Preliminary Assessment of Basic Statistical Concepts

Save this file as Statistics Review to your desktop. While responding, save frequently. When you are done, save it again and email it to me at olsongh3@gmail.com.

Type your name here:

Select the best answer to each of the following items. Type the letter of the answer you select in the space provided to the left.

Multiple Choice
Identify the letter of the choice that best completes the statement or answers the question.

____ 1a. A researcher is interested in the eating behavior of elementary school children and selects a group of 25 children to be tested in a research study. The group of 25 children is a ______.
a. sample
b. statistic
c. population
d. parameter

____ 2b. A researcher is interested in the effect of St. Johns Wort on memory. A group of 25 college students is selected to participate in a research study. The average memory score obtained for the students is a ______.
a. sample
b. statistic
c. population
d. parameter

____ 3c. A researcher is curious about the average IQ of registered voters in the state of Florida. The entire group of registered voters in Florida is an example of a ______.
a. sample
b. statistic
c. population
d. parameter

____ 4c. Statistical techniques that summarize, organize, and simplify data are classified as ____________.
a. population statistics
b. sample statistics
c. descriptive statistics
d. inferential statistics

____ 5a. A characteristic, usually a numerical value, that describes an entire population of scores is a _________.
a. parameter
b. statistic
c. variable
d. constant

____ 6c. How would the following mathematical operation be expressed in summation notation? "Add two points to each score, square the resulting value, then find the sum of the squared numbers."
a. $\Sigma X + 2^2$

b. $(\Sigma X + 2)^2$
c. $\Sigma(X + 2)^2$
d. $\Sigma X^2 + 2$

____ 7c. Frequency distribution polygons are intended for use with ______.
a. either interval or ratio scales of measurement
b. only ratio scales
c. either nominal or ordinal scales
d. only nominal scales

_____ 8c. In a frequency distribution graph, frequencies are presented on the __________ and the scores (categories) are listed on the __________.
a. X axis, Y axis
b. horizontal line, vertical line
c. Y axis, X axis
d. class interval, horizontal line

_____ 9b. In a distribution with positive skew, scores with the highest frequencies(i.e., occur most often) are
a. on the right side of the distribution
b. on the left side of the distribution
c. in the middle of the distribution
d. represented at two distinct peaks

Table 2-2
The following table shows a frequency distribution of quiz scores.

X	f
5	6
4	5
3	5
2	3
1	2

____ 10b. Refer to Table 2-2. How many individuals had a score of X = 2?
a. 1
b. 3
c. 5
d. cannot be determined from the information given

____ 11c. A sample of n = 5 scores has a mean of X ($\bar{X}$) = 9. What is ΣX for this sample?
a. 9/5 = 1.80
b. 5/9 = 0.555
c. 5(9) = 45
d. cannot be determined from the information given

____ 12b. One sample has (n = 4, $\bar{X}$ =10), and a second sample has (n = 8, $\bar{X}$ =20). If the two samples are combined, the mean for the combined sample will be ______.
a. equal to 15
b. greater than 15 but less than 20
c. less than 15 but more than 10
d. none of the above

____ 13a. A sample has (n = 20, $\bar{X}$ =55). After one score is removed from the sample, the mean for the remaining scores is 51. From this information you can conclude that the removed score was ______.
a. greater than 55
b. less than 55
c. It is impossible to estimate the magnitude of the score.

____ 14a. For a distribution of scores, the mean is equal to the median. This distribution is most likely to be ______.
a. symmetrical
b. positively skewed
c. negatively skewed
d. impossible to determine the shape

____ 15c. For any distribution, you can be sure that at least one individual has a score equal to the ______.

a. mean
b. median
c. mode
d. all of the above

____ 16b. For an extremely skewed distribution of scores the best measure of central tendency would be the ______.

a. mean
b. median
c. mode
d. Central tendency cannot be determined for a skewed distribution.

____ 17a. What is the variance (σ^2) of a *population* with N = 5 and sums of squares (SS) = 40?

a. 40/5 = 8
b. $8^2 = 64$
c. 40/4 = 10
d. $10^2 = 100$

____ 18c. A *population* has μ = 40 and σ = 8. If each score is multiplied by 2, the new value of σ will be ______.

a. 4
b. 8
c. 16
d. insufficient information, cannot be determined

____ 19b. You must determine Q1 and Q3 to compute the _______.

a. sum of squared deviations
b. semi-interquartile range
c. variance
d. standard deviation

____ 20d. Of the following z-score values, which one represents the most extreme location on the left-hand side of the distribution?

a. z = +1.00
b. z = +2.00
c. z = -1.00
d. z = -2.00

____ 21a. For a population with μ = 80 and σ = 10, the z-score corresponding to X = 85 would be ______.

a. +0.50
b. +1.00
c. +2.00
d. +5.00

____ 22b. For a population with μ = 60 and σ = 8, the X value corresponding to z = -0.50 would be ______.

a. -4
b. 56
c. 64
d. 59.5

____ 23c. For a population with μ = 60 and σ = 8, the X value corresponding to z = 1.50 would be ______.

a. 12
b. 61.5
c. 72
d. 90

____ 24b. A population of scores has σ = 20. In this population, a score of X = 80 corresponds to z = +0.25. What is the population mean?
a. 70
b. 75
c. 85
d. 90

____ 25d. Under what circumstances would a score that is 15 points above the mean be considered an extreme score?
a. when the population mean is much larger than 15
b. when the population standard deviation is much larger than 15
c. when the population mean is much smaller than 15
d. when the population standard deviation is much smaller than 15

____ 26a. Suppose you earned a score of X = 54 on an exam. Which set of parameters would give you the highest grade?
a. μ = 50 and σ = 2
b. μ = 50 and σ = 8
c. μ = 58 and σ = 2
d. μ = 58 and σ = 8

____ 27d. A population has μ = 50 and σ = 10. If these scores are transformed into z-scores, the population of z-scores will have a mean of ____ and a standard deviation of ____.
a. 50, 10
b. 50, 1
c. 0, 10
d. 0, 1

____ 28b. A distribution with μ = 35 and σ = 8 is being standardized so that the new mean and standard deviation will be μ = 50 and σ = 10. In the new, standardized distribution your score is X = 60. What was your score in the original distribution?
a. X = 45
b. X = 43
c. X = 1.00
d. impossible to determine without more information

____ 29b. A colony of laboratory rats contains 7 albino rats and 23 hooded rats. What is the probability of randomly selecting an albino rat from this colony?
a. 7/23
b. 7/30
c. 23/30
d. 23/7

____ 30d. An introductory psychology class has 9 freshman males, 11 freshman females, 8 sophomore males, and 12 sophomore females. What is the probability of randomly selecting a male from this group?
a. 9/20
b. 8/20
c. 17/20
d. 17/40

____ 31c. For a normal distribution with μ = 40 and σ = 4, what is the probability of sampling an individual with a score less than 46?
a. 0.0668
b. 0.4452
c. 0.9332
d. 0.0548

____ 32c. The distribution of sample means consists of ______.

a. all the scores contained in a particular sample
b. all the scores contained in a specific population
c. the sample means for all the possible random samples (of a specific size) from a particular population
d. the sample means for all the possible random samples (for all sample sizes) from a particular population

____ 33d. The standard error of the mean provides a measure of ______.

a. the maximum possible discrepancy between $\bar{X}$ and μ
b. the minimum possible discrepancy between $\bar{X}$ and μ
c. the exact discrepancy between each specific $\bar{X}$) and μ
d. none of the above

____ 34b. As sample size increases, the standard error of $\bar{X}$ ______.

a. also increases
b. decreases
c. stays constant

____ 35c. A sample of n = 25 scores is determined to have a standard error of the mean (SEM) of 2 points. What is the standard deviation for the population from which the sample was obtained?

a. 2
b. 2/5
c. 10
d. 50

______36a. A random sample is selected from a population with $\mu = 80$ and $\sigma = 10$. To ensure a standard error of 2 points or less, the sample size should be at least ____.

a. n = 5
b. n = 10
c. n = 25
d. It is impossible to obtain a standard error less than 2 for any sized sample.

____ 37a. If sample size (n) is held constant, the SEM will _______ as the population variance increases.

a. increase
b. decrease
c. stay constant
d. cannot answer with the information given

____ 38b. The final step of hypothesis testing is to ______.

a. locate the values associated with the critical region
b. make a statistical decision about H_0
c. collect the sample data and compute the test statistic
d. state the hypotheses and select an alpha level

____ 39a. The *critical region* for a hypothesis test consists of ______.

a. outcomes that have a very low probability if the null hypothesis is true
b. outcomes that have a high probability if the null hypothesis is true
c. outcomes that have a very low probability whether or not the null hypothesis is true
d. outcomes that have a high probability whether or not the null hypothesis is true

____ 40d. In a hypothesis test, an extreme z-score value, like $z = +3$ or $z = +4$, _____.
- a. is probably in the critical region
- b. means that you should probably reject the null hypothesis
- c. is strong evidence of a statistically significant effect
- d. all of the above
- e. none of the above

____ 41a. A Type I error means that a researcher has ______.
- a. falsely concluded that a treatment has an effect
- b. correctly concluded that a treatment has no effect
- c. falsely concluded that a treatment has no effect
- d. correctly concluded that a treatment has an effect

____ 42c. The probability of committing a Type I error ______.
- a. is determined solely by the size of the treatment effect
- b. cannot be controlled by the experimenter
- c. is determined by the level of significance that one chooses
- d. is determined by the value for beta (β) one selects

____ 43d. When $\alpha = .05$, the critical boundaries for a hypothesis test are $z = +1.96$ and -1.96. If the z-score for the sample data is $z = -1.90$, then the correct statistical decision is
- a. fail to reject H_1
- b. fail to reject H_0
- c. reject H_1
- d. reject H_0

____ 44c. A researcher reports that there is a significant difference between two treatments at the .05 level of significance. This means that _______.
- a. if the researcher repeated the experiment, there is a 95% probability that the same decision would be reached
- b. the difference obtained in the experiment is at least 5% larger than the standard error
- c. there is a 5% probability (or less) that the difference is due to chance
- d. the average score for one treatment is at least 5% higher than the average score for the other treatment

____ 45b. What is the *estimated* standard error of the mean, in the i, obtained from a sample with $n = 9$ and $s^2 = 36$?
- a. $\sqrt{36/9}$
- b. $\sqrt{36/8}$
- c. $36/\sqrt{9}$
- d. $36/\sqrt{8}$

____ 46c. One sample has $n = 6$ and $SS = 20$ and a second sample has $n = 6$ and $SS = 30$. What is the pooled variance for the two samples?
- a. 25
- b. 50
- c. 50/10
- d. 50/12

____ 47c. A researcher reports $t(24) = 5$ for a 2-group, independent-measures experiment. How many subjects participated in the entire experiment?
- a. 24
- b. 25
- c. 26
- d. 12

____ 48c. Which of the following most closely resembles the general form of an interval estimate?

a. statistic = parameter ± error
b. statistic = parameter × error
c. parameter = statistic ± error
d. error = parameter ± statistic

____ 49b. If all other factors are held constant, increasing the sample size will ______.

a. increase the standard error
b. decrease the width of the confidence interval
c. increase the width of the confidence interval
d. none of the above

____ 50c. Which combination of factors would definitely reduce the width of a confidence interval?

a. use a larger sample and increase the level of confidence
b. use a smaller sample and increase the level of confidence
c. use a larger sample and decrease the level of confidence
d. use a smaller sample and decrease the level of confidence

____ 51c. A researcher uses analysis of variance to test for mean differences among four treatments with a n = 6 in each treatment. The F-ratio for this analysis would have what df values?

a. df = 3, 5
b. df = 3, 15
c. df = 3, 20
d. df = 4, 24

____ 52b. A researcher reports an F-ratio with df = 3, 36 for an independent-measures experiment. How many treatment conditions were compared in this experiment?

a. 3
b. 4
c. 36
d. 39

____ 53d. An experiment compares two treatment conditions with n = 20 in each treatment. If the data are analyzed with ANOVA, the analysis would have df_{total} = _____.

a. 18
b. 19
c. 38
d. 39

____ 54a. A *treatment effect* refers to differences between scores that are *due to* the different treatment conditions. The differences (or variability) produced by *treatment effects* will contribute to ______.

a. the numerator of the F-ratio
b. the denominator of the F-ratio
c. both the numerator and the denominator of the F-ratio
d. *Treatment effects* do not contribute to the F-ratio because they are removed before the F-ratio is computed.

____ 55b. Under what circumstances are post tests necessary?

a. reject the null hypothesis with k = 2 treatments
b. reject the null hypothesis with k > 2 treatments
c. fail to reject the null hypothesis with k = 2 treatments
d. fail to reject the null hypothesis with k > 2 treatments

____ 56c. A research study compares three treatments with n = 5 in each treatment. If the data were examined using an analysis of variance, the F-ratio would have df = _____.

a. 12
b. 14
c. 2, 12
d. 2, 14

____ 57a. In a two-factor analysis of variance a main effect is defined as ______.

a. the mean differences among the levels of one factor
b. the mean differences among all treatment conditions
c. the mean difference between the two factors
d. the difference between the largest treatment mean and the smallest treatment mean

____ 58d. If the mean and variance are computed for each sample in two-factor experiment, which of the following types of sample data will tend to produce large F-ratios for the two-factor ANOVA?

a. big values for the sample means
b. small values for the sample means
c. big values for the sample variances
d. small values for the sample variances

____ 59a. A set 5 pairs of X and Y values has SS_X = 10, SS_Y = 40 and SSCP = 10. What is the Pearson correlation? ______.

a. r = 10/20 = 0.50
b. r = 10/400 = 0.025
c. $r = \sqrt{10/20} = 0.707$
d. r = 40/100 = 0.40

____ 60a. In the general linear equation Y = bX + a, the value of a is called ______.

a. the slope constant
b. the Y-intercept
c. the X-intercept
d. the beta factor

____ 61a. What information *cannot* be determined from the general linear equation Y = bX + a?

a. the correlation between X and Y
b. the slope
c. the Y intercept
d. a and c

____ 62c. The entrance fee for a theme park is $20. Tickets for each ride and attraction are $3 a piece. Which of the following equations describes the relation between the total cost (Y) and the number of tickets purchased (X) in a single visit to the park?

a. Y = 20X + 3
b. Y = 60X
c. Y = 3X + 20
d. X = 3Y + 20

____ 67d. In the test for goodness of fit, the chi-square statistic has degrees of freedom equal to ______.

a. n - 1
b. n - 2
c. n - C (where C is the number of categories)
d. none of the above

____ 68d. The chi-square test for independence is used to test for ______.

a. a mean difference between two populations
b. a difference between a sample distribution and a population distribution
c. a difference in variance between two populations
d. a relationship between two variables

____ 69c. The chi-square test for independence can be used to evaluate ______.

a. the relationship between two variables
b. differences between two or more population frequency distributions
c. both of the above
d. neither of the above

____ 70d. The chi-square test for independence has degrees of freedom given by _____.

a. R x C
b. R x C(n - 1)
c. n - R x C
d. (R - 1)(C - 1)

____ 71a. A chi-square test for independence is being used to evaluate the relationship between two variables, one of which is classified into 3 categories and the second of which is classified into 4 categories. The chi-square statistic for this test would have df equal to ______.

a. 6
b. 7
c. 10
d. 11

____ 72. Describe the sequence of mathematical operations that would be used to evaluate each of the following expressions.

a. ΣX^2 Sum the squares
b. $(\Sigma X)^2$ Sum then square
c. $\Sigma X + 1$ Sum, then add one
d. $\Sigma(X + 1)$ Add one to each X, then sum
e. $\Sigma(X + 1)^2$ Add one to each X, square the result, then sum

____ 73. Calculate each value requested for the following set of scores. Scores: 0, 1, 4, 2

a. $\Sigma X + 2$ 9
b. $\Sigma(X + 2)$ 15
c. $\Sigma(X + 2)^2$ 65

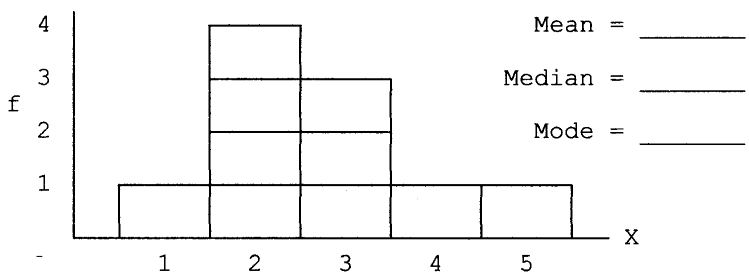
____ 75. A distribution of N = 10 scores has $\mu = 50$ and $\sigma = 10$. If another five individuals, all with scores of X = 50, are added to this distribution, what will happen to the standard deviation? Will it increase, decrease, or stay the same? Explain your answer.

Adding scores at the mean will not change the mean, but it increase the number of scores clustered around the mean, hence the variance (and thus the SD) will decrease.

____ 76. A set of exam scores are reported as X values and z-scores. On this exam a score of X = 75 corresponds to a z-score of z = +2.00, and a score of X = 60 corresponds to a z-score of z = -1.00. What are the values for the mean and standard deviation for this exam? (Hint: Sketch a distribution and locate each of the z-score positions.)

One SD = 5 points. Since 2z is 10 pts above the mean, the mean must be 75 – 2z or 75-10 = 65

77. Find the mean, the median, and the mode for the set of scores in the frequency distribution histogram below.



Mean 27/10 = 2.7

Median 2.5

Node 2

For the next two items, type your answer in the box provided (the box will expand as needed).

78. A set of exam scores are reported as X values and z-scores. On this exam a score of X = 75 corresponds to a z-score of z = +2.00, and a score of X = 60 corresponds to a z-score of z = -1.00. What are the values for the mean and standard deviation for this exam? (Hint: Sketch a distribution and locate each of the z-score positions.)

One SD = 5 points. Since 2z is 10 pts above the mean, the mean must be 75 – 2z or 75-10 = 65

80. State the alternative hypothesis (H_1) for an ANOVA comparing four treatment conditions. Explain why the hypothesis is stated in this way.

A least one pair of means are different (i.e., at least one mean is different from the others). All four means are not equal.

For the next two items, fill in the blanks in the tables.

81. A researcher would like to know whether the change in seasons has any consistent effect on students' academic performance. In the middle of each of the four seasons the researcher selects a random sample of n = 10 students. Each individual in these four separate samples is given a standardized test. The data from this study are examined using an analysis of variance and the results are shown in the summary table below. Fill in all missing values in the table.

Source	SS	df	MS	
Between Treatments	_60_	_ 3 _	_20_	F = _10_
Within Treatments	_72_	_36_	2	
Total	132	_39_		

82. The following table summarizes the results of a two-factor ANOVA evaluating an independent-measures experiment with two levels of factor A, three levels of factor B, and n = 5 subjects in each separate sample. Fill in all missing values in the table. (Hint: Start with the df column.)

Source	df	SS	MS	
Between Treatments	_5 _	60		
Factor A	_1 _	_10_	_10_	F_A = 5
Factor B	_2 _	_20_	_10_	F_B = _5__
A x B	_2 _	30	_15_	F_{AxB} = _7.5_
Within Treatments	_24_	_48_	2	
Total	_29_	_108_		

Save the file and email it to me at olsongh3@gmail.com.