for the chi-squared distribution, when  $\alpha = 0.05$  and  $v = 4$ ?

- a) The area to the right of 9.488 is 0.05.
- b) The area to the left of 9.488 is 0.95.
- c) The total area under the chi-squared curve is 9.488.
- d) All of the above statements are true.
- e) None of the above statements are true.

188. In testing

$$H_0 : p_1 = p_{10}, p_2 = p_{20}, \dots, p_5 = p_{50}$$

verses alternative  $H_a$  that states that at least one  $p_i$  does not equals  $p_{i0}$

rejection of  $H_0$  is appropriate is at 0.10 significance level when the test statistic value  $\chi^2$  is

- a) Greater than or equal to 9.236.
- b) smaller than or equal to 11.070
- c) between 9.236 and 11.070
- d) smaller than or equal to 7.779
- e) greater than or equal to 7.779

189. The main difference between a  $\chi^2$  test for independence and a  $\chi^2$  test for homogeneity of proportions is which of the following?

- a) They are based on a different number of degrees of freedom.
- b) One of the tests is for a two-sided alternative and the other is for a one-sided alternative.
- c) In one case, two variables are compared within a single population. In the other case, two

populations are compared in terms of a single variable.

- d) For a given value of  $\chi^2$ , they have different P-values.
- e) There are no differences between the tests. They measure exactly the same thing.

190. Which of the following statements are true?

- a) The  $\chi^2$  goodness-of-fit test can be used when the number of categories k is two or more.
- b) If  $Z \sim N(0,1)$ , then  $Z^2$  has a distribution with one degree of freedom.
- c) The chi-squared tests in this chapter are not all upper-tailed.
- d) The P-value for an upper-tailed chi-squared test is the area under the chi-squared curve with v degrees of freedom to the left of the calculated  $\chi^2$  test statistic value.
- e) All of the above statements are true.

191. Which of the following statements are not true?

- a) In testing

$H_0 : p_1 = p_{10}, \dots, p_k = p_{k0}$  versus  $H_a : H_0$  is not true, the null hypothesis is simple hypothesis in the sense that each  $p_{i0}$  is a specified number.

- b) In testing

$H_0 : p_1 = \theta^2, p_2 = 2\theta(1-\theta), p_3 = (1-\theta)^2$  versus  $H_a : H_0$  is not true, the null hypothesis is composite in the sense that knowing  $H_0$  is true does not uniquely determine the cell probabilities and expected cell counts but only their general form.

- c) A general rule of thumb for degrees of freedom in a chi-

- squared test is that  $\chi^2$  d.f. is the difference between the number of freely determined cell counts and the number of independent parameters estimated.
- d) All of the above statements are true  
e) None of the above statements are true
192. The  $\chi^2$  goodness-of-fit test statistic, when there are  $k$  categories and  $m$  parameters to be estimated from the sample data, has approximately a chi-squared distribution with  $v$  degrees of freedom, where  $v$  equals
- a)  $m-k-1$   
b)  $k-m$   
c)  $k-m-1$   
d)  $m+k-1$   
e)  $k-m+1$
193. Which of the following statements are not true?
- a) The chi-squared goodness-of-fit test can be used to test whether the sample comes from a specified family of continuous distributions, such as the normal family, but it cannot be used to test whether the sample comes from a specified discrete distribution, such as Poisson.  
b) A normal probability plot is used for checking whether any member of the normal distribution family is plausible.  
c) The sample correlation coefficient  $r$  is a quantitative measure of the extent to which points cluster about a straight line.  
d) The null hypothesis of population normality is rejected if the sample correlation coefficient  $r$  is less than or equal to,  $c_\alpha$  where  $c_\alpha$  is a critical value chosen to yield the desired significance level  $\alpha$ .  
e) All of the above statements are true.
194. The number of degrees of freedom for a two-way contingency table with  $I$  rows and  $J$  columns is
- a)  $IJ$   
b)  $(I-1)J$   
c)  $I(J-1)$   
d)  $(I-1)(J-1)$   
e)  $I+J-1$
195. In a two-way contingency table with 3 rows and 5 columns, assume that the second row total is 120 and the fourth column total is 50, and the total number of observations is 600. Then, the estimated expected count in cell (2, 4) is
- a) 50  
b) 40  
c) 30  
d) 20  
e) 10
196. The chi-squared test for homogeneity can safely be applied as long as each estimated expected count  $\hat{e}_{ij}$  for all cells in the contingency table must be
- a) At least 5  
b) At most 10  
c) At least 10  
d) At most 15  
e) Any number between 10 and 15
197. Which of the following statements are not true?
- a) The chi-squared test statistic used in testing for independence is identical to that used in testing for homogeneity.  
b) In general, the number of degrees of freedom when testing for independence is larger than those used in testing for homogeneity.  
c) The chi-squared test for independence can safely be applied as long as the estimated expected count  $\hat{e}_{ij}$  for all cells in the contingency table is larger than or equal to 5.  
d) The rejection region in testing for homogeneity at significance level  $\alpha$  is that the test statistic value

168 When the standard for accepting the difference was at p-value of 0.05 and the calculated value was 0.01, the null hypothesis was rejected by the researcher. What do you think of results?

- a) Wrongly rejected
- b) Significant difference
- c) No difference
- d) Alternate hypothesis is wrong
- e) Sample size was small

169 Which of the following statements are false about the chi-squared distribution with  $v$  degrees of freedom?

- a) It is a discrete probability distribution with a single parameter  $v$ .
- b) It is positively skewed (long upper tail)
- c) It becomes more symmetric as  $v$  increases.
- d) All of the above statements are true.
- e) All of the above statements are false.

170 Which of the following statements are true about the percentiles of a chi-squared distribution with 20 degrees of freedom?

- a) The 5th percentile is 31.410
- b) The 95th percentile is 10.851
- c) The 10th percentile is 12.443
- d) The 90th percentile is 37.566
- e) All of the above statements are true.

171 The lower limit of a 95% confidence interval for the variance  $\sigma^2$  of a normal population using a sample of size  $n$  and variance value  $s^2$  is given by:

- a)  $(n-1)s^2 / \chi^2_{0.05, n-1}$
- b)  $(n-1)s^2 / \chi^2_{0.025, n-1}$

c)  $(n-1)s^2 / \chi^2_{0.95, n-1}$

d)  $(n-1)s^2 / \chi^2_{0.975, n-1}$

e) None of the above answers are correct.

172 As the degrees of freedom increase (and especially when the degrees of freedom are more than 90), the graph of the chi-square distribution becomes more:

- a) Symmetrical
- b) Skewed to right
- c) Skewed to left
- d) Asymmetrical

173 The Goodness-of-Fit hypothesis test is typically a:

- a) Two-tailed test
- b) Wagging-tailed test
- c) Left-tailed test
- d) Right-tailed test

174 The standard statistical test for testing normality is

- a) F test
- b) The Kolmogorov-Smirnov test
- c) T-test
- d) Non of these

175 A higher value of Anderson-Darling test of normality test statistic indicates that

- a) Good fit
- b) Lack of fit
- c) Both a) and b)
- d) None of these

176 Critical value for Anderson-Darling test statistic is calculated by

- a) Chi-Square distribution
- b) Normal distribution
- c) Exponential distribution
- d) Binomial distribution

177 Which of the following statements are true?

- i. In order to use a  $\chi^2$  procedure, the expected value for each cell of a one- or two-way table must be at least 5.

- ii) In order to use  $\chi^2$  procedures, you must have at least 2 degrees of freedom.
- iii) In a  $4 \times 2$  two-way table, the number of degrees of freedom is 3.
- I only
  - I and III only
  - I and II only
  - III only
  - I, II, and III
175. The critical value in a chi-square goodness-of-fit test depends on
- The number of categories.
  - The normality of the population.
  - The value of the test statistic.
  - All of the above.
179. For a chi-square test, a  $4 \times 5$  contingency table will have how many degrees of freedom?
- 12
  - 8
  - 9
  - 6
180. We sometimes combine two categories in a chi-square test if
- The sample size is less than 30.
  - Their observed frequencies are below 5.
  - Their expected frequencies are below 5.
  - The p-value is less than  $\alpha$ .
181. Association between two nominal variables is measured by
- Linear regression
  - Odds ratio
  - Spearman rank correlation
  - All of the above
  - None of the above
182. If one finding is statistically significant at  $p < 0.01$  and a second at  $p < 0.05$ , it would be logical to say that:
- Finding 1 is more significant than finding 2.
  - Finding 2 is more significant than finding 1.
  - Finding 1 and finding 2 are equally significant.
- d) You can have greater confidence in rejecting the null hypothesis for finding 1 than finding 2.
183. A data transformation that changes the value of numbers, but not the scale of measurement are called
- Nonlinear transformations
  - Geometric transformations
  - Linear transformations
  - Simple transformations
184. A legitimate reason for transforming your data is:
- To help a non significant finding become significant.
  - When your data do not meet assumptions of a parametric statistic and no nonparametric alternative is available.
  - To reduce the effects of extraneous variables.
  - All of the above
185. If for some reason you cannot use inferential statistics, you may have to:
- Establish reliability through replication.
  - Redo your experiment so that you can use inferential statistics.
  - Simply "eyeball" your results to determine reliability.
  - Ignore reliability issues and interpret your data anyway.
186. Which of the following statements are not true?
- The chi-squared distribution is used to obtain a confidence interval for the variance  $\sigma^2$  of a normal population.
  - Provided that  $np_i \geq 5$  for every  $i$  ( $i = 1, 2, \dots, k$ ), the  $\chi^2$  goodness-of-fit test statistic when all  $k$  category probabilities are completely specified has approximately a t distribution with  $k-1$  degrees of freedom.
  - A multinomial experiment generalizes a binomial experiment by allowing each trial to result in one of  $k$  possible outcomes, where

- k>2. In general, we refer to these outcomes as categories.
- d) All of the above statements are correct.  
e) None of the above statements are correct.
187. Which of the following statements are true regarding the critical value  $\chi_{v,\alpha}$ , for the chi-squared distribution, when  $\alpha = 0.05$  and  $v = 4$ ?
- a) The area to the right of 9.488 is 0.05.  
b) The area to the left of 9.488 is 0.95.  
c) The total area under the chi-squared curve is 9.488.  
d) All of the above statements are true.  
e) None of the above statements are true.
188. In testing  $H_0 : p_1 = p_{10}, p_2 = p_{20}, \dots, p_5 = p_{50}$  versus alternative  $H_a$  that states that at least one  $p_i$  does not equals  $p_{i0}$  rejection of  $H_0$  is appropriate at 0.10 significance level when the test statistic value  $\chi^2$  is
- a) Greater than or equal to 9.236.  
b) smaller than or equal to 11.070  
c) between 9.236 and 11.070  
d) smaller than or equal to 7.779  
e) greater than or equal to 7.779
189. The main difference between a  $\chi^2$  test for independence and a  $\chi^2$  test for homogeneity of proportions is which of the following?
- a) They are based on a different number of degrees of freedom.  
b) One of the tests is for a two-sided alternative and the other is for a one-sided alternative.  
c) In one case, two variables are compared within a single population. In the other case, two populations are compared in terms of a single variable.
- d) For a given value of  $\chi^2$ , they have different P-values.  
e) There are no differences between the tests. They measure exactly the same thing.
190. Which of the following statements are true?
- a) The  $\chi^2$  goodness-of-fit test can be used when the number of categories k is two or more.  
b) If  $Z \sim N(0,1)$ , then  $Z^2$  has a distribution with one degree of freedom.  
c) The chi-squared tests in this chapter are not all upper-tailed.  
d) The P-value for an upper-tailed chi-squared test is the area under the chi-squared curve with v degrees of freedom to the left of the calculated  $\chi^2$  test statistic value.  
e) All of the above statements are true.
191. Which of the following statements are not true?
- a) In testing  $H_0 : p_1 = p_{10}, \dots, p_k = p_{k0}$  versus  $H_a : H_0$  is not true, the null hypothesis is simple hypothesis in the sense that each  $p_{i0}$  is a specified number.
- b) In testing  $H_0 : p_1 = \theta^2, p_2 = 2\theta(1-\theta), p_3 = (1-\theta)^2$  versus  $H_a : H_0$  is not true, the null hypothesis is composite in the sense that knowing  $H_0$  is true does not uniquely determine the cell probabilities and expected cell counts but only their general form.
- c) A general rule of thumb for degrees of freedom in a chi-

- squared test is that  $\chi^2$  d.f. is the difference between the number of freely determined cell counts and the number of independent parameters estimated.
- d) All of the above statements are true  
e) None of the above statements are true
192. The  $\chi^2$  goodness-of-fit test statistic, when there are  $k$  categories and  $m$  parameters to be estimated from the sample data, has approximately a chi-squared distribution with  $v$  degrees of freedom, where  $v$  equals
- a)  $m-k-1$   
b)  $k-m$   
c)  $k-m-1$   
d)  $m+k-1$   
e)  $k-m+1$
193. Which of the following statements are not true?
- a) The chi-squared goodness-of-fit test can be used to test whether the sample comes from a specified family of continuous distributions, such as the normal family, but it cannot be used to test whether the sample comes from a specified discrete distribution, such as Poisson.  
b) A normal probability plot is used for checking whether any member of the normal distribution family is plausible.  
c) The sample correlation coefficient  $r$  is a quantitative measure of the extent to which points cluster about a straight line.  
d) The null hypothesis of population normality is rejected if the sample correlation coefficient  $r$  is less than or equal to,  $c_\alpha$  where  $c_\alpha$  is a critical value chosen to yield the desired significance level  $\alpha$   
e) All of the above statements are true.
194. The number of degrees of freedom for a two-way contingency table with  $I$  rows and  $J$  columns is
- a)  $IJ$   
b)  $(I-1)J$   
c)  $I(J-1)$   
d)  $(I-1)(J-1)$   
e)  $I+J-1$
195. In a two-way contingency table with 3 rows and 5 columns, assume that the second row total is 120 and the fourth column total is 50, and the total number of observations is 600. Then, the estimated expected count in cell (2, 4) is
- a) 50  
b) 40  
c) 30  
d) 20  
e) 10
196. The chi-squared test for homogeneity can safely be applied as long as each estimated expected count  $\hat{e}_{ij}$  for all cells in the contingency table must be
- a) At least 5  
b) At most 10  
c) At least 10  
d) At most 15  
e) Any number between 10 and 15
197. Which of the following statements are not true?
- a) The chi-squared test statistic used in testing for independence is identical to that used in testing for homogeneity.  
b) In general, the number of degrees of freedom when testing for independence is larger than those used in testing for homogeneity.  
c) The chi-squared test for independence can safely be applied as long as the estimated expected count  $\hat{e}_{ij}$  for all cells in the contingency table is larger than or equal to 5.  
d) The rejection region in testing for homogeneity at significance level  $\alpha$  is that the test statistic value

- $\chi^2 \geq \chi^2_{\alpha(I-1)(J-1)}$  where I and J are the number of rows and columns, respectively, in the two-way contingency table.
- e) All of the above statements are true.
198. The number of degrees of freedom in testing for independence when using a contingency table with 6 rows and 4 columns is:
- a) 24      b) 10  
 c) 15      d) 20  
 e) 12
199. Association of two attributes is measured by which of the following
- a) Coefficient of correlation  
 b) Coefficient of rank correlation  
 c) Coefficient of association  
 d) None of these
200. Two attributes A and B are independent if
- a)  $(AB) = \frac{(A)(B)}{N}$   
 b)  $(AB) < \frac{(A)(B)}{N}$   
 c)  $(AB) = \frac{(A)(B)}{N}$   
 d) None of these
201. For two attributes A and B if  $(AB) > \frac{(A)(B)}{N}$ , then
- a) A and B are independent  
 b) A and B are Positively associated  
 c) A and B are negatively associated  
 d) None of these
202. For two attributes A and B if  $(AB) < \frac{(A)(B)}{N}$ , then
- a) A and B are independent  
 b) A and B are Positively associated  
 c) A and B are negatively associated  
 d) None of these
203. Yule has proposed an other measure of attributes association that is
- a) Coefficient of colligation  
 b) Coefficient of contingency  
 c) Cramer's coefficient of contingency  
 d) None of these
204. The range of coefficient of measure of association is
- a)  $0 \leq Q \leq 1$   
 b)  $-1 \leq Q \leq 1$   
 c)  $-1 \leq Q \leq 1$   
 d) None of these
205. The formula of coefficient of association is
- a)  $Q = \frac{(AB)(\alpha\beta) + (A\beta)(\alpha B)}{(AB)(\alpha\beta) - (A\beta)(\alpha B)}$   
 b)  $Q = \frac{(AB)(\alpha\beta) + (A\beta)(\alpha B)}{(AB)(\alpha\beta) - (A\beta)(\alpha B)}$   
 c) Both a) and b) are true  
 d) None of these
206. For two attributes A and B if  $Q=0$ . Then A and B are
- a) Independent  
 b) Positively associated  
 c) Negatively associated  
 d) None of these
207. If two attributes A and B are positively associated, then coefficient of association is
- a) 0  
 b) Between 0 and -1  
 c) Between 0 and 1  
 d) -1  
 e) None of these
208. Two attributes A and B are negatively associated if
- a)  $(AB)(\alpha\beta) = (A\beta)(\alpha B)$   
 b)  $(AB)(\alpha\beta) > (A\beta)(\alpha B)$   
 c)  $(AB)(\alpha\beta) < (A\beta)(\alpha B)$   
 d) None of these

- 109 If cell frequency ( $\alpha/\beta$ ) = 0, then coefficient of association
- 0
  - 1
  - 1
  - Between 0 and -1
  - None of these

- 210 If cell frequency ( $\alpha\beta$ ) = 0, then coefficient of association
- 0
  - 1
  - 1
  - Between 0 and -1
  - None of these

- 211 If cell frequency ( $\alpha\beta$ ) = 0, then the value of coefficient of colligation
- 0
  - 1
  - 1
  - Between 0 and -1
  - Undefined

- 212 If cell frequency ( $\alpha B$ ) = 0, then the value of coefficient of colligation
- 0
  - 1
  - 1
  - Between 0 and -1
  - Undefined

- 213 An other measure of coefficient of contingency is
- Coefficient of colligation
  - Coefficient of contingency
  - Cramer's coefficient of contingency
  - Rank correlation
  - None of these

- 214 If  $N = 150, (A) = 30, (B) = 60, (AB) = 12$ , then A and B are
- Positively associated
  - Negatively associated

- 215 If  $N = 1000, (AB) = 110, (B) = 200, (A) = 400$ , then A and B are
- Positively associated
  - Negatively associated
  - Independent
  - None of these

## ANSWERS

1	b	2	e	3	e	4	a	5	b	6	d	7	b	8	a	9	c	10	c	11	c	12	a	13	14	15	c	16	a	17	a	18	d	19	b	20	c	21	a	22	b	23	c	24	c	25	c	26	c	27	c	28	c	29	d	30	c	31	c	32	c	33	b	34	a	35	c	36	b	37	c	38	b	39	c	40	b	41	d	42	e	43	b	44	c	45	b	46	e	47	a	48	b	49	b	50	d	51	a	52	a	53	d	54	b	55	e	56	b	57	a	58	e	59	a	60	d	61	d	62	d	63	a	64	c	65	d	66	e	67	a	68	b	69	d	70	c	71	b	72	e	73	b	74	b	75	a	76	d	77	c	78	b	79	e	80	d	81	c	82	b	83	a	84	c	85	d	86	b	87	e	88	a	89	c	90	d	91	c	92	d	93	b	94	a	95	e	96	d	97	b	98	c	99	a	100	d	101	b	102	c	103	e	104	c	105	d	106	d	107	a	108	e	109	d	110	c	111	d	112	c	113	d	114	c	115	b	116	e	117	c	118	d	119	d	120	c	121	d	122	e	123	e	124	c	125	c	126	c	127	c	128	b	129	d	130	b	131	c	132	c	133	d	134	c	135	c	136	a	137	e	138	a	139	b	140	c	141	d	142	a	143	e	144	e	145	d	146	b	147	b	148	b	149	a	150	c	151	d	152	c	153	b	154	b	155	b	156	d	157	a	158	a	159	d	160	c
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	---	----	---	----	---	----	----	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---	-----	---

161	c	162	c	163	d	164	d	189	c	190	a	191	d	192	c
165	d	166	b	167	b	168	b	193	a	194	d	195	e	196	a
169	a	170		171	d	172	a	197	b	198	c	199	c	200	c
173	a	174	b	175	b	176	b	201	b	202	c	203	a	204	c
177	b	178	a	179	a	180	c	205	b	206	a	207	c	208	c
181	b	182	d	183	c	184	b	209	b	210	d	211	e	212	b
185	a	186	b	187	c	188	e	213	c	214	c	215	a		

---

## **DESIGN OF EXPERIMENT**

1. Which of the following are important in designing an experiment?
  - I. Control of all variables that might have an influence on the response variable.
  - II. Randomization of subjects to treatment groups.
  - III. Use of a large number of subjects to control for small-sample variability.
  - a) I only
  - b) I and II only
  - c) II and III only
  - d) I, II, and III
  - e) II only
2. Which of the following statements are not true?
  - a) The analysis of variance, or more briefly ANOVA, refers broadly to a collection of experimental situations and statistical procedures for the analysis of quantitative responses from experimental units.
  - b) The simplest ANOVA problem is referred to as two-way ANOVA.
  - c) Single-factor ANOVA focuses on a comparison of more than two population or treatment means.
  - d) All of the above statements are true.
  - e) None of the above statements are true.
3. Analysis of variance
  - a) Has the prime aim of comparing variances
  - b) Has the prime aim of comparing means
  - c) Is based on a linear model
  - d) Removes systematic variation in a set of data
  - e) Compares a set of experimental treatments
  - f) Analysis of Variance Splits up a total sum of square
4. Which of the following statements about regression analysis is/are true?
  - b) Is carried out on frequency data
  - c) Assumes a normal distribution of residuals
  - d) Helps to fit a line to a set of points on a graph
  - e) Estimates the variance of residuals.
5. A mean square is
  - a) The square of a treatment mean
  - b) A sum of squares divided by its degrees of freedom
  - c) The smallest sum of squares in an ANOVA table
  - d) The average of the squares of all the observations
  - e) A corrected sum of squares.
6. One-way Analysis of Variance (ANOVA) is used to fit a straight line to data
  - b) Removes one source of systematic variation from the data
  - c) Requires equal replication of all experimental treatments
  - d) Tests the Null Hypothesis that treatment means are different
  - e) Can be used when treatments have different variances
7. In a single-factor ANOVA problem involving five populations or treatments, which of the following statements are true about the alternative hypothesis?
  - a) All five population means are equal.
  - b) All five population means are different.
  - c) At least two of the population mean are different.
  - d) At least three of the population mean are different.
  - e) At most, two of the population means are equal.

8. Which of the following statements are true?
- In some experiments, different samples contain different numbers of observations. However, the concepts and methods of single-factor ANOVA are most easily developed for the case of equal sample sizes.
  - The population or treatment distributions in single-factor ANOVA are all assumed to be normally distributed with the same variance  $\sigma^2$ .
  - In one-way ANOVA, if either the normality assumption or the assumption of equal variances is judged implausible, a method of analysis other than the usual F test must be employed.
  - The test statistic for single-factor ANOVA is  $F = \text{MSTr}/\text{MSE}$ , where MSTr is the mean square for treatments, and MSE is the mean square for error.
  - All of the above statements are true.
9. In single-factor ANOVA, MSTr is the mean square for treatments, and MSE is the mean square for error. Which of the following statements are not true?
- MSE is a measure of between-samples variation.
  - MSE is a measure of within-samples variation.
  - MSTr is a measure of between-samples variation.
  - The value of MSTr is affected by the status of  $H_0$  (true or false).
  - All of the above statements are true.
10. In single-factor ANOVA, MSE is the mean square for error, and MSTr is the mean square for treatments. Which of the following statements are not true?
- The value of MSTr is affected by the status of  $H_0$  (true or false).
- b) When  $H_0$  is true,  $E(\text{MSTr}) = E(\text{MSE}) = \sigma^2$  where  $\sigma^2$  is the common population variance
- c) When  $H_0$  is false,  $E(\text{MSTr}) > E(\text{MSE}) = \sigma^2$  where  $\sigma^2$  is the common population variance
- d) The value of MSE is affected by the status of  $H_0$  (true or false).
- e) All of the above statements are true.
11. In a single-factor ANOVA problem involving four populations/treatments, the four sample standard deviations are 25.6, 30.4, 28.7, and 32.50. Then, the mean square error is
- 29.3
  - 117.2
  - 864.865
  - 29.409
  - None of the above answers are correct.
12. In a single-factor ANOVA problem involving five populations/treatments with a random sample of four observations from each one, it is found that  $SSTr = 16.1408$  and  $SSE = 37.3801$ . Then the value of the F statistic is
- 1.619
  - 2.316
  - 0.432
  - 1.522
  - 4.248
13. In one-way ANOVA, which of the following statements are true?
- $SST$  is a measure of the total variation in the data.
  - $SSE$  measures variation that would be present within treatments even if  $H_0$  were true, and is thus the part of total variation that is unexplained by the truth or falsity of  $H_0$ .
  - $SSTr$  is the amount of variation between treatments that can't be explained by the model.

- explained by possible differences in the population or treatments' means.
- d) If explained variation is large relative to unexplained variation, then  $H_0$  is rejected in favor of  $H_a$ .
- e) All of the above statements are true.
14. Which of the following statements are not true?
- An F distribution arises in connection with a ratio in which there is one number of degrees of freedom associated with the numerator  $v_1$  and another number of degrees of freedom associated with the denominator  $v_2$ .
  - For single-factor ANOVA, a value of the test statistic  $F = MSt/MSE$  that is greatly smaller than 1, casts considerable doubt on  $H_0$ .
  - The unbiasedness of MSE in single-factor ANOVA is a consequence of  $E(S^2_i) = \sigma^2$  (the common population variance) whether  $H_0$  is true or false.
  - In a single-factor ANOVA problem, the F critical value that captures upper-tail area .05 under the F curve with  $v_1 = 6$  and  $v_2 = 4$  is 6.16.
  - All of the above statements are true.
15. The distribution of the test statistic in single-factor ANOVA is the
- Binomial distribution
  - Normal distribution
  - T distribution
  - F distribution
  - None of the above answers are correct.
16. In a single-factor ANOVA problem involving 3 treatments, the sample means were 5, 6, and 9. If each observation in the third sample was increased by 20, the test statistic value F would
- Increase
  - Decrease
  - Remain the same
  - Increase by 20
  - Decrease by 10
17. Which of the following statements are not true?
- When the computed value of the F statistic in single-factor ANOVA is significant, the analysis is terminated because no differences among the population means have been identified.
  - When  $H_0$  is rejected in single-factor ANOVA, further analysis is carried out by applying the multiple comparisons procedure.
  - Tukey's multiple comparisons procedure involves the use of probability distribution called the Studentized range distribution
  - All of the above statements are true.
  - None of the above statements are true.
18. In a single-factor ANOVA problem involving five populations or treatments with a random sample of nine observations from each one, suppose that  $H_0$  is rejected at .05 level. Which of the following values are correct for the appropriate  $Q_{\alpha}$  critical value needed to perform Tukey's procedure?
- 4.76
  - 3.79
  - 4.04
  - 3.85
  - 4.80
19. Which of the following statements are not correct?
- The simultaneous confidence level is controlled by Tukey's method.
  - The Tukey intervals are based on independent samples.

- c) To obtain a 95% simultaneous confidence level using the Tukey method, the individual level for each interval must be considerably larger than 95%.
- d) All of the above statements are true.
- e) None of the above statements are true.
20. If three 90% confidence intervals for a population proportion  $p$  are calculated based on three independent samples selected randomly from the population, then the simultaneous confidence level will be
- About 73%
  - Exactly 90%
  - Exactly 81%
  - Exactly 270%
  - None of the above answers are correct.
21. Which of the following statements are not true?
- The two-sample t test is more flexible than the F test when the number of treatments or populations is 2.
  - The two-sample t test is valid without the assumption that the two population variances are equal.
  - The two-sample t test can be used to test  $H_a: \mu_1 > \mu_2$  or  $H_a: \mu_1 < \mu_2$  as well as  $H_a: \mu_1 \neq \mu_2$ .
  - The F test can be used to test  $H_a: \mu_1 > \mu_2$  or  $H_a: \mu_1 < \mu_2$  as well as  $H_a: \mu_1 \neq \mu_2$ .
  - When the number of treatments or populations is at least 3, there is no general test procedure known to have good properties without assuming equal populations variances.
22. In a single-factor ANOVA problem involving 4 populations, the sample sizes are 7, 5, 6, and 6. If  $SST = 65.27$  and  $SSTR = 23.49$ , then the statistic value f is
- 3.75
  - 2.09
  - 7.83
  - 0.56
  - 6.67
23. Two-way ANOVA
- Compares two means
  - Is used to fit a quadratic curve to a set of data
  - Assumes treatments have been allocated to units completely at random
  - Is used when there are two explanatory variables in regression
  - Removes two sources of systematic variation from the data
24. In performing an analysis of variance for a single-factor experiment, a fundamental assumption is made about the treatment
- Means are equal.
  - Means are unequal.
  - Variances are equal.
  - Variances are unequal.
25. One way ANOVA model is
- Fixed effect model
  - ANOVA type I
  - Random effect model
  - Both a) and b)
  - None of these
26. ANOVA type II model is
- One way ANOVA model
  - Fixed effect model
  - Random effect model
  - Both a) and c)
  - None of these
27. In a two-factor experiment, factor A consists of 4 levels, factor B consists of 3 levels, and there is one observation on each of the treatments, which of the following statements are not true?
- $SST$  has 12 degrees of freedom
  - $SSA$  has 3 degrees of freedom
  - $SSB$  has 2 degrees of freedom
  - $SSE$  has 6 degrees of freedom

- e) None of the above statements are correct.
- 3 The primary interest of designing a randomized block experiment is:
- To reduce the variation among blocks
  - To increase the between-treatments variation to more easily detect differences among the treatment means
  - To reduce the within-treatments variation to more easily detect differences among the treatment means
  - To increase the total sum of squares
  - All of the above statements are true.
- 29 Which of the following is not a longitudinal design?
- Panel
  - Cross-sectional
  - Trend
  - Both a and c are longitudinal designs
- 30 The type of design is one where all participants participate in all experimental treatment conditions.
- Factorial design
  - Repeated measures design
  - Replicated design
  - Pretest-posttest control-group design
- 31 Which of the following is the primary difference between an experiment and an observational study?
- Experiments are only conducted on human subjects; observational studies can be conducted on nonhuman subjects.
  - In an experiment, the researcher manipulates some variable to observe its effect on a response variable; in an observational study, he or she simply observes and records the observations.
  - Experiments must use randomized treatment and control groups;
- treatment and control groups, but they do not need to be randomized.
- d) Experiments must be double-blind; observational studies do not need to be.
- e) There is no substantive difference—they can both accomplish the same research goals.
- 32 If you are contemplating doing many post hoc, unplanned comparisons, you must be concerned with:
- Per-comparison error.
  - Beta errors.
  - Family wise error.
  - Probability funneling.
- 33 When you have specific pre experimental hypotheses, you can do \_\_\_\_\_ after finding a significant effect with an ANOVA.
- Planned comparisons
  - Unplanned comparisons
  - Ad hoc comparisons
  - Unweighted comparisons
- 34 If you have unequal sample sizes, you would use an unweighted means analysis if:
- Your experimental procedure caused the unequal sample sizes.
  - Your experimental procedure did not cause the unequal sample sizes.
  - The size of the sample in one group did not exceed any of the others by more than three participants.
  - Both a and b
- 35 A pharmaceutical manufacturer is designing an experiment to test four different capsule ingredients designed to reduce dissolution time. Each ingredient will be tested at 10 milligrams and 40 milligrams. A full or complete factorial design is used with five replications per run. The number of levels, factors, and runs is, respectively:
- 2, 4, 16

- b) 4, 5, 8  
c) 10, 4, 40  
d) 2, 20, 16

36. (Refer to the previous problem.) The variance of the five replications for each run is calculated. Most of these variances are approximately equal but two are significantly lower than the others. The experimenters would be especially interested in those two runs if they want to optimize:

- a) Dissolution time.  
b) Interactions.  
c) Main effects.  
d) Robustness.  
e) Degrees of freedom.

37. (Refer to the previous problem.) To estimate the within-treatment variance, the experimenters would calculate the variances of:

- a) All 80 readings.  
b) The five replications for each run.  
c) The runs for which a factor is at its lowest level.

38. You are developing a new strain of strawberries (say, Type X) and are interested in its sweetness as compared to another strain (say, Type Y). You have four plots of land, call them A, B, C, and D, which are roughly four squares in one large plot for your study (see the figure below). A river runs alongside of plots C and D. Because you are worried that the river might influence the sweetness of the berries, you randomly plant type X in either A or B (and Y in the other) and randomly plant type X in either C or D (and Y in the other). Which of the following terms best describes this design?

A	C
B	D

- a) A completely randomized design  
b) A randomized study  
c) A randomized observational study  
d) A block design, controlling for the strain of strawberry

39. In the randomized block design for ANOVA where the single factor of primary interest has  $l$  levels, and  $b$  blocks are created to control for extraneous variability in experimental units or subjects, the number of degrees of freedom for SSE (error sum of squares) is given by

- a)  $lb-1$   
b)  $(l-1)(b-1)$   
c)  $l-1$   
d)  $b-1$   
e)  $l+b-1$

40. In the fixed effects model with interaction, assume that there are 5 levels of factor A, 4 levels of factor B, and 3 observations (replications) for each of the 20 combinations of levels of the two factors. Then the number of degrees of freedom of the interaction sum of squares (SSAB) is

- a) 60  
b) 20  
c) 15  
d) 12  
e) 59

41. In the fixed effects model with interaction, assume that there are 4 levels of factor A, 3 levels of factor B and 3 observations for each of the 12 combinations of levels of the two factors. Then, the number of degrees of freedom for the error sum of squares (SSE) is

- a) 36  
b) 35  
c) 24  
d) 10  
e) 9

42. An experiment has seven factors, two levels each. The experiment has eight runs. This experimental design is called:  
 a) Full-factorial design  
 b) Half-fractional factorial design  
 c) Interaction.  
 d) None of the above.

- 43 In the fixed effects model with interaction, assume that there are 3 levels of factor A, 2 levels of factor B, and 3 observations for each of the six combinations of levels of the two factors. Then the critical value for testing the null hypothesis of no interaction between the levels of the two factors at the .05 significance level is
- 3.49
  - 3.89
  - 3.00
  - 3.55
  - 3.11
- 44 Which of the following statements are true regarding a two-factor experiment?
- In some experiments, the levels of either factor may have been chosen from a large population of possible levels, so that the effects contributed by the factor are random rather than fixed.
  - If both factors contribute random effects, the model is referred to as a random effects model.
  - If one factor is fixed, and the other contributes random effects, a mixed effects model results.
  - If both factors are fixed, the model is referred to as a fixed effects model.
  - All of the above statements are true.
- 45 The following equation  $SST = SSA + SSB + SSAB + SSE$  applies to which ANOVA model?
- One-factor ANOVA
  - Two-factor ANOVA with interaction
  - Three-factor ANOVA
  - Randomized block design
  - All of the above
- 46 In a two-factor ANOVA problem, there are 4 levels of factor A, 5 levels of factor B, and 2 observations (replications) for each combination of levels of the two factors. Then, the number of treatments in this experiment is
- 40
  - 11
  - 10
  - 20
  - 8
- 47 In the three-factor fixed effects model assumes that there are 3 levels for each of the three factors A, B, and C, and 2 observations for each combination of levels of the three factors. Then the number of degrees of freedom for the error sum of squares (SSE) is
- 54
  - 27
  - 11
  - 18
  - 16
- 48 In the three-factor fixed effects model, assume that there are 4 levels of factor A, 2 levels of factor B, 4 levels of factor C, and 3 observations for each combination of levels of the three factors. Then, the number of degrees of freedom for the three-factor interaction sum of squares (SSABC) is
- 32
  - 10
  - 9
  - 12
  - 13
- 49 The following equation  $SST = SSA + SSB + SSC + SSAB + SSAC + SSBC + SSABC + SSE$  applies to which ANOVA model?
- One-factor ANOVA
  - Two-factor ANOVA with interaction
  - Three-factor ANOVA with interactions
  - Randomized block design
  - Latin square design

50. Which of the following statements are not true?
- Tukey's multiple comparison procedure can be used in two-factor ANOVA but not in three-factor (or more) ANOVA.
  - When several factors are to be studied simultaneously, an experiment in which there is at least one observation for every possible combination of levels is referred to as complete layout.
  - A three-factor experiment, with I levels of factor A, J levels of factor B, and K levels of factor C, in which fewer than IJK observations are made is called an incomplete layout.
  - There are some incomplete layouts in which the pattern of combinations of factors is such that the analysis is straightforward. One such three-factor design is called a Latin square.
  - All of the above statements are true.
51. The following equation  $SST = SSA + SSB + SSC + SSE$  applies to which ANOVA model?
- One-factor ANOVA
  - Two-factor ANOVA with interaction
  - Three-factor ANOVA with interactions
  - Latin square design
  - Randomized block design
52. The principle disadvantage of fractional factorial experiments is that:
- Experimental error is high.
  - Robustness is compromised.
  - Effects are confounded.
  - Measurements are less precise.
  - Analysis is more difficult.
53. A cell is a combination of two or more \_\_\_\_\_ in a factorial design.
- Research designs
  - Research measurements
  - Dependent variables
  - Independent variables
54. A factorial design is one in which:
- Only one independent variable is studied to determine its effect on the dependent variable
  - Only two independent variables are simultaneously studied to determine their independent and interactive effects on the dependent variable
  - Two or more independent variables are simultaneously studied to determine their independent and interactive effects on the dependent variable
  - Two dependent variables are studied to determine their interactive effects
55. Which of the following statements are not true?
- An experiment in which there are p factors, each at two levels, is referred to as a  $p^2$  factorial experiment.
  - A  $2^p$  factorial experiment provides a simple setting for introducing the important concepts of confounding and fractional replications.
  - A  $2^p$  experiment, with four factors A, B, C, and D, has 16 different experimental conditions.
  - All of the above statements are true.
  - None of the above statements are true.
56. Which of the following statements are true?
- Blocking is always effective in reducing variation associated with extraneous sources.
  - It is often not possible to carry out all  $2^p$  experimental conditions in a  $2^p$  factorial experiment in a homogeneous experimental environment.
  - When the  $2^p$  experimental conditions are placed in  $2^r$  homogeneous blocks ( $r < p$ ), the price paid for this blocking is  $(r^2)^{1/p}$ .

- $2^1 - 1$  of the factor effects cannot be estimated.
- d) All of the above statements are true.
- e) None of the above statements are true.
57. Which of the following statements are not true?
- If the two three-factor interactions BCD and CDE are chosen for confounding, then their generalized interaction is BE.
  - If the two three-factor interactions ABC and CDE are chosen for confounding, then their generalized interaction is ABCDE.
  - When the number  $p$  of factors is large, a single replicate of a  $2^3$  experiment can be expensive and time consuming.
  - All of the above statements are true.
  - None of the above statements are true.
58. \_\_\_\_\_ refers to the influence of a single independent variable.
- Interaction effect
  - Reactive effect
  - Main effect
  - Proactive effect
59. Which of the following is possible in a factorial design with two independent variables?
- There is only one main effect present
  - There are two main effects present
  - There are two main effects and an interaction effect present
  - All of the above are possible
60. Which of the following is a factorial design where different participants are randomly assigned to the levels of one independent variable but participants take all levels on another independent variable?
- One-group pretest-posttest
  - Pretest-posttest control-group design
- c) Factorial design
- d) Factorial design based on a mixed model
61. For a treatment to be deemed effective when used in the context of an A-B-A single case design, what has to occur?
- Behavior should change as the treatment is implemented
  - Behavior should return to baseline levels when the treatment is removed
  - When the treatment is removed, behavior should stay at the level that was created by the treatment rather than revert back to the baseline
  - Both a) and b)
62. Which of the following statements are true?
- For experimental situations with more than three factors, there are often no replications, so sums of squares associated with non confounded higher-order interactions are usually pooled to obtain an error sum of squares that can be used in the denominators of the various F statistics.
  - One replicate of a  $2^6$  factorial experiment involves an observation for each of the 64 different experimental conditions.
  - If an experimenter decides to include only  $2^p - 1$  of the  $2^p$  possible conditions in the experiment; this is usually called a half-replicate.
  - The first step in selecting half-replicate is to select a defining effect as the non estimable effect.
  - All of the above statements are true.
63. Consider  $2^3$  factorial experiment in a block of size  $b$  with  $r$  replications. When ABC, AB and AC are partially confounded in each replication then degree of freedom for total is
- $r - 1$



- a) I only                    b) II only  
 c) III only                d) I and III  
 e) II and III

72. Which of the following are true statements?

- In an experiment some treatment is intentionally forced on one group to note the response.
  - In an observational study information is gathered on an already existing situation.
  - Sample surveys are observational studies, not experiments.
- a) I and II                b) I and III  
 c) II and III              d) I, II, and III  
 e) None of the above gives the complete set of true responses.

73. Which of the following are true statements?

- In an experiment researchers decide how people are placed in different groups.
  - In an observational study, the people themselves select which group they are in.
  - A control group is most often a self-selected grouping in an experiment.
- a) I and II                b) I and III  
 c) II and III              d) I, II and III  
 e) None of the above gives the complete set of true responses.

74. In one study on the effect of niacin on cholesterol level, 100 subjects who acknowledged being long time niacin takers had their cholesterol levels compared with those of 100 people who had never taken niacin. In a second study, 50 subjects were randomly chosen to receive niacin and 50 were chosen to receive a placebo.

- a) The first study was a controlled experiment, while the second was an observational study.  
 b) The first study was an observational study, while the second was a controlled experiment.  
 c) Both studies were controlled experiments.

- d) Both studies were observational studies.  
 e) Each study was part controlled experiment and part observational study.

75. In one study subjects were randomly given either 500 or 1000 milligrams of vitamin C daily, and the number of colds they came down with during a winter season was noted. In a second study people responded to a questionnaire asking about the average number of hours they sleep per night and the number of colds they came down with during a winter season.

- a) The first study was an experiment without a control group, while the second was an observational study.  
 b) The first was an observational study, while the second was a controlled experiment.  
 c) Both studies were controlled experiments.  
 d) Both studies were observational studies.  
 e) None of the above is a correct statement.

76. In a 1992 London study, 12 of 20 migraine sufferers were given chocolate whose flavor was masked by peppermint, while the remaining eight sufferers received a similar-looking, similar-tasting tablet that had no chocolate. Within 1 day, five of those receiving chocolate complained of migraines, while no complaints were made by any of those who did not receive chocolate. Which of the following is a true statement?

- a) This study was an observational study of 20 migraine sufferers in which it was noted how many came down with migraines after eating chocolate.  
 b) This study was a sample survey in which 12 out of 20 migraine sufferers were picked to receive peppermint-flavored chocolate.

- c) A census of 20 migraine sufferers was taken, noting how many were given chocolate and how many developed migraines.
- d) A study was performed using chocolate as a placebo to study one cause of migraines.
- e) An experiment was performed comparing a treatment group that was given chocolate to a control group that was not.
77. Suppose you wish to compare the average class size of mathematics classes to the average class size of English classes in your high school. Which is the most appropriate technique for gathering the needed data?
- a) Census  
 b) Sample survey  
 c) Experiment  
 d) Observational study  
 e) None of these methods is appropriate.
78. Which of the following are true statements?
- Based on careful use of control groups, experiments can often indicate cause-and-effect relationships.
  - While observational studies may suggest relationships, great care must be taken in concluding that there is cause and effect because of the lack of control over lurking variables.
  - A complete census is the only way to establish a cause-and-effect relationship absolutely.
- a) I and II      b) I and III  
 c) II and III      d) I, II, and III  
 e) None of the above gives the complete set of true responses
79. Two studies are run to compare the experiences of families living in high-rise public housing to those of families living in townhouse subsidized rentals. The first study interviews 25 families program for at least 1 year, while the second randomly assigns 25 families to each program and interviews them after 1 year. Which of the following is a true statement?
- a) Both studies are observational studies because of the time period involved.  
 b) Both studies are observational studies because there are no control groups.  
 c) The first study is an observational study, while the second is an experiment.  
 d) The first study is an experiment, while the second is an observational study.  
 e) Both studies are experiments.
80. A  $4^3$  experiment means that we are considering:
- a) Three levels of four factors  
 b) Four dependent variables and three independent variables.  
 c) Four go/no-go variables and three continuous variables.  
 d) Four levels of three factors.

## ANSWERS

1	d	2	d	3	a	4	b
5	b	6	b	7	c	8	e
9	a	10	d	11	c	12	a
13	e	14	b	15	d	16	a
17	a	18	c	19	b	20	a
21	d	22	a	23	e	24	c
25	d	26	d	27	a	28	c
29	d	30	b	31	b	32	c
33	a	34	b	35	a	36	d
37	b	38	e	39	b	40	d
41	c	42	d	43	b	44	e
45	b	46	d	47	b	48	c
49	c	50	a	51	d	52	c
53	d	54	c	55	a	56	d
57	b	58	c	59	d	60	d
61	d	62	e	63	c	64	c
65	d	66	c	67	d	68	d
69	b	70	a	71	e	72	d
73	a	74	b	75	a	76	e
77	a	78	a	79	c	80	d

# MULTIVARIATE ANALYSIS

1. If A and B are two  $n \times n$  matrices, which of the following is not always true?
  - $A+B = B+A$
  - $\text{Trace}(A+B) = \text{Trace}(B+A)$
  - $\text{Trace}(AB) = \text{Trace}(BA)$
  - $\text{Rank}(-AB) = \text{Rank}(AB)$
  - $\text{Rank}(AB) = \text{Rank}(A)$
2. The set of all linear combination of  $X_1, X_2, \dots, X_k$  is called
  - Linearity
  - Linear span
  - Linear equation
  - None of the above
3. A set of vectors  $X_1, X_2, \dots, X_n$  are linearly independent if
  - $a_1X_1 + a_2X_2 + \dots + a_kX_k = 0$
  - $a_1X_1 + a_2X_2 + \dots + a_kX_k \leq 0$
  - $a_1X_1 + a_2X_2 + \dots + a_kX_k \geq 0$
  - None of the above
4. Length of vector  $\underline{X}$  is calculated as
  - $L_x = \sqrt{X_1^2 + X_2^2 + \dots + X_k^2}$
  - $L_x = \sqrt{X_1^2 + X_2^2 + \dots + X_k^2}$
  - $L_x = \sqrt{X_1 + X_2 + \dots + X_k}$
  - None of the above
5. A matrix in which no. of rows and columns are equal, is called
  - Singular matrix
  - Square matrix
  - Triangular matrix
  - None of the above
6. A matrix  $A_{(m \times n)}$  is defined to be orthogonal if
  - $A'A = I$
  - $A^{-1} = A'$
  - $A'A = AA' = I$
  - All of the above
7. If A is a square matrix of order ( $m \times m$ ) then the sum of diagonal elements is called
  - Sum of matrix
  - Trace of matrix
  - Determinant of matrix
  - None of the above
8. The rank of a matrix
 
$$\begin{bmatrix} 1 & 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 & 2 \\ 1 & 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 1 & 3 \end{bmatrix}$$
 is
  - 1
  - 2
  - 3
  - 4
  - 5
9. If A is a Square matrix, then  $\det(A - \lambda I) = 0$  is known as
  - Eigenvector
  - Eigenvalues
  - Characteristic equation
  - Both b) & c)
  - None of the above
10. In multivariate Analysis var-cov matrix is
  - $\Sigma = E(X - \mu)(X - \mu)'$
  - $\Sigma = E(X - \mu)(X - \mu)'$
  - Both a & b
  - None of the above

11. In the relation  $\sum = V^2 \rho V^2$ ,  $V^2$  is called  
 a) Variance matrix  
 b) Standard deviation matrix  
 c) Covariance matrix  
 d) None of the above
12. If  $X \sim N_p(\mu, \Sigma)$  then  $(X - \mu)' \Sigma^{-1} (X - \mu)$  is distributed as  
 a)  $N_p(\mu, \Sigma)$   
 b)  $\chi^2$  with  $P$  degree of freedom  
 c) Beta distribution  
 d) None of the above
13. In multivariate analysis,  $n(\bar{x} - \mu)' S^{-1} (\bar{x} - \mu)$  is called  
 a) Var-cov matrix  
 b) Hotelling T Square  
 c) Multivariate Normal Distribution  
 d) None of the above
14. In multivariate analysis the distribution of  $\bar{X}$  is  
 a) Chi-Square  
 b) Normal  
 c) T-Square  
 d) None of the above
15. In multivariate analysis the distribution of sample covariance matrix is  
 a) T-Square distribution  
 b) Chi-Square distribution  
 c) Wishart distribution  
 d) Normal distribution
16. In Factor analysis the reliable variance  
 a) Total variance + Error variance  
 b) Total variance - Error variance  
 c) Common variance + Specific Variance  
 d) Both b) & c)  
 e) None of the above
17. In principal component analysis (PCA), the first component contains  
 a) Minimum variance  
 b) Maximum variance  
 c) Both a) and b)  
 d) None of these
18. In principal component analysis components are  
 a) Correlated  
 b) Uncorrelated  
 c) Both a) and b)  
 d) None of these
19. In PCA, when the variables measure in different units then PC's represent the basis of  
 a) Variance covariance matrix  
 b) Correlation matrix  
 c) Communalties  
 d) Specific variance
20. The goal of multiple regression is to  
 a) Explain variation in one variable based on variation in a set of others.  
 b) Reduce a large number of correlated measures to a smaller number.  
 c) Develop causal explanations based on correlational data  
 d) Test the effect of set of independent variables on a single dependent variable.
21. A multivariate statistic that allows you to investigate the relationship between two sets of variables is  
 a) Multiple regression.  
 b) MANOVA  
 c) Canonical correlation.  
 d) Discriminant analysis.
22. Correlational multivariate analyses include  
 a) Discriminant analysis.  
 b) Multiple regression.  
 c) Canonical correlation.  
 d) Both a) and b).  
 e) All of the above
23. An advantage of using an experimental multivariate design over separate univariate designs is that using the multivariate analysis:  
 a) Allows you to look at more complex relationships than does the univariate strategy.  
 b) Provides a more powerful test of your hypotheses.

- c) Allows you not to worry about meeting restrictive assumptions characteristic of univariate statistics.
- d) Both a) and b)
- 24 A multivariate statistic that allows you to analyze several dependent variables from an experimental design simultaneously is:
- ANOVA.
  - MANOVA.
  - Factor analysis.
  - Path analysis.
- 25 \_\_\_\_\_ is used for causal analysis.
- Path analysis
  - MANOVA
  - Factor Analysis
  - Canonical Correlation
- 26 The presence of outliers in your data:
- Affects the magnitude of the correlations calculated but not the slope of the regression line.
  - Affects the slope of the regression line but not the magnitude of the correlations calculated.
  - Affects both the slope of the regression line and the magnitude of the correlations calculated.
  - Affects both the slope of the regression line and the magnitude of the correlations calculated.
- 27 An effective way of detecting multivariate outliers is to:
- Convert raw scores to z scores and evaluate the degree of deviance of the z scores.
  - Conduct individual Pearson correlations on your data before conducting any multivariate test.
  - Do nothing; outliers do not significantly affect multivariate statistics.
  - Both a) and b)
- 28 Generally speaking, multivariate analysis requires:
- Fairly large samples.
  - Small samples.
- c) Less concern over meeting assumptions than do univariate tests.
- d) Sampling from a population that is not normally distributed.
- 29 In a factor analysis, the correlation between an individual variable and an underlying dimension is a:
- Discriminant function.
  - Factor loading.
  - Squared semipartial correlation.
  - Canonical function.
- 30 The factors extracted in a factor analysis are made more clear and interpretable by:
- Converting raw scores to z scores before analysis.
  - Eliminating variables that have low correlations with other variables.
  - Applying a square root transformation to the raw data before analysis.
  - Statistically rotating factors.
- 31 In factor analysis, \_\_\_\_\_ assumes that your measures are correlated.
- Parallel rotation
  - Orthogonal rotation
  - Oblique rotation
  - All of the above
- 32 According to Tabachnick and Fidell (2001), principal components analysis could be used to:
- Help infer causality from correlational data.
  - Extract as many factors as possible from your data before a factor analysis.
  - Experiment with different communality values after an exploratory factor analysis.
  - Determine the degree of contribution of a variable in a multiple regression analysis.
- 33 \_\_\_\_\_ is a statistical technique used to evaluate the relationship between two variables statistically controlling the effects of third variable
- Discriminant analysis
  - Canonical correlation

- c) Partial correlation  
d) Factor analysis
34. A statistical technique that involves entering multiple predictor variables into an equation according to a specified order determined by theory is:
- Hierarchical regression.
  - Simple regression.
  - Stepwise regression.
  - None of the above
35. The use of stepwise regression techniques is frowned on because:
- Only three predictor variables can be entered at a time.
  - It tends to be too sensitive to causal relationships among variables.
  - It tends to capitalize on chance and may be limited to a particular sample.
  - All of the above
36. If you want to estimate how much of the variability in the criterion variable can be accounted for by variation in each predictor in multiple regression analysis, you should use:
- Unstandardized regression weights.
  - Standardized regression weights.
  - Factor loadings.
  - Squared semi partial correlations.
37. If you have multiple predictor variables and a dichotomous dependent variable, the most appropriate multivariate test is:
- Stepwise regression.
  - Canonical correlation.
  - Factor analysis.
  - Discriminant analysis.
38. Loglinear analysis:
- Is a nonparametric statistic.
  - Works much like chi-square.
  - Can be used in place of ANOVA, MANOVA, or multiple regression where your data are categorical.
  - All of the above.
39. If you have two sets of variables to correlate, the most appropriate multivariate test is:
- Stepwise regression.
  - Canonical correlation.
  - Factor analysis.
  - Discriminant analysis.
40. After finding a significant effect with MANOVA, you would use \_\_\_\_\_ to assess each dependent variable's contribution to the significant effect.
- Roy-Bargman stepdown analysis.
  - Univariate ANOVAs
  - T test
  - Loglinear analysis
41. Using a MANOVA in place of a univariate analysis for a within-subjects experiment is advantageous because MANOVA:
- Allows you to circumvent some of the restrictive assumptions of the univariate within-subjects ANOVA.
  - Allows you to include more than two independent variables in your analysis.
  - Uses separate error terms to test effects rather than a pooled error term.
  - None of the above
42. In research situations in which you want to measure or manipulate categorical variables, an appropriate alternative to statistics such as ANOVA, MANOVA, or multiple regression would be:
- Canonical correlation.
  - Multiple t tests.
  - Path analysis.
  - Multivariate frequency analysis.
43. The statistic used to evaluate data in a loglinear analysis is:
- $G^2$ .
  - F.
  - d.
  - Chi-square.
44. Path analysis is:
- A unique statistical test, allowing you to evaluate multiple dependent variables in one test.

- b) An application of multiple regression to investigating causal relationships among variables.

c) Not used to investigate causal relationships, but is a multivariate statistic.

d) An extension of the Pearson  $r$  to multivariate designs.

In path analysis, \_\_\_\_\_ variables start a causal sequence.

a) Endogenous

b) Latent

c) Exogenous

d) Exothermic

In structural equation modeling, \_\_\_\_\_ variables are variables in a model that are not directly observable.

a) Endogenous

## NON PARAMETRIC TESTS

1. In non-parametric method, which of the following assumption is not hold?
  - Homoscedasticity
  - Normality
  - Randomness
  - None of these
  
2. In Non-parametric method, which of the following is used as a measure of central tendency
  - Mean
  - Mode
  - Median
  - Weighted mean
  
3. When normality is doubtful then ANOVA and Kruskal-Wallis tests give
  - Same conclusion
  - Different conclusion
  - Identical conclusion
  - Both a) and c)
  
4. When symmetry is in doubt, and then the test is recommended.
  - Fisher test
  - Wilcoxon signed rank test
  - Both a) and b)
  - None of these
  
5. The Mann-Whitney test is a nonparametric test median of
  - One group
  - Two group
  - Two or more groups
  - None of these
  
6. Kruskal-Wallis is a generalization of
  - Fisher test
  - Wilcoxon signed rank test
  - Mann-Whitney test
  - None of these
  
7. Critical value for Kruskal-Wallis test is calculated by
  - Normal distribution
  
8. Which of the following tests would be an example of a nonparametric procedure?
  - Wilcoxon signed-rank test
  - The  $t$  test for population mean  $\mu$
  - The  $F$  test for population means  $\mu_1, \mu_2, \dots, \mu_k$
  - All of the above tests are correct
  - Only b) and c) are correct tests
  
9. When ranking data in a Wilcoxon signed-rank test, the data value that receives a rank of 1 is the
  - Largest value regardless of its size
  - Smallest value regardless of its size
  - Middle value regardless of its size
  - 25th percentile value
  - 75th percentile value
  
10. Which of the following statements are not true?
  - The  $t$  and  $F$  procedures are not "distribution-free" procedures because they require the distributed assumption of normality.
  - The  $t$  and  $F$  procedures are not "nonparametric" procedures because they are based on the normal parametric family of distribution.
  - Distribution-free and nonparametric procedures are valid for very few different types of underlying distributions
  - Generally speaking, the distribution-free procedures perform almost as well as their  $t$  and  $F$  counterparts on the "home

"ground" of the normal distribution, and will often yield a considerable improvement under nonnormal conditions.

- e) All of the above statements are true.
11. A random sample of size 15 is drawn from a continuous and symmetric probability distribution with mean  $\mu$ . In testing  $H_0: \mu = 300$  versus  $H_a: \mu \neq 300$  using the Wilcoxon signed-rank test with approximate level of significance of .05, the rejection region for the test is
- $s_+ \geq 95$
  - $s_+ \leq 25$
  - Either  $s_+ \geq 95$  or  $s_+ \leq 25$
  - $s_+ \leq 95$
  - $s_+ \geq 25$
12. Which of the following statements are true?
- When the data consists of pairs  $(X_1, Y_1), \dots, (X_n, Y_n)$  and the differences  $D_i = X_i - Y_i$  ( $i = 1, 2, \dots, n$ ) are not assumed to be normally distributed, hypotheses tests about the expected differences  $\mu_D$  can be tested by using the Wilcoxon signed-rank test on the  $D_i$ 's provided that the distribution of the differences is continuous and symmetric.
  - When the sample size  $n$  is larger than 20, it can be shown that the Wilcoxon signed-rank test statistic shows approximately a normal distribution when the null hypothesis is true.
  - When the underlying distribution being sampled is normal, either the  $t$  test or the Wilcoxon signed-rank test can be used to test a

hypothesis about the population mean  $\mu$ .

- A number of different efficiency measures have been proposed by statisticians, one that many statisticians regard as credible is called asymptotic relative efficiency (ARE).
  - All of the above statements are true.
13. Which of the following statements are not true?
- When the underlying distribution being sampled has "heavy tails", that is, when observed values lying far from population mean  $\mu$  are relatively more likely than they are when the distribution is normal, the  $t$  test can perform poorly.
  - If the asymptotic relative efficiency (ARE) of one test with respect to a second equals .50, then when sample sizes are large, twice as large a sample size will be required of the first test to perform as well as the second test.
  - When the underlying distribution is normal, the asymptotic relative efficiency of the Wilcoxon signed-rank test with respect to the  $t$  test is approximately 0.95.
  - For any distribution, the asymptotic relative efficiency will be at least .86, and for many distributions it will be much greater than 2.
  - All of the above statements are true.
14. Which of the following statements are true?
- In large-sample problems, the Wilcoxon signed-rank test is never very much less efficient than the  $t$  test and may be much more efficient if the underlying distribution is far from normal.
  - The Wilcoxon signed-rank test statistic for large-sample is  $(1)/4$ .

$$z = \frac{s_+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}, \text{ where } n$$

is the sample size and  $S$  is the sum of the ranks associated with the positive observations.

- c) When the sample size  $n > 20$ , the Wilcoxon signed-rank test statistic shows approximately a normal distribution with mean and variance given by  $n(n-1)/4$  and  $n(n-1)(2n-1)/24$ , respectively
- d) All of the above statements are true.
- e) None of the above statements are true.

15. A random sample of size 24 is drawn from a continuous and symmetric probability distribution with mean  $\mu$ . In testing  $H_0: \mu = 200$  versus

$H_a: \mu \neq 200$  it can be shown that the Wilcoxon signed-rank test statistic shows approximately a normal distribution with mean and standard deviation given, respectively, by

- a) 150 and 35
- b) 25 and 300
- c) 150 and 300
- d) 35 and 25
- e) 25 and 35

16. Which of the following tests would be an example of a distribution-free procedure?

- a) The  $t$  test for population mean  $\mu$
- b) The paired  $t$  test for the expected difference  $\mu$
- c) The  $F$  test for two or more population means
- d) The Wilcoxon rank-sum test
- e) Only A and B are correct tests

17. Which of the following statements are not true?

- a) When at least one of the sample sizes in a two-sample problem is small, the  $t$  test requires the

- b) assumption of normality (at least approximately). The Wilcoxon rank-sum test statistic  $W$  is the sum of the ranks in the combined  $(X, Y)$  sample associated with  $X$  observations
- c) Because the Wilcoxon rank-sum test statistic  $W$  has a continuous probability distribution, there will always be a critical value corresponding exactly to one of the usual levels of significance
- d) All of the above statements are true.
- e) None of the above statements are true.

18. Which of the following statements are true?

- a) The Wilcoxon rank-sum test procedure is not distribution-free because it will not have the desired level of significance for a very large class of underlying distributions.
- b) If there are three observed values of  $x$  and five observed values of  $y$ , then the smallest possible value of the Wilcoxon rank-sum test statistic  $W$  is  $w = 6$  and the largest possible value is  $w = 21$ .
- c) When the distributions being sampled are both normal with  $\sigma_1 = \sigma_2$ , and therefore have the same shapes and spreads, only the pooled  $t$  test can be used in testing

- $H_0: \mu_1 - \mu_2 = \nu_0$  whereas the Wilcoxon rank-sum test should not be used because it is distribution-free.
- d) When normality and equal variances both hold, the Wilcoxon rank-sum test is approximately 75% as efficient as the pooled  $t$  test in large samples.
  - e) All of the above statements are true.

18. Two independent random samples of sizes 5 and 7 are selected from two continuous distributions with means  $\mu_1 = \mu_2$  and that the two distributions have the same shape and spread. In testing  $H_0: \mu_1 - \mu_2 = 0$  versus  $H_a: \mu_1 - \mu_2 \neq 0$  using the Wilcoxon rank-sum test with approximate significance level of .05, the rejection region for the test is

- a) Either  $w \geq 20$  or  $w \leq 45$
- b) Either  $w \geq 10$  or  $w \leq 20$
- c) Either  $w \geq 45$  or  $w \leq 10$
- d) Either  $w \geq 10$  or  $w \leq 45$
- e) Either  $w \geq 45$  or  $w \leq 20$

20. A sample of size 8 is selected at random from a continuous symmetric distribution. A 95% Wilcoxon signed-rank interval (actually 94.5%) has the form

- a)  $\bar{x}_{32}, \bar{x}_{72}$
- b)  $\bar{x}_{32}, \bar{x}_{38}$
- c)  $\bar{x}_3, \bar{x}_{32}$
- d)  $\bar{x}_4, \bar{x}_5$
- e)  $\bar{x}_{36}, \bar{x}_{72}$

21. Which of the following statements are not true?

- a) The efficiency of the Wilcoxon signed-rank interval relative to the t interval is roughly the same as that for the Wilcoxon test relative to the t test.
- b) For large samples when the underlying population is normal, the Wilcoxon signed-rank interval will tend to be slightly longer than the t interval.
- c) For large samples when the underlying population is quite nonnormal (symmetric but with heavy tails), then the Wilcoxon signed-rank interval will tend to be much shorter than the t interval.

d) All of the above statements are true.

e) None of the above statements are true.

22. The Kruskal-Wallis test is always

- a) Two-tailed test.
- b) One-tailed test.
- c) Used with one sample.
- d) Used when the populations are normally distributed.
- e) Used with match-pairs samples

23. The Friedman's test is always

- a) Two-tailed test
- b) One-tailed test
- c) Used with one sample
- d) Used when the populations are normally distributed.
- e) Used with matched-pairs samples

24. Which of the following distributions approximate the Kruskal-Wallis test statistic  $K$ , when  $H_0: \mu_1 = \mu_2 = \dots = \mu_l$  is true, and the number of populations or treatments  $l=3$  and the sample sizes  $J_i \geq 6$  ( $i = 1, 2, 3$ ) or  $l > 3$  and  $J_i \geq 5$  ( $i = 1, 2, \dots, l$ )?

- a) Standard normal distribution
- b) T distribution with  $l-1$  degrees of freedom
- c) F distribution with  $l-1$  and  $J - l$  degrees of freedom.
- d) Chi-squared distribution with  $l-1$  degrees of freedom.
- e) Either B or C.

25. Which nonparametric test is analogous to a parametric t-test for independent sample means?

- a) Wald-Wolfowitz test.
- b) Wilcoxon signed rank test.
- c) Mann-Whitney test
- d) Kruskal-Wallis test.

26. Which nonparametric test is analogous to a one-factor ANOVA?

- a) Kruskal-Wallis test.
- b) Friedman test.
- c) Mann-Whitney test
- d) Wilcoxon signed rank test

- 27 Which nonparametric test is analogous to a parametric one-sample t-test for differences in paired data?
- Wald-Wolfowitz test.
  - Wilcoxon signed rank test
  - Mann-Whitney test.
  - Kruskal-Wallis test.
- 28 Which parametric test resembles the nonparametric Spearman's rank test?
- The t-test of a correlation coefficient.
  - The t-test of two sample means.
  - The t-test of one sample mean.
  - The one factor ANOVA.
- 29 Which is not true of the one-sample runs test?
- It is also called the Wald-Wolfowitz test after its inventors.
  - Its purpose is to detect non-randomness.
- c) It cannot be applied to sequential binary data  
d) It is similar to test for autocorrelation.
30. Which nonparametric test is used to compare one sample median with a benchmark?
- Wald-Wolfowitz test.
  - Wilcoxon signed rank test.
  - Mann-Whitney test.
  - Kruskal-Wallis test.

**ANSWERS**

1	b	11	c	21	d
2	c	12	e	22	b
3	d	13	d	23	b
4	a	14	d	24	d
5	b	15	a	25	c
6	c	16	d	26	a
7	b	17	c	27	b
8	a	18	b	28	a
9	b	19	e	29	c
10	c	20	c	30	b

## **STATISTICAL QUALITY CONTROL**

- 1 A quality control program is considered to be:
- A collection of quality control procedures and guidelines.
  - A step-by-step listing of all quality control check points.
  - A summary of company quality control policies.
  - A system of activities to provide quality of products and service.
- 2 The Juran trilogy for quality management consists of:
- Planning, execution, reward.
  - Process, product, shipment.
  - Planning, control, improvement.
  - Production, perspective, payment.
- 3 A method of dealing with an inspector found to be falsifying the results of inspection of borderline product is to:
- Criticize the inspector on the basis that the pattern of reading does not follow the laws of chance.
  - Review the procedure for evaluating and reporting borderline product.
  - Review the inspector's results against the expected results calculated from a normal curve.
  - Criticize the inspector for not knowing how to read the inspection equipment.
- 4 ISO 9000 is an example of:
- A domestic standard.
  - An industry association standard.
  - An international standard.
  - An isometric standard.
  - None of the above.
- 5 Which of the following are perceived benefits of being ISO certified?
- Employees must share information and agree on which practices are best.
  - Inventory is reduced because there are smaller batch sizes.
  - Training costs are reduced since the processes are well documented.
  - Both a) and b)
  - Both a) and c)
- 6 Which of the following are problems associated with gaining ISO accreditation?
- Resources must be devoted to something not on the value stream.
  - Managers may be accused of "tree hugging" because fear can be useful.
  - Employees rarely feel stifled because of bureaucratic hurdles are eliminated.
  - Both a) and b)
  - Both a) and c)
- 7 Deming estimated that \_\_\_\_\_% of quality problems are due to the system rather than the workers.
- 15
  - 50
  - 75
  - 85
- 8 The probability that defect type A occurs is .83, the probability that defect type B occurs is .83, and the probability that both occur is .83. Find the probability that at least one of the defects occurs.
- .83
  - .92
  - .33
  - 1.05
  - .029

The probability of observing at least one defective in a random sample of size 10, drawn from a population that has been producing, on the average, 10 percent defective units, is:

- a)  $(0.10)^{10}$       b)  $(0.90)^{10}$   
c)  $1 - (0.10)^{10}$       d)  $1 - (0.90)^{10}$
10. Let the random variable  $X$  denotes the number of defective items in the lot,  $A$  denotes the event that the lot is accepted, and  $p$  denotes the proportion of defective items in the lot. Which of the following statements is not true?
- If the sample size  $n$  is large relative to the lot size  $N$ , then the probability of accepting the lot,  $P_A$ , is calculated using the Hypergeometric distribution.
  - When the sample size  $n$  is small relative to the lot size  $N$  (the rule of thumb i.e.  $n \leq 0.05N$ ), then the probability of accepting the,  $P_A$ , is calculated using the binomial distribution.
  - If the probability of accepting the lot,  $P_A$ , is large only when  $p$  is small (this, of course, depends on the specified critical value  $c$ ), then the Poisson approximation to the binomial distribution is justified.
  - The larger value of  $p$ , the larger the probability  $P_A$  of accepting the lot.
  - All of the above statements are true.
11. If prevention costs are increased to pay for engineering work in quality control, and this result in a reduction in the number of product defects, this yields a reduction in:
- Appraisal costs.
  - Operating costs.
  - Quality costs.
  - Failure costs.
12. The primary reason for evaluating and maintaining surveillance over a supplier's quality program is to
- Perform product inspection at the source.
  - Eliminate incoming inspection costs.
  - Motivate suppliers to improve quality.
  - Make sure the supplier's quality program is functioning effectively.
13. When planning a total quality system, one key objective is to provide a means of guaranteeing the maintenance of product integrity. Which of the following quality systems provisions is designed to most directly provide such a guarantee?
- Drawing and print control
  - Calibration and maintenance of test equipment
  - Identification and segregation of nonconforming material
  - Specification change control
14. The most desirable method of evaluating a supplier is:
- A history evaluation.
  - A survey evaluation.
  - A questionnaire.
  - A discussion with the quality manager on the phone.
15. In recent months, several quality problems have resulted from apparent changes in design specifications by engineering, including material substitutions. This has only come to light through quality engineering's failure analysis system. You recommend establishing which of the following quality system provisions as the best corrective action?
- A formal procedure for initial design review
  - A formal procedure for specification change control (sometimes called an ECO or SCO system)

- c) A formal system for drawing and print control
- d) A formal material review board (MRB) system
16. One of the most important reasons for a checklist in an "in-process" audit is to:
- Assure that the auditor is qualified.
  - Obtain relatively uniform audits.
  - Minimize the time required to complete the audit.
  - Notify the audited function prior to the audit.
17. You contract with a company to audit one of their offshore suppliers. This type of audit is called a:
- Client audit.
  - Third-party audit.
  - Follow-up audit.
  - Registration audit.
  - Certification audit.
18. The step of the audit that is most often poorly completed is:
- Audit team training.
  - Corrective action and verification.
  - Notification and preparation.
  - Execution.
  - Exit meeting.
19. In most audit situations, it is best for the auditor to:
- Notify the auditee in advance, unless prohibited by regulations.
  - Conduct a surprise audit.
  - Provide the auditee with the audit checklist at, but not before, the opening meeting.
  - None of the above.
20. In an audit team meeting following the exit meeting, a team member suggests an additional finding. It is appropriate to:
- Add the finding to the final report.
  - Omit the finding since it was not mentioned in the exit session.
21. The main purpose of a quality audit is to:
- Evaluate management commitment.
- b) Determine whether products conform to specifications
- c) Assure that the organization is financially sound
- d) Obtain information regarding the quality system
- e) Determine whether quality personnel are certified by a third party
22. ASQ certification is:
- Licensure.
  - Peer recognition.
  - Registration.
  - Governmentally regulated.
23. Which line below is not found in the ASQ Code of Ethics?
- Will do whatever I can to promote the reliability and safety of all products that come within my jurisdiction
  - Will not disclose information concerning the business affairs or technical processes of any present or former Will strive to improve each process I encounter
  - Will take care that credit for the work of others is given to those to whom it is due
24. The ASQ definition of quality is:
- The efficient production of products that the customer expects.
  - Conformance to expectations.
  - Conformance to specifications.
  - The composite of marketing, engineering, and manufacturing, through which the product or service will meet the expectations of the customer.
  - The totality of features and characteristics of a product that affect its ability to satisfy a given need.
25. The primary purpose of audit working papers is to provide:
- Evidence of analysis of internal control.
  - Support for the audit report.

- c) A basis for evaluating audit personnel.
- d) A guide for subsequent audits of the same areas.
26. A quality engineer, when asked about details of a technical process used by his employer, is bound by the ASQ Code of Ethics to:
- Obtain the consent of the employer first.
  - Divulge no information.
  - Inform his employer of the request.
  - Provide information verbally but provide no documents.
27. It is usually true that increasing the resources devoted to prevention efforts will:
- Increase total cost of quality.
  - Decrease failure costs.
  - Increase quality department visibility.
  - Cause cost accounting miscalculations.
28. Which of the following elements is least necessary to a good corrective action feedback report?
- What caused the failure
  - Who caused the failure
  - What correction has been made
  - When the correction is effective
29. A classification of characteristics makes it possible to:
- Separate the "vital few" from the "trivial many" kinds of defects.
  - Direct the greatest inspection effort to the most important quality characteristics.
  - Establish inspection tolerances.
  - Allow the inspector to choose what to inspect and what not to inspect.
30. Dodge-Romig tables are designed to minimize which parameter?
- AOQL
  - AQL
  - ATI
  - AOQ
31. The ASQ Code of Ethics includes the following four sections:
- Fundamental Principles, Relations with the Public, Relations with Employers and Clients, Relations with Peers.
  - Fundamental Principles, Legal Obligations, Relations with Employers and Clients, Relations with Peers.
  - Fundamental Principles, Relations with the Public, Relations with Employers and Clients, Rights and Responsibilities.
  - Appropriate Use of Knowledge, Relations with the Public, Relations with Employers and Clients, Relations with Peers.
  - None of the above.
32. The process of studying "best practices" of other organizations is called:
- Good manufacturing practice (GMP).
  - Quality initiative analysis.
  - Policy study and deployment (PS&D).
  - Benchmarking.
  - None of the above.
33. The people responsible for leading quality initiatives need to:
- Involve all those impacted in the decision-making process.
  - Consider the unanticipated consequences of any proposed changes.
  - Establish a mechanism to verify that the proposed change has the desired affect.
  - Communicate the results to others.
  - All of the above.
34. When strong disagreements arise between team members:
- They should be encouraged to ignore their differences for the sake of team harmony.
  - The team should explore data collection schemes that would help each member take a different position. The

- team should vote as soon as possible.
- c) The team should seek outside mediation to solve the problem.  
d) The team should report that it is unable to reach a conclusion.
- 35 The most useful measure of the effectiveness of training is:  
a) The score on a final examination.  
b) The opinion survey at the end of the class.  
c) A pre-test and post-test comparison.  
d) The impact on processes and products.
36. The expenditure for new equipment to be used for measuring functionality of products belongs in which quality cost category?  
a) Appraisal  
b) Prevention  
c) Internal failure  
d) External failure  
e) None of the above
- 37 Expenses incurred for a quality engineer's visit to determine whether a supplier is meeting the specifications on an existing purchase order belong under which quality cost category?  
a) Appraisal  
b) Prevention  
c) Internal failure  
d) External failure  
e) none of the above
- 38 The quality leader most often associated with robust design is:  
a) Juran  
b) Deming  
c) Taguchi  
d) Ishikawa
- 39 The quality improvement cycle referred to as either PDCA or PDSA is credited to which two people?  
a) Deming and Shewhart  
b) Juran and Taguchi
- 40 Robustness of processes and products refers to:  
a) The resistance to change caused by environmental variables  
b) The ability to produce the first part within specifications  
c) The tendency to use less than half the total tolerance  
d) The likelihood that the product or service will exceed customer expectations.
- 41 Customer complaints for products with low unit price are usually:  
a) Easily satisfied.  
b) Not serious.  
c) Underreported.  
d) Due to subterfuge.  
e) Due to shortages.
42. When gathering data about product performance, one should keep in mind that customer complaint:  
a) Provide reliable performance data.  
b) Should be the only source for product performance data.  
c) Are poor measures of product performance?  
d) Are useless as an indicator of customer satisfaction.
43. Deming recommends single suppliers for purchased parts. Advantages of this approach include size of contract will be larger and therefore may command more attention from supplier.  
I. Minimal disruption due to strikes.  
II. Simplified communication.  
III. Increased competition between suppliers for each order.  
a) I and II  
b) I and IV  
c) III and IV  
d) I and III  
e) II and IV
44. Joseph Juran delineated three managerial processes that are interrelated as the Juran Trilogy. These may be summarized as:

- III. Quality training  
 IV. Quality mission and vision statement.  
 V. Quality planning.  
 VI. Quality control.  
 a) IV, V, III  
 b) V, VI, I  
 c) IV, II, VI  
 d) II, IV, V  
 e) II, V, I
45. Where inspector efficiency is defined as the ratio of correct decisions to the total decisions regarding individual items, most inspection operations performed by human inspectors are approximately:  
 a) 40-55% efficient.  
 b) 55-70% efficient.  
 c) 70-95% efficient.  
 d) 95-100% efficient.
46. In a visual inspection situation, one of the best ways to minimize deterioration of the quality level is to:  
 a) Retrain the inspector frequently.  
 b) Have a program of frequent eye exams.  
 c) Add variety to the task.  
 d) Have a standard to compare against as an element of the operation.
47. The quality engineer should be concerned with the human factors of a new piece of in-house manufacturing equipment as well as its operational effects because it:  
 I. May speed the line to the point where a visual operator inspection is impossible.  
 II. May require the operator's undivided attention at the controls so the product cannot be fully seen.  
 III. May remove an operator formerly devoting some portion of time to inspection.  
 a) I only  
 b) II only  
 c) I and III only  
 d) I, II, and III
48. One element of a quality system is product verification. This activity includes:  
 a) Use of test and inspection points in processes to verify conformance.  
 b) Verification of incoming materials.  
 c) Final product verification.  
 d) All of the above.  
 e) a) and c) only.
49. Requirements for document control include all of the following except:  
 a) A process for generation, approval, and distribution of documents.  
 b) A process to ensure ready availability of documents where they are needed.  
 c) A process for distribution of revisions and the removal of obsolete documents.  
 d) A process for making the documents available to the general public.
50. The difference between an internal and an external audit is:  
 a) An internal audit is conducted indoors while an external audit is conducted outside.  
 b) For an internal audit, the auditors are employees of the auditee organization, whereas for an external audit they are not.  
 c) An external audit is always conducted by a third party.  
 d) An internal audit is done by the auditee in preparation for an external audit.
51. One difference between a "finding" and an "observation" in an audit report is:  
 a) An observation is supported by one or more findings.  
 b) A finding is supported by one or more observations.  
 c) Findings detail weaknesses found in the documentation and observations refer to weaknesses in the implementation of the documentation.

2. Findings are discovered accidentally while observations are intentionally sought.
5. The technique of seeking out and studying the best products and processes in other divisions or companies with the intent of continuous improvement is called:
- Quality function deployment.
  - Benchmarking
  - Using Baldridge criteria
  - Quality auditing.
53. Precontrol starts a process specifically centered between:
- Process limits.
  - Safety lines.
  - Normal distribution limits.
  - Three-sigma limits.
  - Specification limits.
54. A large number of lots have shown up on a shipping dock, and their quality has not been ascertained. Which method(s) would be obviously helpful?
- Acceptance sampling
  - DOE
  - Formal optimization
  - Both a) and b)
  - Both a) and c)
55. Which of the following statements are true?
- Until quite recently, control chart procedures and acceptance sampling techniques were regarded by practitioners as equally important parts of quality control methodology, but this is no longer the case.
  - Acceptance sampling deals with what has already been produced and thus does not provide for any direct control over process quality.
  - The most straightforward type of acceptance sampling plan involves selecting a single random sample of size  $n$  and then rejecting the lot if the number of defectives in the sample exceeds a specified critical value  $c$ .
  - We want an operating characteristic (OC) curve that is higher for very small  $p$  (proportion of defective items) and lower for larger  $p$ . This can be achieved by increasing the sample size  $n$  and the specified critical value  $c$ .
  - All of the above statements are true.
56. In acceptance sampling, the risk of accepting a poor quality lot is considered a
- Type I error
  - Consumer's risk
  - Producer's risk
  - Bad luck lot
  - None of the above.
57. In acceptance sampling, the risk of rejecting a good quality lot is considered a
- Type II error
  - Consumer's risk
  - Producer's risk
  - Bad luck lot
  - None of the above
58. The technique most directly associated with guaranteeing that all measurement equipment are capable and critical characteristics are being monitored is
- Process mapping
  - Benchmarking
  - Design of Experiments (DOE)
  - Control Planning
  - Acceptance Sampling
59. Which of the following is most relevant to cost-effective evaluation of many units?
- Benchmarking
  - Control planning
  - Acceptance sampling
  - Design of Experiments (DOE)
  - A reaction plan
60. In ANSI/ASQ Z1.4-2003, the AQL is always determined at what probability of acceptance ( $P_a$ ) on the OC curve?
- 0.05
  - 0.10
  - 0.90
  - 0.95
  - None of the above

- 61 ANSI/ISO/ASQ Q9001-2000 replaces
- ANSI/ISO/ASQC Q9001-1994.
  - ANSI/ISO/ASQC Q9002-1994.
  - ANSI/ISO/ASQC Q9003-1994.
  - All of the above.
  - None of the above.
- 62 In comparison with attributes sampling plans, variables sampling plans
- Have the advantage of greater simplicity.
  - Usually require a larger sample size for comparable assurance as to the correctness of decisions in judging a single quality characteristic.
  - Have the advantage of being applicable to either single or multiple quality characteristics.
  - Provide greater assurance, for the same sample size, as to the correctness of decisions in judging a single quality characteristic.
- 63 An operation requires shipments from your vendor of small lots of fixed size. The attribute sampling plan used for receiving inspection should have its OC curve developed using
- The binomial distribution.
  - The Gaussian (normal) distribution.
  - The Poisson distribution.
  - The hypergeometric distribution.
- 64 Using ANSI/ASQ Z1.4-2003, determine the sample size for a lot of 300. AQL = 1.0%, level II normal inspection, double sampling
- 50, 50
  - 20, 20
  - 32, 32
  - 80, 100
- 65 The acronym "AQL," as used in sampling inspection means
- That level of lot quality for which there is a small risk of rejecting the lot.
  - The average quality limit
- c) The maximum percent defective that can be considered satisfactory as a process average.
- d) The quality level.
- 66 Lots of 75 parts each are inspected to an AQL of 0.25% using normal inspection, level II single sampling. Assuming general inspection level II, what is the sample size for ANSI/ASQ Z1.4-2003?
- 13
  - 20
  - 50
  - 75
- 67 Complete inspection is (roughly speaking) a single sampling plan with
- $n = d$
  - $N = c$
  - $N = n$
  - $c = d$
  - None of these describe complete inspection even roughly speaking
- 68 A team studies a coil steel banding process and makes five changes resulting in productivity improvements of 2%, 2.8%, 2.4%, 2%, and 3% respectively. These improvements are best described by which approach to problem solving?
- 5S
  - Pa-yé
  - Kaizen
  - PDCA
  - Reengineering
- 69 Which of the following methods involve generating statistical evidence?
- Formal optimization and QFD generally create statistical evidence.
  - Acceptance sampling, regression, and SPC create DOE evidence.
  - Process mapping and QFD generally create statistical evidence.
  - Both a) and b)
  - Both a) and c)

- 70 A company is trying to design a new product and wants to systematically study its competitor's products. Which methods are obviously helpful (*i.e.*, the method description mentions related goals)?
- Gauge R&R
  - QFD
  - Formal Optimization
  - Both a) and b)
  - Both a) and c)
- 71 A company has implemented a new design into production. Now it is interested in prioritizing which inspection areas need more attention and in documenting a complete safety system. Which methods are obviously helpful (*i.e.*, the method description mentions related goals)?
- FMEA
  - QFD
  - Control planning
  - Both a) and b)
  - Both a) and c)
- 72 Which methods are obviously helpful for evaluating measurement systems (*i.e.*, the method description mentions related goals)?
- Gauge R&R
  - DOE
  - Formal Optimization
  - Both a) and b)
  - Both a) and c)
- 73 A company is trying to design a new product and wants to study input combinations to develop input-output predictive relationships. Which methods are obviously helpful (*i.e.*, the method description mentions related goals)?
- Regression
  - DOE
  - Control planning
  - Both a) and b)
  - Both a) and c)
- 74 A team is in a problem-solving phase in which the objectives and responsibilities have been established but the state of the current system
- has not been measured which method(s) would be obviously helpful
- SPC charting
  - Gauge R&R
  - DOE
  - Both a) and b)
  - Both a) and c)
- 75 The operators of a manufacturing cell work out a more orderly arrangement for tool storage and establish a schedule to maintain cleanliness on a daily basis. These improvements are best described by which approach to problem solving?
- Pa-yé
  - Kaizen
  - PDCA
  - Reengineering
- 76 A quality engineer employed by a hospital is asked to improve the process of medication storage in locked cabinets near patient doors. One defect that occurs rarely is that the medication caddy is left out when the cabinet is relocked. The engineer installs a gravity-activated arm that will not permit the door to close when the caddy isn't inside. This improvement is best described by which approach to problem solving?
- 5S
  - Pa-yé
  - Kaizen
  - PDCA
  - Reengineering
- 77 A team is investigating ways to reduce power outages. They determine that an outage can occur in only three ways: grid failure, local transformer failure, or local overload. They then investigate each of these three events for possible causes. They draw a diagram that "fans out," using the power outage as the handle of the fan. These improvements are best described by which approach to problem solving?
- Affinity diagram
  - Interrelationship digraph
  - Tree diagram

- d) Process decision program chart  
 e) Matrix diagram  
 f) Prioritization matrix  
 g) Activity network diagram
78. A team's goal is to improve information flow in a payroll function. They make 33 Post-It notes, each listing an issue for further investigation. After some discussion, they group them into four categories: mandated record keeping, privacy concerns, insurance concerns, and transfer concerns. This grouping process is best described by which approach to problem solving?  
 a) Affinity diagram  
 b) Interrelationship digraph  
 c) Tree diagram  
 d) Process decision program chart  
 e) Matrix diagram  
 f) Prioritization matrix  
 g) Activity network diagram
79. The team in the above problem draws arrows from Post-It notes that are causes to notes that are the effects of these causes. This step is best described by which approach to problem solving?  
 a) Affinity diagram  
 b) Interrelationship digraph  
 c) Tree diagram  
 d) Process decision program chart  
 e) Matrix diagram  
 f) Prioritization matrix  
 g) Activity network diagram
80. A team working with plant relocation is tasked with designing a process for moving 180 pieces of equipment. Incoming orders may need to be filled during the move at either the old site or the new one. Transportation equipment availability is uncertain. Construction schedules at the new site are very weather-dependent. The team designs a chart that attempts to cover these and other contingencies with appropriate measures for each. The tool best fitted for this task is:  
 a) Affinity diagram  
 b) Interrelationship digraph
- c) Tree diagram  
 d) Process decision program chart  
 e) Matrix diagram  
 f) Prioritization matrix  
 g) Activity network diagram
81. Which problem-solving technique is derived from PERT (program evaluation and review technique) and CPM (critical path method) for project management?  
 a) Affinity diagram  
 b) Interrelationship digraph  
 c) Tree diagram  
 d) Process decision program chart  
 e) Matrix diagram  
 f) Prioritization matrix  
 g) Activity network diagram
82. Which department(s) could possibly use SPC charting?  
 a) Production  
 b) Marketing  
 c) Sales and logistics, for monitoring delivery times of truckers  
 d) All of the above are correct.
83. Which of the following is true about engineering specification limits?  
 a) They are associated with the " $\pm$ " given on blueprints.  
 b) They can fail to reflect actual performance in that nonconforming defective.  
 c) They are always written in dimensionless units.  
 d) Both a) and b)  
 e) Both a) and c)
84. Which of the following is correct about engineering specifications?  
 a) They are sometimes made up by engineers who do not know the implications.  
 b) They are often used in contracts between procurement and suppliers.  
 c) They could be so wide as to raise no production concerns.  
 d) All of the above are correct.  
 e) Both a) and c)

- 55** A potential scope for the sales subsystem for a lemonade stand is
- Improve the taste of a different type of lemonade by adjusting the recipe.
  - Increase profit through reducing raw optimizing over the price.
  - Reduce "cycle time" between purchase of materials and final product delivery.
  - All of the above fit the definition as used in the text.
  - Both b) and c) are scope objectives.
- 56** Which constitute relevant tangible deliverables from a taste improvement project?
- A gallon of better-tasting lemonade
  - Documentation giving the improved recipe
  - An equation predicting the taste rating as a function of ingredients
  - All of the above are tangible deliverables.
  - Both a) and b) are tangible deliverables.
- 57** Which of the following are possible deliverables from a wood process project?
- Ten finished chairs
  - Posters comparing relevant competitor chairs
  - Settings that minimize the amount of wasted wood
  - All of the above are tangible deliverables.
  - Both a) and c) are tangible deliverables.
- 58** Why might using rework and scrap costs to evaluate the cost of nonconformities be inaccurate?
- Rework generally does not require expense
  - If there are many nonconforming units, some inevitably reach customers
  - Production defects increase lead times, resulting in lost sales.

- All of the above are possible reasons.
  - Both b) and c) are correct.
- 89.** Why are go-no-go decisions utilized?
- Eliminating design concepts early in a design process can save tooling costs.
  - More design concepts exist than can be investigated, due to budget limitations.
  - Decisive choices can be made, potentially related to multiple product lines.
  - All of the above are possible uses.
  - Both b) and c) are possible uses.
- 90.** An engineer might use a Pareto chart to uncover what type of information?
- Prioritization of nonconformity types identify the relevant subsystem.
  - Pareto charts generally highlight the most recent problems discovered on the line.
  - Pareto charting does not involve attribute data.
  - All of the above are correct.
  - Both b) and c) result from a Pareto chart.
- 91.** What properties are shared between reproducibility and repeatability errors?
- Both derive from mistakes made by people and/or equipment.
  - Neither is easily related to true errors in relation to standards.
  - Estimation of both can be made with a crossed gauge R&R.
  - All of the above are correct.
  - Both a) and b)
- 92.** Which are differences between reproducibility and systematic errors?
- Systematic errors are between a generic process and the standard values; reproducibility errors are differences between each appraiser and the average appraiser.
  - Evaluating reproducibility errors relies more on standard values

- c) Systematic errors are more easily measured without standard values.
- d) All of the above are differences between reproducibility and systematic errors.
- e) Both b) and c) are true differences
93. Which of the following is NOT a benefit of SPC charting?
- Charting helps in thorough evaluation of system quality
  - It helps identify unusual problems that might be fixable.
  - It encourages people to make continual adjustments to processes.
  - It encourages a principled approach to process meddling (only after evidence)
  - Without complete inspection, charting still gives a feel for what is happening.
94. Control chart types are: (Subject GRE, 2010)
- Two
  - Three
  - Four
  - Five
  - None of the above.
95. Which of the following describes the relationship between common cause variation and local authorities?
- Local authority generally cannot reduce common cause variation on their own.
  - Local authority has the power to reduce only common cause variation.
  - Local authority shows over-control when trying to fix assignable causes.
  - All of the above are correct.
  - All of the above are correct except a) and d).
96. Which of the following is correct and most complete?
- False alarms are caused by assignable causes.
- b) The charts often alert local authority to assignable causes which they fix.
- c) Charts seek to judge the magnitude of average assignable cause variation.
- d) All of the above are correct.
- e) All of the above are correct except a) and d).
97. Nonconforming control charts are based on (Subject GRE, 2010)
- Binomial distribution
  - Hypergeometric Distribution
  - Poisson distribution
  - Negative distribution
  - None of the above
98. Which of the following is (are) true of *u*-chart process capability?
- It is the usual average number of nonconformities per item
  - It is the fraction of nonconforming units under usual conditions
  - It necessarily tells less about a system than *p*-chart process capability.
  - All of the above are true.
  - All of the above are correct except a) and d)
99. Which of the following relate run rules to six sigma goals?
- Run rules generate additional false alarms improving chart credibility
  - Sometimes  $LCL = 0$  so run rules offer a way to check project success.
  - Run rules signal assignable causes in start-up. improving capability measures.
  - All of the above are correct
  - All of the above are correct except a) and d)
100. What is the upper control limit for a *p* chart (proportion defective) when the average daily production is 2500 units with an established fraction defective of 0.05?
- 0.054
  - 0.058
  - 0.063
  - 0.066

101 Which is not a characteristic of a p-chart?

- a) It shows the number of defects per item being inspected.
- b) It measures the fraction of non-conforming items in a sample.
- c) It is based on the binomial distribution (or its normal approximation).
- d) It will have varying control limits if the sample size is changing.

102 An  $\bar{x}$  and  $R$  chart was prepared for an operation using 20 samples with five pieces in each sample.  $\bar{x}$  was found to be 33.6 and  $\bar{R}$  was 6.20. During production, a sample of five was taken and the pieces measured 36, 43, 37, 25, and 38. At the time this sample was taken:

- a) Both the average and range were within control limits.
- b) Neither the average nor range were within control limits.
- c) Only the average was outside control limits.
- d) Only the range was outside control limits.

103 In a normal distribution, what is the area under the curve between  $+0.7$  and  $+1.3$  standard deviation units?

- a) 0.2903
- b) 0.7580
- c) 0.2580
- d) 0.1452

104 Given process average = 1.64, average range = .05,  $n = 5$ , assuming statistical control and a normal population what proportion of the population will meet specifications of  $1.65 \pm .05$ ?

- a) 80%
- b) 86%
- c) 97%
- d) 93%

105 The control chart that is most sensitive to variations in measurements is

- a) p chart
- b) np chart
- c) c chart
- d)  $\bar{x}$  and R chart

106 Which of the following statements are not true?

- a) Raising quality levels can lead to decreased costs, a greater degree of consumer satisfaction, and thus increased profitability.
- b) Control Charting is now used extensively in industry as a diagnostic technique for monitoring production processes to identify instability and unusual circumstances.
- c) The basis for most of control charts lies in our work concerning probability distributions of various statistics such as the sample mean  $x$  and sample proportion  $p = X/n$ .
- d) All of the above statements are true.
- e) None of the above statements are true.

107 Which of the following statements are not true?

- a) Control charts and acceptance sampling plans were first developed in the 1960's and 1970's.
- b) Statisticians and engineers have recently introduced many statistical methods for identifying types and levels of production inputs that will ensure high-quality output.
- c) There is a large body of material known as "Taguchi methods" named after Japanese engineer/statistician G Taguchi.
- d) All of the above statements are true.
- e) None of the above statements are true.

108 Which of the following are not examples of assignable causes of variation?

- a) Contaminated material
- b) Incorrect machine settings
- c) Environmental factors
- d) Unusual machine tools wear
- e) All of the above

109 Which of the following statements are true?

- a) We might think of "natural random variation" as uncontrollable background noise.
- b) Control charts provide a mechanism for recognizing situations where assignable causes may be adversely affecting product quality.
- c) Once a control chart indicates an out-of-control situation, an investigation can be launched to identify causes and take corrective action.
- d) A basic element of control charting is that samples have been selected from the process of interest at a sequence of time points.
- e) All of the above statements are true.

110 Which of the following statements are not true?

- a) The basis for the choice of a center line for a control chart is sometimes a target value or design specification, for example a desired value of the bearing diameter.
- b) An "in-control" process is a process that "meets design specifications or tolerance".
- c) An in-control process is simply one whose behavior with respect to variation is stable over time, showing indications of unusual extraneous causes.
- d) If the points on a control chart all lie between the two control limits, the process is deemed to be in control.
- e) All of the above statements are true.

111 Which of the following statements are not true?

- a) An in-control process is simply one that is operating in a stable fashion, reflecting only natural random variation.

- b) An out-of-control "signal" occurs whenever a plotted point falls outside the two control limits.
- c) There is a strong analogy between the logic of control charting and hypothesis testing. The null hypothesis here is that the process is out-of-control.
- d) The two control limits are designed so that an in-control process generates very few false alarms whereas a process not in control quickly gives rise to a point outside the limits.
- e) All of the above statements are true.

112 Which of the following statements are not true?

- a) To construct an  $X$  chart, there are two different commonly used methods for estimating the unknown process standard deviation  $\sigma$ ; one based on the  $k$  sample standard deviations and the other on the  $k$  sample ranges.
- b) In the case of a normal population distribution, the estimator of the unknown process standard deviation  $\sigma$  based on sample standard deviations  $S$  is more efficient than that based on the sample range.
- c) The sample standard deviation  $S$  is an unbiased estimator of the population standard deviation  $\sigma$ , that is,  $E(S) = \sigma$ .
- d) All of the above statements are true.
- e) None of the above statements are true.

113 Which of the following statements are not true?

- a) One important use of control charts is to see whether some measure of location of the variable's distribution remains stable over time.
- b) It is highly unlikely that for an in-control process, the sample mean will fall within 3 standard

deviations  $\left(\frac{3\sigma}{\sqrt{n}}\right)$  of the process mean  $\mu$ .

- c) The use of control charts based on 3 standard deviation limits is traditional, but tradition is certainly not inviolable.
- d) All of the above statements are true.
- e) None of the above statements are true.

114. Assume that for an in-control process, the random variable of interest  $X$  has a normal distribution with mean value  $\sigma$  and standard deviation  $\sigma$ . If  $\bar{X}$  denotes the sample mean for a random sample of size  $n$  selected at a particular time, then

- a)  $E(\bar{X}) = \mu$
- b)  $\sigma_{\bar{X}} = \left(\frac{\sigma}{\sqrt{n}}\right)$
- c)  $X$  has a normal distribution.
- d) All of the above are true.
- e) None of the above are true.

115. Suppose that at each of the time points 1, 2, 3, ..., a sample of size  $n$  is selected at random from a normal distribution with known mean  $\mu$  and standard deviation  $\sigma$ . In order to construct a 3-sigma  $X$  chart, we need to

- a)  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots$ ; calculated values of the corresponding sample means.
- b) Plot  $\bar{x}_i$ 's over time; that is, plot the points  $(1, \bar{x}_1), (1, \bar{x}_2), (1, \bar{x}_3), \dots$  and so on.
- c) Draw horizontal lines across the plot as  $\mu \pm \frac{3\sigma}{\sqrt{n}}$
- d) All of the above are needed.
- e) Only A and B are needed.

116. Which of the following statements are true?

- a) Generally speaking, a control chart will be effective if it gives very few out-of-control signals when the process is in control, but shows a point outside the control limits almost as soon as the process goes out of control.
- b) One assessment of a control chart's effectiveness is based on the notion of "error probabilities"  $\alpha$  and  $\beta$ .
- c) The use of a 3-sigma limits for an  $X$  chart makes it highly unlikely that an out-of-control signal will result from an in-control process.
- d) One assessment of a control chart's effectiveness involves the average run length (ARL) needed to observe an out-of-control signal.
- e) All of the above statements are true.

117. The inability of  $X$  charts with 3-sigma limits to quickly detect small shifts in the process mean has prompted investigators to develop procedures that provide improved behavior in this respect. Which of the following conditions need to be satisfied for an appropriate intervention to take corrective action?

- a) Two out of three successive points fall outside 2-sigma limits on the same side of the center line.
- b) Four out of five successive points fall outside 1-sigma limits on the same side of the center line.
- c) Eight successive points fall on the same side of the center line.
- d) All of the above.
- e) None of the above.

118. Which of the following statements are true?

- a) It is important to ensure that a process is under control with respect to location (equivalently central tendency) as well as variation.

- b) Most practitioners recommend that control of a process be established on variation prior to constructing an X chart or any other chart for controlling location.
- c) Charts for variation are based on the sample standard deviation S and also based on the sample range R.
- d) Charts for variation that are based on the sample standard deviation S are generally preferred over charts that are based on the sample range R because the standard deviation gives a more efficient assessment of variation than does the range.
- e) All of the above statements are true

119 Suppose that  $k$  independently selected samples are available, each one consisting of  $n$  observations on a normally distributed variable. Denote the sample standard deviations by  $s_1, s_2, s_3, \dots, s_k$ . The values  $s_1, s_2, s_3, \dots, s_k$  are plotted in sequence on an S chart. The center line of the chart will be at height equals to the

a) Average of the values  $s_1, s_2, s_3, \dots, s_k$

b) Range of the values  $s_1, s_2, s_3, \dots, s_k$

c) Standard deviation of the values  $s_1, s_2, s_3, \dots, s_k$

d) First quartile of the values  $s_1, s_2, s_3, \dots, s_k$

e) third quartile of the values  $s_1, s_2, s_3, \dots, s_k$

120 Suppose there are 10 sample obtained at equally spaced time points and  $n=4$  observations in each sample. If the sum of the 20 sample standard deviations is 42, and that the tabulated value of  $\chi^2$  is 921, then

- UCL for an S control chart are respectively
- a) 1.062 and 3.138
- b) 3.138 and 4.765
- c) 0.0 and 4.765
- d) -3.54 and 4.554
- e) -5.65 and 4.765

- 121 Suppose there are 28 samples obtained at equally spaced time points, and  $n=7$  observations in each sample. If the sum of the 28 sample ranges is 98 and that the tabulated values of  $b_3$  and  $c_3$  are 2.706 and 8.33, respectively, then the 3-sigma control limits LCL and UCL for an R control chart are respectively
- a) 0 and 6.73
- b) 27 and 6.73
- c) -30.61 and 37.61
- d) 0.0 and 37.61
- e) None of the above answers are correct

- 122 Which of the following statements are not correct?
- a) For a 3-sigma X chart, where the process mean  $\mu$  and standard deviation  $\sigma$  are known, the probability that a point on the chart falls above the UCL is .013, as is the probability that the point falls below the LCL.
- b) For a 3-sigma S chart,  $P(S > UCL) = P(S < LCL) = 0.0013$
- c) For a 3-sigma R chart,  $P(R > UCL) = P(R < LCL) = 0.0013$
- d) Only B and C are not correct
- e) Only A and C are not correct

- 123 Consider a sample of  $n$  items obtained at a particular time, and let  $X$  be the number of defectives and  $\hat{p} = \frac{X}{n}$ . Which of the following statements are not true?
- a)  $E(X) = np$
- b)  $V(X) = p(1-p)$
- c)  $E(\hat{p}) = p$

$$d) v(\hat{p}) = \frac{p(1-p)}{n}$$

- e) If  $np \geq 10$  and  $n(1-p) \geq 10$ ,  $\hat{p}$  has approximately a normal distribution.

124 Suppose that 25 samples, each of size 200, were selected from what is believed to be an in-control process, and that  $\sum \hat{p}_i$  is the fraction of defective items in sample. The 3-sigma control limits, LCL and UCL, of the p chart for the fraction of defective items are respectively

- a) 0.236 and .244
- b) .0462 and .0942
- c) 0 and .0564
- d) 0 and .0942
- e) -.0084 and .0564

125 Assume that the total number of defects in 50 samples are 450. Which of the following statements are not true regarding the c chart for the number of defectives?

- a) The center line of the chart is at height 9
- b) The lower control limit (LCL) is 0
- c) The upper control limit (UCL) is 18
- d) All of the above statements are true
- e) None of the above statements are true

126 Which of the following statements are true?

- a) The CUSUM procedures are used for controlling process location.
- b) There are CUSUM procedures for controlling process variation.
- c) There are CUSUM procedures for attribute data.
- d) None of the above statements are true
- e) All of the above statements are true

127 Which of the following statements are true?

- a) A defect on the traditional  $\bar{X}$  chart is its inability to detect a relatively small change in a process mean
- b) Whether a process is judged out of control at a particular time depends only on the sample at that time, and not on the past history of the process
- c) The computational version of a cumulative sum (CUSUM) procedure is used almost exclusively in practice, but the logic behind the procedure is most easily grasped by first considering the graphical form.
- d) All of the above statements are true
- e) None of the above statements are true.

128 Which of the following statements are not true?

- a) A particular V-mask is determined by specifying the "lead distance"  $d$  and "half angle"  $\theta$  or equivalently, by specifying  $d$  and the length  $h$  of the vertical line segment from 0 to the lower (or to the upper) arm of the mask.
- b) One method for deciding which V-mask to use involves specifying the size of a shift in the process mean that is of particular concern to an investigator, then the parameters of the mask are chosen to give desired values of  $\alpha$  and  $\beta$ ; the false-alarm probability and the probability of not detecting the specified shift, respectively.
- c) One method for deciding which V-mask to use involves selecting the mask that yields specified values of the ARL (average run length) both for an in-control process and for a process in which the mean has shifted by a designated amount.

- d) All of the above statements are true.
- e) None of the above statements are true.
- 129 Which of the following statements are not true?
- If we let  $\Delta$  denote the size of a shift in a process mean  $\mu$  that is to be quickly detected using a CUSUM procedure, then it is common practice to let  $k = 2\Delta$ , where  $k$  denotes the slope of the lower arm of the V-mask.
  - A quality control practitioner may specify a desired value of an ARL (average run length) when the process is in control  $\mu = \mu_0$ .
  - A quality control practitioner may specify a desire value of an ARL (average run length) when the process is out of control because the mean has shifted by  $\Delta$  ( $\mu = \mu_0 + \Delta$  or  $\mu = \mu_0 - \Delta$ )
  - All of the above statements are true.
  - None of the above statements are true.
30. If a process is out of control, the theoretical probability that a single point on the  $\bar{x}$  chart will fall between plus one sigma and the upper control limit is:
- 0.2240
  - 0.1587
  - Unknown
  - 0.3413
31. In order to be effective, the quality audit function ideally should be:
- An independent organizational segment in the quality control function.
  - An independent organizational segment in the production control function.
  - An independent organizational segment in the manufacturing operations function.
  - All of the above.

132. Ishikawa diagrams are
- Based on the Japanese character for quality.
  - Cause-and-effect diagrams.
  - Similar to Pareto charts.
  - An alternative to CPM for project management.
  - An alternative to flowcharts.
133. Which of the following quality indices is likely to have the greatest appeal to top management as an indicator of relative costs?
- Quality cost per unit of product.
  - Quality cost per hour of direct production labor.
  - Quality cost per unit of processing cost.
  - Quality cost per unit of sales.
  - Quality cost per dollar of direct production labor.
134. In spite of the quality engineer's best efforts, situations may develop in which his or her decision is overruled.
135. The most appropriate action would be to:
- Resign the position based on convictions.
  - Report findings to an outside source, such as a regulatory agency or the press.
  - Document findings, report to superiors, and move on to the next assignment.
  - Discuss findings with coworkers in order to gain support, thereby forcing action.
136. Review of purchase orders for quality requirements falls into which one of the following quality cost segments?
- Prevention
  - Appraisal
  - Internal failures
  - External failures
137. Failure costs include costs due to:
- Quality control engineering.
  - Inspection setup for tests.
  - Certification of special-process suppliers.
  - Supplier analysis of nonconforming hardware.

- 138 A quality system must include two main items—the preparation of documented quality system procedures and instructions, and:
- The effective implementation of the documented quality system procedures and instructions.
  - Management responsibility.
  - Purchasing department involvement.
  - Assessment of subcontractors.
- 139 Product verification is an important element of a quality system. Product verification:
- Is the procedure used to verify shipping routes and rates.
  - Is primarily a design function and is not used once the design has been finalized.
  - Consists of procedures that inform top management that numerical quotas are being met.
  - Uses tests and inspection in the manufacturing process to verify conformance.
  - Is a tool used mostly by third-party auditors.
- 140 Unless otherwise designated, changes to quality documents should be reviewed and approved by:
- Top management.
  - The same organization/function that produced the original document.
  - A team with representation from the quality function.
  - The chief quality officer.
  - All persons impacted by the changes.
- 141 The "layers" of a quality manual are:
- Policies, procedures, instructions, and records.
  - Executive, middle management, team leader, and team members.
  - Financial, product integrity, supplier relations, and process control.
  - Vision, goals, objectives, and projects.
- 142 In most systems of classification of quality characteristics:
- There are five categories.
  - The highest categories make reference to safety and health.
  - The lowest category refers to functional failure.
  - Fit and finish are not considered in any of the categories.
  - None of the above.
- 143 Strategic plans:
- Should be determined after tactical plans are established.
  - Tend to be more devious than operational plans.
  - Are longer terms than tactical plans.
  - Are less controversial than tactical plans.
  - Are used to implement goals established by the tactical plans.
- 144 A material traceability system provides information on:
- The flow of material through manufacturing processes.
  - The raw material used for a particular manufactured item.
  - The purchase order to invoice payment process.
  - Location of items material to pending legal action.
  - None of the above.
- 145 The principle purpose of a material review board (MRB) is to:
- Determine the vendor of purchased material.
  - Make decisions regarding nonconforming material.
  - Review traceability records for outgoing materials.
  - Establish policies and procedures for inspecting incoming material.
- 146 Product traceability is usually used to:
- Identify sources of the material used for the product.
  - Maintain a paper trail of financial transactions regarding the product.
  - Track the design changes affecting the product.

- d) Maintain payroll records, clock times, and so on, regarding the product.
- 147 A lot of size 2000 is to be inspected using ANSI/ASQ Z1.4-2003, with an AQL of 65%. Use a single, normal level II plan. The sample size is:
- 20
  - 40
  - 60
  - 80
  - none of the above
- 148 An attribute sampling plan lists the accept and reject values as  $Ac = 3$ ,  $Re = 6$ . Using the plan correctly, the inspector finds five defectives. The appropriate action is to:
- Reject the lot since the number of defectives exceeds the  $Ac$  value.
  - Accept the lot since the number of defectives is less than the  $Re$  value.
  - Draw another sample.
  - Use a different sampling plan.
  - None of the above
- 149 A lot is inspected using ANSI/ASQ Z1.4-2003 with an AQL of 65%. The lot passes the inspection. Does that guarantee that the lot has .65% or fewer defectives?
- Yes, because the sampling plans are statistically valid
  - Yes, assuming the sample was randomly selected
  - No, because Z1.4 isn't appropriate for use with defectives
  - No, because type I error could have occurred
  - No, because type II error could have occurred
- 150 A lot has 15% defective parts. The attribute sampling plan is based on an AQL of 10%. In this situation, the probability of rejection is .936. Find  $b$ .
- 0.936
  - 93.6%
  - 0.468
  - 0.064
  - none of the above
- 151 A lot of size 2000 is to be inspected using ANSI/ASQ Z1.4-2003, with an AQL of 15%. Use a single, normal level II plan. The sample size is:
- 20
  - 40
  - 60
  - 80
  - None of the above



- 152 The diagram shows operating characteristic (OC) curves for two sampling plans. The dashed curve is:
- Better
  - Worse
  - Better for the consumer
  - Better for the producer

- 153 If the lot is much worse than the AQL, then the sampling plan most likely to detect this at the lowest cost is:
- Single
  - Double

- 154 If the same characteristic is measured 10 times, the precision of the measurement system refers to the:
- Amount of variation between the 10 readings
  - Proximity of the average of the 10 readings to the true value of the characteristic
  - Smallest unit measurable
  - General quality of the equipment used

- 155 If the same characteristic is measured 10 times, the accuracy of the measurement system refers to the:
- Amount of variation between the 10 readings
  - Proximity of the average of the 10 readings to the true value of the characteristic

- c) Smallest unit measurable.  
 d) General quality of the equipment used.
- 156 FMEA is most beneficial, when used:  
 a) To analyze warranty data.  
 b) To analyze data from the manufacturing process.  
 c) During the design phase.  
 d) In documentation analysis.
- 157 A technique for translating customer demands into product characteristics is:  
 a) QFD.  
 b) SPC.  
 c) FMEA.  
 d) AQP.  
 e) DOE.
- 158 Derating as a design tool is the:  
 a) Reduction of the rated reliability of a product or component based on new data.  
 b) Assignment of a product to operate at stress levels below its normal rating.  
 c) Shortening of the warranty period in order to lower consumer expectations.  
 d) Removal of posted ratings from electronic components.
- 59 FMECA includes tools for:  
 a) Determining root causes of failure.  
 b) Assessing the probability that a product will function for a stated period of time.  
 c) Approximating the capability of a process to hold a certain dimension within tolerance.  
 d) Examining a proposed design for possible ways it can fail.
- 60 In calculating the RPN in an FMEA, the following three values are multiplied together:  
 a) Severity, opportunity, difficulty.  
 b) Severity, occurrence, design.  
 c) Severity, occurrence, detection.  
 d) Sensitivity, opportunity, difficulty.
- 161 Each failure mode and effects analysis (FMEA) report should contain:  
 a) A department to be assigned.  
 b) Recommended corrective actions.  
 c) Deadline dates.  
 d) Return on investment.  
 e) Associated project management procedures.
- 162 A quality team is given a process problem. The team responds by carefully selecting a solution from among several that are proposed. The team next installs the solution and collects the resulting data. Based on this data, the team modifies the solution slightly, then installs it as a permanent part of the process. The best description of the procedure the team has employed is:  
 a) kaizen  
 b) CI  
 c) PDSA  
 d) C.A.R.  
 e) FMEA
- 163 Value Stream Mapping can be viewed as an extension of which activity?  
 a) Gauge R&R  
 b) Benchmarking  
 c) Design of experiments  
 d) Process mapping  
 e) None of the above is correct.
- 164 What are possible benefits associated with U-shaped cells?  
 a) Parts are produced before demands are placed, for readiness.  
 b) Parts are produced in batches of one.  
 c) Personal accountability for product quality is returned to the worker.  
 d) All of the above are correct.  
 e) All of the above are correct except a) and d).
- 165 Which is correct and most complete?  
 a) Mixed production results in fewer setups than ordinary batch production.

- b) Kanban cards can limit the total amount of inventory in a plant at any time.
- c) U-shaped cells cause workers to perform only a single specialized task well.
- d) All of the above are correct.
- e) All of the above are correct except a) and d).
166. Which of the following is correct and most complete?
- a) Design of experiments does not require new testing.
- b) Screening can start where C&E matrices end and further shorten the KIV list.
- c) Strictly speaking, DOE is essential for proof in the statistical sense.
- d) All of the above are correct.
- e) All of the above are correct except a) and d).
167. Which of the following is correct and most complete?
- a) FMEA focuses on the manufacturing system with little regard to measurement.
- b) FMEA is based on quantitative data measured using physical equipment.
- c) FMEA helps to clarify the vulnerabilities of the current system.
- d) All of the above are correct.
- e) All of the above are correct except a) and d).
168. Which of the following is the most complete and correct?
- a) FMEA is primarily relevant for identifying nonvalue added operations.
- b) Both C&E and FMEA activities generate prioritized lists.
- c) Process mapping helps identify cause and effect relationships.
- d) All of the above are correct.
- e) All of the above are correct except a) and d).
169. The Deming (or Shewhart) cycle is a guide for:
- a) The steps followed in continuous quality improvement.
- b) Milestones in project or process management.
- c) The procedure used for setting up control charts.
- d) The Design of quality products.
- e) Quality team facilitation.
170. Kaizen activity usually:
- a) Only involves personnel from accounting functions.
- b) Results in a single change that has a large impact on improvement.
- c) Focuses on warranty and customer satisfaction analysis.
- d) Tends to produce many small incremental improvements.
171. An improvement technique known as the Shewhart cycle has the initials PDCA or PDSA. Which element of the cycle is being executed when a team tries out a proposed procedure for improving data integrity?
- a) P
- b) D
- c) C (or S)
- d) A
172. An improvement technique known as the Shewhart cycle has the initials PDCA or PDSA. Which element of the cycle is being executed when a team decides to make a change in standard operating procedure?
- a) P
- b) D
- c) C (or S)
- d) A
173. An affinity diagram provides a technique to:
- a) Determine the extent to which two components will operate synergistically.
- b) Evaluate the way different failure modes cascade into one another in a domino effect.

73. Determine the area under the normal curve as the number of standard deviations increases without bound.
74. A team wants to reduce the failure rate of a complex system. A useful tool for analyzing the problem would be a:
- Histogram.
  - Scatter diagram.
  - Gantt chart.
  - Tree diagram.
75. A team needs to decide at which points in a process to measure quality characteristics. The quality tool they would use is a:
- Control chart.
  - Check sheet.
  - Scatter diagram.
  - Flowchart.
  - Pareto chart.
76. A team wants to illustrate which defect types are occurring most frequently. The quality tool they would use is a:
- Control chart.
  - Check sheet.
  - Scatter diagram.
  - Flowchart.
  - Pareto chart.
77. A team wants to determine whether it might be useful to calculate the correlation coefficient between two variables. The quality tool they would use is a:
- Control chart.
  - Check sheet.
  - Scatter diagram.
  - Flowchart.
  - Capability analysis.
78. A team wants to determine the "shape" of the data to help determine whether it is bimodal as suspected. The quality tool they would use is a:
- d) Flowchart.  
e) Pareto chart.
179. When installing corrective action for a quality problem, it is important that
- Documents are updated.
  - All involved personnel are informed.
  - The process is monitored for recurrence of the problem.
  - All of the above.
  - None of the above.
180. When completing a corrective action cycle, the most frequently underemphasized step is:
- Recurrence control.
  - Effectiveness assessment.
  - Problem identification.
  - Correction.
181. The steps in a corrective action cycle, in alphabetical order, are: 1. Correction, 2. Effectiveness assessment, 3. Problem identification, 4. Recurrence control. List these in the order a team should execute them.
- 3, 1, 2, 4
  - 3, 4, 1, 2
  - 3, 4, 2, 1
  - 3, 2, 4, 1
182. In most lists of problem-solving steps, the first step is to:
- Install a quick fix.
  - Analyze the data.
  - Identify the problem.
  - Propose several alternative solutions.
183. Many organizations "successfully" solve problems only to have the same problems reoccur at a later date. This is due to failure of:
- Recurrence control.
  - Problem identification.
  - Cause-and-effect analysis.
  - Flowchart application.
  - Matrix diagram implementation.

- 184 Robust design refers to the ability of a product to:
- Fulfill its function for extended periods of time without failure
  - Fulfill its function despite changes in operating environment
  - Continue operating even though it has suffered damage
  - Operate well for various functions in addition to the one for which it was designed.
  - None of the above
- 185 "Pa-ye," or error-proofing, and "robust design" have the following in common:
- Both require statistical analysis of data
  - Both are preventive in nature
  - Both deal exclusively with products rather than processes
  - Neither are used in nonmanufacturing processes
- 186 Most barriers to quality improvement can be attributed to:
- Poor attitudes of hourly employees.
  - Unreasonable requirements by suppliers.
  - Inadequate statistical knowledge by quality professionals.
  - Insufficient commitment by management.
  - Lack of computer power.
- 187 Deming called the technique of studying a sample to gain understanding of the distribution of a population an "enumerative study." His main objection to these studies was that
- They are too difficult to perform correctly
  - They require extensive use of computers.
  - They assume a stable distribution.
  - Random samples are expensive to obtain
  - These studies have a high probability of type II error
- 188 An automatic gauging system is to be installed in a process. The gage will insert data values into a database from which machine adjustments will be made automatically. A critical factor in specifying the equipment is:
- The communication link between the gage and the computer
  - The compatibility of software in the gage and in the computer
  - Adequate manual overrides
  - All of the above.
- 189 The distribution on which *p* charts are based is called:
- Chi-square
  - Poisson
  - Bimodal
  - Binomial
  - Exponential
- 190 A capability study is done on the diameter of stainless steel rods turned on an automatic lathe. The most appropriate probability distribution to use is:
- Normal
  - Poisson
  - Hypergeometric
  - Binomial
  - Exponential
- 191 A process needs a constant pH of between 5 and 9 to operate efficiently. A control chart is used to monitor the acidity of a well-stirred vat of liquid. An *x*- and *R* chart is used with  $n = 4$ . The four readings in each sample are collected from the four compass points approximately 4" from the outside of the round vat. The *x* and *R* chart is not a good choice for this situation because it will probably show
- Very few points outside control limits even when the process is not stable.
  - A large number of points outside the control limits even when the process is stable.

- 1) The process is not normally distributed  
2) The average run length (ARL) is excessive

A stable, normally distributed process with specification  $3.50 \pm .03$  has  $m = 3.51$  and  $s = .016$ . What percent of the production violates specification?

- a) 16.43%
- b) 12.62%
- c) 18.58%
- d) 11.18%

(Refer to the previous problem.) Find  $C_{pk}$  and  $C_{pkL}$ .

- a) 1.21 and .85
- b) .85 and 1.21
- c) .35 and .63
- d) .63 and .42
- e) None of the above

Management of a trendy leather goods shop decides upon a 250% markup on handbags using no data and judgment only. This represents:

- a) Formal decision-making
- b) A House of Quality application
- c) Anecdotal information about the retail industry
- d) None of the above is correct

5 Which of the following is correct and most complete?

- a) Seat-of-the-pants decision-making is rarely (if ever) supported by anecdotal information.
- b) Inspecting a HOQ while making decisions is moderately formal.
- c) Performing DOE and using a solver to generate recommendations is informal.
- d) All of the above are correct.
- e) All of the above are correct except a) and d).

6 Formal optimization often requires:

- a) Subjectively factoring considerations not included in the formulation
- b) Clarifying as a group the assumptions and data for making decisions

- c) Plotting the objective function in the vicinity of the solutions for insight
- d) All of the above are correct.
- e) All of the above are correct except a) and d).

197 According to the six sigma literature a project for improving an existing system ends with which phase?

- a) Define
- b) Analyze
- c) Improve
- d) Verify
- e) Control

198 Which of the following is correct and most complete?

- a) Filling out each row of a control plan could require performing a gauge R&R
- b) According to the text, reaction plans are an optional stage in control planning.
- c) All characteristics on blueprints are critical characteristics
- d) All of the above are correct.
- e) All of the above are correct except a) and d).

199 The text implies that FMEAs and control plans are related in which way?

- a) FMEAs can help clarify whether characteristics should be declared critical.
- b) FMEAs determine the capability values to be included in the control plan.
- c) FMEAs determine the optimal reaction plans to be included in control plans.
- d) All of the above are correct.
- e) All of the above are correct except a) and d).

200 When considering sampling policies, the risks associated with accepting an undesirable lot grows with:

- a) Larger rational subgroup size
- b) Decreased tolerance of nonconformities in the rational subgroup (e.g., lower  $c$ )

- c) Increased tolerance of nonconformities in the overall lot (e.g., higher  $c$ )  
 d) Decreased overall lot size  
 e) None of the above
201. Which of the following is correct and most complete?  
 a) Correct document control requires the implementation of control plans.  
 b) Often, projects complete with revised SOPs are implemented corporation-wide.  
 c) Documenting findings can help capture all lessons learned in the project.  
 d) All of the above are correct.  
 e) All of the above are correct except a) and d).
202. Which of the following is correct and most complete?  
 a) EWMA control charts typically plot attribute data such as the number nonconforming.  
 b) Hotelling  $T^2$  charts are sometimes called multivariate control charts.  
 c) Multivariate control charts offer an alternative to several applications of Xbar & R charts.  
 d) All of the above are correct.  
 e) All of the above are correct except a) and d).
203. Which of the following is correct and most complete?  
 a) Multivariate charting typically involves the calculation of a large number of parameters during the startup phase.  
 b) Critical characteristics can vary together because they share a common cause.  
 c) EWMA charting often but not always discovers problems more quickly than Xbar & R charting.  
 d) All of the above are correct.  
 e) All of the above are correct except a) and d).
204. Which of the following is correct and most complete?  
 a) The  $\bar{R}$  in EWMA charting can be adjusted based on a desire to detect small shifts slowly or big shifts more quickly.  
 b) EWMA generally detects shifts faster than Xbar & R charting.  
 c) EWMA is particularly relevant when critical characteristics correlate.  
 d) All of the above are correct.  
 e) All of the above are correct except a) and d).
205. Which of the following is correct and most complete?  
 a) Multiple Xbar & R charts do not help in assignable cause identification.  
 b) Specific assignable causes might be associated with large values of certain linear combinations of quality characteristic values.  
 c) Multivariate charting could be applied to college admissions.  
 d) All of the above are correct.  
 e) All of the above are correct except a) and d).
206. Suppose that top face voids were much more expensive to fix than other voids. Which charting method would be most appropriate?  
 a)  $p$ -charting  
 b)  $u$ -charting  
 c) Xbar & R charting  
 d) Demerit charting  
 e) EWMA charting
207. Which of the following is the most correct and complete?  
 a) According to the definition of six sigma, monetary benefits must be measured.  
 b) Charting is often helpful for measuring benefits and translating to dollars.

- c) Gauge R&R might or might not be needed in a control plan.  
 d) All of the above are correct.  
 e) All of the above are correct except a) and d).

8 While evidence showed that the project resulting system inputs helped save money, which of the following is the safest criticism of the approach used?

- a) The team could have applied design of experiments methods.  
 b) A cause & effect matrix could have clarified what was important to customers.  
 c) Having a charter approved by management could have shorted the DOE time.  
 d) Computer assisted optimization would improve decision-making in this case.

9 Which charting method is most relevant for measuring performance?

- a) X bar & R charting is the most relevant since there is a high quality level.  
 b) P-charting is the most important since we are given the fraction of nonconforming units only.  
 c) U-charting is the most important since we are given count of nonconformities and no weighting data.

10 Which likely explains why formal gauge R&R was not used?

- a) The count of voids is attribute data, and it was not clear whether standard methods were applicable.  
 b) The engineers were not aware of comparison with standards methods since it is a relatively obscure method.  
 c) The project was not important enough for formal gauge R&R to be used.  
 d) Answers in parts a) and b) are both reasonable explanations.

- 211 Which is the safest critique of methods used in the wire harness study?  
 a) A cost Pareto chart would have been better since cost reduction was the goal.  
 b) QFD probably would have been more effective than DOE.  
 c) Charting of void count should not have been dropped since it always helps.  
 d) An FMEA could have called attention to certain void locations being missed.  
 e) Pareto charting must always be applied in the define phase.

#### ANSWERS

1	d	2	c	3	b	4	c
5	e	6	a	7	d	8	a
9	d	10	d	11	d	12	d
13	c	14	b	15	b	16	b
17	b	18	b	19	a	20	b
21	d	22	b	23	c	24	e
25	b	26	a	27	b	28	b
29	b	30	c	31	a	32	d
33	e	34	b	35	d	36	e
37	a	38	c	39	a	40	a
41	c	42	c	43	d	44	b
45	c	46	d	47	d	48	d
49	d	50	b	51	b	52	b
53	e	54	a	55	e	56	b
57	c	58	d	59	c	60	e
61	d	62	d	63	d	64	c
65	c	66	c	67	c	68	c
69	a	70	b	71	e	72	a
73	d	74	a	75	a	76	b
77	c	78	a	79	b	80	d
81	g	82	d	83	a	84	d
85	b	86	a	87	d	88	e
89	d	90	a	91	d	92	a
93	c	94	a	95	a	96	b
97	a	98	a	99	e	100	c
101	a	102	d	103	d	104	c
105	d	106	d	107	a	108	c
109	e	110	b	111	c	112	c
113	b	114	d	115	d	116	e
117	d	118	e	119	a	120	c
121	b	122	d	123	b	124	c
125	d	126	e	127	d	128	n

## 34 Compendium of Statistics

133	d	134	c	135	a	136	d	173	d	174	d	175	b	176	c
137	a	138	d	139	b	140	a	177	b	178	d	179	a	180	a
141	b	142	c	143	b	144	b	181	c	182	a	183	b	184	b
145	a	146	e	147	c	148	e	185	d	186	c	187	d	188	d
149	d	150	d	151	d	152	b	189	a	190	b	191	d	192	d
153	a	154	b	155	c	156	a	193	c	194	b	195	d	196	e
157	b	158	d	159	c	160	b	197	a	198	a	199	c	200	e
161	c	162	d	163	e	164	b	201	e	202	d	203	a	204	d
165	e	166	c	167	e	168	a	205	d	206	d	207	c	208	c
169	d	170	b	171	d	172	d	209	d	210	d	211	b		

## **INDEX NUMBERS**

1. Index number reveal the state of:
  - a) Inflation
  - b) Deflation
  - c) Both a) and b)
  - d) Either a) or b)
2. Index numbers are also known as;
  - a) Economics barometers
  - b) Sign and guide posts
  - c) Both a) and b)
  - d) Either a) or b)
3. Index number is a:
  - a) Measure of relative change
  - b) A special type of an average
  - c) A percentage relative
  - d) All of above
4. Index numbers are expressed:
  - a) In percentages
  - b) In ratios
  - c) In term of absolute value
  - d) All the above
5. Index number help:
  - a) In framing of economic policies
  - b) In assessing the purchasing power of money
  - c) For adjusting national economics
  - d) All the above
6. The error(s) involved in the construction of index numbers is/are:
  - a) Error of sampling
  - b) Formula error
  - c) Error in collected data
  - d) All the above
7. Element of subjectivity is involved in the index number due to:
  - a) Choice of base period
  - b) Selection of weight
  - c) Choice of commodities
  - d) All of the above
8. Most commonly used index number is:
  - a) Diffusion index number
  - b) Price index number
  - c) Value index number
  - d) None of the above
9. One of the limitation in the construction of index numbers is:
  - a) The choice of the type of the average
  - b) Choice of investigator
  - c) Choice of variable to be studied
  - d) All of above
10. Consumer price index number is constructed for:
  - a) A well defined section of people
  - b) All people
  - c) Factory workers only
  - d) All the above
11. Diffusion index number reveals the changes in:
  - a) Elite
  - b) Industrial production
  - c) A group of time series
  - d) None of above
12. the first and fore most step in the construction of index numbers is:
  - a) Choice of base period
  - b) Choice of weight
  - c) To define the purpose of index numbers
  - d) All of above
13. Base period of an index number should be:
  - a) A year only
  - b) A normal period
  - c) A period at distant past
  - d) None of above

14. Data of index number should be collected from:  
 a) The retailers  
 b) The wholesale dealers  
 c) The selected group of people  
 d) None of above
15. Most preferred type of average for index number is:  
 a) Arithmetic mean  
 b) Geometric mean  
 c) Harmonic mean  
 d) None of these
16. Most frequently used index number formulae are:  
 a) Weighted formulae  
 b) Unweighted formulae  
 c) Fixed weighted formulae  
 d) None of the above
17. The unweighted price index formula based on n items is:  
 a)  $\sum_{i=1}^n \frac{P_{it}}{P_{0i}}$   
 b)  $\sum_{i=1}^n \frac{P_{it}}{P_{0i}} \times 100$   
 c)  $\frac{\sum_{i=1}^n P_{it}}{\sum_{i=1}^n P_{0i}} \times 100$   
 d) None of these
18. Unweighted price index formula is:  
 a) Most frequently used  
 b) Seldom used  
 c) The best  
 d) All of above
19. Laspeyre's index formula uses the weights of the:  
 a) Base year  
 b) Current year  
 c) Both of above  
 d) None of above
20. The discrepancy  $(P_{01} \times P_{10} - 1)$  is termed as:  
 a) Joint error  
 b) Homogeneity error
- c) Formula error  
 d) None of the above
21. Factor reversal test was invented by:  
 a) Walsh  
 b) A.L. Dowley  
 c) John I. Griffin  
 d) Irving Fisher
22. The condition for the factor reversal test to be satisfied with usual notation is:  
 a)  $P_{01} \times Q_{01} = V_{01}$  b)  $\frac{P_{01} \times Q_{01}}{V_{01}} = 1$   
 c)  $\frac{P_{01} \times Q_{01}}{V_{01}} - 1 = 0$  d) All the above
23. The condition for the price indices to satisfy the circular test for four year data is:  
 a)  $P_{01} P_{12} P_{23} P_{30} = 1$   
 b)  $P_{01} P_{12} P_{23} P_{34} = 1$   
 c)  $P_{01} + P_{12} + P_{23} = P_{30}$   
 d)  $P_{12} + P_{23} + P_{34} = 1$
24. Circular test for prices indices is satisfied by the formula:  
 a) Based on geometric mean of price relatives  
 b) Obtained by Kelly's fixed weight method  
 c) Both (a) and (b)  
 d) Neither (a) nor (b)
25. Fisher's ideal formula does not satisfy:  
 a) Time reversal test  
 b) Circular test  
 c) Factor reversal test  
 d) Unit test
26. Year-to-year indices in the chain-base method are called  
 a) Chain indices  
 b) Link relatives  
 c) Fixed base indices  
 d) All the above

7. The relation between fixed base indices  $P_{01}, P_{02}, P_{03}$  and the chain base indices  $P_{01}, P_{12}, P_{23}$  is:
- $P_{03} = P_{01}P_{12}P_{23}$
  - $P_{03} = P_{12}P_{23}$
  - $P_{01}P_{12} = P_{02}P_{23}$
  - All the above
8. Indices calculated by the chain-base method are almost free from:
- Homogeneity error
  - Seasonal variations
  - Rigidity of weights
  - All the above
9. When the indices given for a number of years are to be worked out to a new base period, this phenomenon is known as:
- Splicing
  - Base shifting
  - Both a) and b)
  - Neither a) nor b)
10. Combining of two index number series having different base periods into one series with common base period is known as:
- Splicing
  - Base shifting
  - Both (a) and (b)
  - Neither (a) nor (b)
11. If the old is connected, with the new series of index numbers, it is known as:
- Base shifting
  - Backward splicing
  - Forward splicing
  - None of the above
12. If the new series is connected, with the old series, it is known as:
- Base shifting
  - Backward splicing
  - Forward splicing
  - None of the above
13. Factor reversal test permits the interchange of:
- Base periods
- b) Price and quantity  
 c) Weights  
 d) None of the above
34. The consumer price index in 1990 increases by 80% as compared to the base 1980. A person in 1980 getting Rs. 60,000 per annum should now get:
- Rs 1,08000 per annum
  - Rs.72,000 per annum
  - Rs. 54,000 per annum
  - None of the above
35. If the family spends on food, housing and clothing in the ratio of 5:3:2 and experiences the rise in prices of these heads by 40,30, and 20 percent respectively, the family budget will be increased by:
- 33%
  - 30%
  - 27%
  - None of the above
36. If the index number for 1990 to the base 1980 is 250, the index number for 1980 to the base 1990 is:
- 4
  - 40
  - 400
  - None of the above
37. If Laspeyres's price index is 324 and Paasche's price index 144, then Fisher's ideal index is:
- 234
  - 180
  - 216
  - None of the above
38. If the consumer price index for 1994 is 800, then the purchasing power of a rupee is:
- 0.125 paise
  - 12.5 paise
  - 8 paise
  - None of the above
39. If the consumer price index numbers for 1981 and 1982 to the base 1972 are 320 and 400 respectively. The consumer price index for 1981 to the base 1982 is:
- 125

- b) 80
- c) 128
- d) None of the above

40. The index number for 1985 to the base 1980 is 125 and for 1980 is 80. The given index satisfy:

- a) Time reversal test
- b) Circular test
- c) Factor reversal test
- d) All of the above

41. The price relatives for the three commodities are 125, 120, and 130 with their respective weights 5, w, and 8. If the price index for the set is 1125.25, the value of w is:

- a) 6
- b) -7
- c) 7
- d) None of the above

42. If the group indices are 80, 120, and 125 and their respective group weights are 60, 20, and 20, the consumer price index is:

- a) 108.33
- b) 97.00
- c) 98.49
- d) None of the above

43. Which index satisfies factor reversal test?

- a) Paasche's index
- b) Laspeyere's index
- c) Fisher's ideal index
- d) Walsh price index

44. If the sub index is zero, then the total index must:

- a) Be zero
- b) Be infinity
- c) Indeterminate
- d) Not be zero

45. If the unit of measurement of a commodity changes, the value of index number:

- a) Also changes
- b) Remains same
- c) Both a) and b)
- d) None of these

46. If all the sub indices are equal to p, then the total indices will be:

- a) Equal to P
- b) Equal to 100
- c) Equal to 1
- d) Equal to 0

47. The property that on changing the unit of measure of a commodity, the index number does not change, is called

- a) Test of equality
- b) Test of homogeneity
- c) Test of commensurability
- d) Test of proportionality

48. The property that a subindex is zero or infinity but the total index is not zero or infinity is known as:

- a) Test of randomness
- b) mness
- c) Factor reversal test
- d) Test of definiteness
- e) None of these

49. The property that in a case of all sub-indices being P resulting into total index equal to P is termed as:

- a) Test of randomness
- b) Test of proportionality
- c) Test of definiteness
- d) test of commensurability

50. Index numbers are special type of

- a) Averages
- b) Percentage relatives
- c) Ratios
- d) All of above

51. The index that indicates the progress or recession of an economic cycle in a group of time series

- a) Value index
- b) Quantitative index
- c) Diffusion index
- d) All of above

52. Index numbers generally have types

- a) One
- b) Three
- c) Two
- d) None of these

53. Which of the following is the step is necessary for the construction of index number?  
 a) Purpose  
 b) Selection of commodities  
 c) Selection of base year  
 d) All of the above  
 e) None of these
54. Index numbers are computed for:  
 a) Price of commodities  
 b) Quantity of commodities  
 c) Both a) and b)  
 d) None of these
55. Simple index number is computed for:  
 a) Only for one variable  
 b) For two variables  
 c) Three variables  
 d) Multivariates  
 e) None of these
56. The geometric mean of Laspeyere's and Passche's price indices is also known as:  
 a) Fisher's price index  
 b) Kelly's price index  
 c) Marshal-Edgeworth index number  
 d) Bowley's price index
57. An appropriate method for working out consumer price index is:  
 a) Weighted aggregate expenditure method  
 b) Family budget method  
 c) Price relative method  
 d) None of the above
58. The weights used in Passche's formula belong to  
 a) The base period  
 b) The given period  
 c) To any arbitrary chosen period  
 d) None of the above
59. Which of the following method is used to compute unweighted index number?  
 a) Simple aggregate index  
 b) Average lative index  
 c) Both a) and b)  
 d) None of these
60. Which of the following method is used to compute weighted index numner?
- a) Laspeyre's method  
 b) Paasche's method  
 c) Marshall-Edgeworth method  
 d) All of the above  
 e) None of the above
61. Which of the following is true for quantity index numbers?  
 a) Prices are constant quantities are changed  
 b) Prices are changed quantities are constant  
 c) Both prices and quapties are changed  
 d) Both prices and quantities are constant  
 e) None of these
62. Which of the following method is upwar bias?  
 a) Laspeyre's method  
 b) Paasche's method  
 c) Chain base method  
 d) CPI method  
 e) None of these
63. Which of the following method is downward bias?  
 a) Laspeyre's method  
 b) Paasche's method  
 c) Chain base method  
 d) CPI method  
 e) None of these

## ANSWERS

1	c	2	c	3	d	4	a
5	d	6	d	7	d	8	b
9	a	10	a	11	c	12	c
13	b	14	d	15	b	16	a
17	c	18	b	19	a	20	a
21	d	22	d	23	a	24	c
25	b	26	b	27	d	28	d
29	b	30	a	31	c	32	b
33	c	34	a	35	a	36	b
37	c	38	b	39	b	40	a
41	c	42	b	43	c	44	d
45	b	46	a	47	c	48	c
49	b	50	d	51	c	52	c
53	d	54	c	55	a	56	a
57	b	58	b	59	c	60	d
61	a	62	a	63	b		