Name: _____
ID: 810-_____

Assignment 1 (solution)

Part I

1. Using the rules of logarithms, solve the following:

   a. $\log\left(\frac{47}{23}\right)$ b. log(13 * 27) c. $\log(8^6)$

   a. log (47) – log (23) =0.3103700219
   b. log (13) + log (27) = 2.545307116
   c. 6 * log (8) = 5.418539922

2. Using the rules of exponents, solve the following:

   a. $4^2 * 4^6$ b. $(3^4)^3$ c. $\frac{7^5}{7^2}$ d. $(7/13)^8$

   a. $4^{(2+6)} = 65536$
   b. $3^{(4*3)} = 531441$
   c. $7^{(5-2)} = 343$
   d. $\frac{7^8}{13^8} = 0.0070670392$

3. Students' scores on an exam range from 35 to 98, with a mean score of 74. Which of the following is the most realistic value for the standard deviation: -10, 0, 3, 12, or 63? Clearly explain what's unrealistic about each of the other values.

   12 is the most realistic value. If 12 was the standard deviation, almost all of the observations would fall within 3 standard deviations of the mean.
   The standard deviation cannot be a negative number, so -10 is not realistic.
   If the standard deviation was zero, then the range would be zero.
   63 is too large and 3 is too small, given the context and the range given.

§ Read the following and answer the questions (4-7)

Jesse and Mallory each run an average of 4 miles per day. However, they don't always the run the same number of miles as each other each day.

4. Determine 7 observations that could possibly be the number of miles Jesse ran each day this week.

   Answers will vary: [2 3 4 4 4 5 6]

5. Determine 7 observations that could be how many miles Mallory ran each of these days, if the distance of Mallory's runs varies more than Jesse's.

   Answers will vary: [1 1 2 4 6 7 7], [1 2 3 4 5 6 7], [2 2 2 2 2 9 9]

6. Determine 7 observations that could be how many miles Mallory ran each of these days, if the distance of Mallory's runs has the same variance as the distance of Jesse's runs.

   Answers will vary: [1 4 4 4 4 4 7]

7. In this context, explain why the variance is a useful descriptive statistic to use in addition to the mean to describe the running habits of Jesse and Mallory.

The variance tells you how the lengths of the daily runs differ from the average length of a daily run. In the context of running, even though they both average 4 miles per day, one runner may run more consistently around 4 miles every day [4 4 4 4 4 4 4] and one runner may run, for example, fewer miles Monday through Friday and more miles Saturday and Sunday when they have more time [2 2 2 2 2 9 9]. Those running habits are different and may determine differences in other factors of interest to a runner like conditioning or weight loss.

8. In what case or cases would the mean, mode, and median all equal each other?

9. In the case of severe skewness what is the best measure of central tendency and why?

10. Fill the table below to calculate the following terms (show your work)

student	language art (X)	reading (Y)	X^2	Y^2	XY
Karl	5	4	25	16	20
Jennifer	4	5	16	25	20
Andy	3	5	9	25	15
Sue	1	3	1	9	
Thomas	2	2	4	4	4
sum	15	19	55	79	62

a. Mean and standard deviation of Language Art scores
Mean=15/5=3
variance=[55-(15)^2/5]/4=2.5
SD=1.58

b. Mean and standard deviation of Reading scores
Mean 19/5=3.8
variance=[79-(19)^2/5]/4=1.7
SD=1.30

c. Covariance and correlation of Language Art and Reading scores
covariance =[62-(15*19)/5]/4=1.25
correlation=1.25/(1.58*1.3)=0.60856 (0.606339)

Part II (SPSS)

You have just administered your second biology quiz of the semester. You have scanned your 'bubble sheets' through the optical scanner and you now have a tab delimited file in the structure presented below. Answer each question below by providing SPSS output and/or a brief response.

1. Import the tab delimited file into SPSS according to the file layout
    a. Add variable labels
    b. Add value labels
    c. Recode all missing values to '9' (missing values are present in ethnic and Required)
        i. Set 9 as 'Not specified' within the variable view
    d. Why are the start and end positions presented in the layout blank? In what instance would they be critical?

Field Description	Variable Name	Start Position	End Position	Format	Valid Values
Student Identifier	ID			A2	
Student Gender	Gender			F1.0	1=Female 2=Male
Student Ethnicity	Ethnic			F1.0	1=Asian 2=African American 3=Hispanic 4=Native American 5=White
Course Requirement Flag	Required			F1.0	0=No, Course is NOT required 1=Yes, Course IS required
Score Student Received on Quiz 1	Q1Score			F2.0	
Item Responses for Quiz 2	Item1-Item12			12(F1.0)	0=Incorrect 1=Correct

2. Compute a total QUIZ 2 Score for each student

3. Using SPSS, describe your data set in any way you see fit (must use the concepts we discussed in class—you may want to describe the students, the items, the total scores etc.)

4. Select and run descriptive statistics:
    a. Select the females from the data file. What is their mean Total Quiz 2 Score?
    b. Select the males from the data file. What is their mean Total Quiz 2 Score?
    c. Use an alternate SPSS Menu option that will reduce the number of steps to compare group means.

5. Which student(s) achieved the highest grade? How can you describe the score(s), above and beyond a raw score, within the data file?
   Report your transformed score for the highest achieving student(s).

KEY SPSS Component

1) Import the tab delimited file into SPSS according to the file layout
   a. Add variable labels
   b. Add value labels
   c. Recode all missing values to '9' (missing values are present in ethnic and Required)
      i. Set 9 as 'Not specified' within the variable view
   d. Why are the start and end positions presented in the layout blank? In what instance would they be critical?

[a-c may be apparent through their output tables under #3. D) Start/End positions are not necessary for a file that is tab/character delimited. They would be critical for a raw text/fixed/ASCII file]

2) Compute a total QUIZ 2 Score for each student
[will be apparent from their answer to #4]

3) Using SPSS, describe your data set in any way you see fit (must use the concepts we discussed in class—you may want to describe the students, the items, the total scores etc.)
[As long as they have attempted to run a few of the menu commands we reviewed and interpreted it correctly, its correct]

Student Ethnicity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Asian	4	12.5	13.3	13.3
	African American	10	31.3	33.3	46.7
	Hispanic	4	12.5	13.3	60.0
	Native American	1	3.1	3.3	63.3
	White	11	34.4	36.7	100.0
	Total	30	93.8	100.0	
Missing	Not Specified	2	6.3		
Total		32	100.0		

Course Required Flag

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	10	31.3	32.3	32.3
	Yes	21	65.6	67.7	100.0
	Total	31	96.9	100.0	
Missing	Not Specified	1	3.1		
Total		32	100.0		

Descriptives

			Statistic	Std. Error
Quiz2	Mean		8.2500	.38626
	95% Confidence Interval for Mean	Lower Bound	7.4622	
		Upper Bound	9.0378	
	5% Trimmed Mean		8.2569	
	Median		7.5000	
	Variance		4.774	
	Std. Deviation		2.18499	
	Minimum		4.00	
	Maximum		12.00	
	Range		8.00	
	Interquartile Range		3.00	
	Skewness		.134	.414
	Kurtosis		-1.119	.809

Correlations

		Score Student Recieved on Quiz 1	Quiz2
Score Student Recieved on Quiz 1	Pearson Correlation		
	Sig. (2-tailed)		
	N		
Quiz2	Pearson Correlation	.770**	
	Sig. (2-tailed)	.000	
	N	32	

**. Correlation is significant at the 0.01 level (2-tailed).

4) Select and run descriptive statistics:
   a. Select the females from the data file. What is their mean Total Quiz 2 Score?
   b. Select the males from the data file. What is their mean Total Quiz 2 Score?
   c. Use an alternate SPSS Menu option that will reduce the number of steps to compare group means.

Report

Quiz2

Student Gender	Mean	N	Std. Deviation
Female	7.6471	17	1.65609
Male	8.9333	15	2.54858
Total	8.2500	32	2.18499

5) Which student(s) achieved the highest grade? How can you describe the score(s), above and beyond a raw score, within the data file?
   Report your transformed score for the highest achieving student(s).

[Student 11 & Student 25 both answered all items correct. You can add z scores into a data file easily via checking the 'Save standardized values as variables' within descriptives--z score 1.72]