Lab Assignment No. 2: Answer Key

1) Using the dataset (t5_2.sas7bdat), what is the correlation matrix between all response variables?

Pearson Correlation Coefficients, N = 89

	sshist	verbal	science
sshist	1.00000	0.30025	0.53493
	0.0043	<.0001	
verbal	0.30025	1.00000	0.33390
	0.0043		0.0014
science	0.53493	0.33390	1.00000
	<.0001	0.0014	

Based on the computed correlation matrix all three variables appear to be correlated with one another at the $p < .01$ level. Given that these data are analyzed in the multivariate environment some underlying communality is expected. In fact these three assessments are subcomponents of a larger overall assessment. Therefore significant correlations would be expected.

2) Evaluate univariate normality. Take appropriate steps. Justify your decisions.

Tests for Normality for sshist

Test		--Statistic---	-----p Value------	
Shapiro-Wilk	**W**	**0.987798**	**Pr < W**	**0.5781**
Kolmogorov-Smirnov	D	0.093009	Pr > D	0.0566

For the *sshist* variable we observe a Shapiro-Wilk statistic of $W = .99$, $p > .05$. We therefore conclude that the variable appears to be normally distributed.

Tests for Normality for verbal

Test		--Statistic---	-----p Value------	
Shapiro-Wilk	**W**	**0.653503**	**Pr < W**	**<0.0001**
Kolmogorov-Smirnov	D	0.169537	Pr > D	<0.0100

For the *verbal* variable we observe a Shapiro-Wilk statistic of $W = .65$, $p < .001$. We therefore conclude that the variable appears to be non-normally distributed. Further investigation of the corresponding extreme observations table, box plot,

and stem-and-leaf plot found a single observation (id = 89) that was unrepresentatively larger than the remaining bulk of the data. Because of the extremely unlikely (and in fact impossible) score the observation was deleted from the dataset and excluded from further analyses. The normality test was rerun.

Tests for Normality for verbal

Test		--Statistic---		-----p Value------
Shapiro-Wilk	**W**	**0.967803**	**Pr < W**	**0.0274**
Kolmogorov-Smirnov	D	0.084575	Pr > D	0.1214

The deletion of the extreme outlier improved the variables distribution. However, it did not make it normal. The extreme outlier table was also consulted:

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
25	50	70	83
28	66	71	63
28	48	72	10
32	17	73	73
35	59	75	51

Screening the five lowest and highest remaining values in the non-normally distributed variable did not provide further evidence against any one observation. Since multivariate normality is not guaranteed by univariate normality the variable was no further altered.

Tests for Normality for science

Test		--Statistic---		-----p Value------
Shapiro-Wilk	**W**	**0.985687**	**Pr < W**	**0.4386**
Kolmogorov-Smirnov	D	0.079914	Pr > D	>0.1500

For the *science* variable we observe a Shapiro-Wilk statistic of $W = .98$, $p > .05$. We therefore conclude that the variable appears to be normally distributed.

3) Evaluate multivariate normality (Leverage Values, Mahalonobis Distances).

4) What are the Mardia's coefficients for these data? Interpret. Take appropriate steps. Justify your decisions.

For the three variables consistent means have been observed over the past several years. Subsequently the means of 520 for *sshist*, 55 for *verbal* and 22 for *science* are believed to be the population means.

5) Conduct univariate one-sample t-tests. What are your conclusions?

6) Compute the corresponding Hoteling's T^2.

7) What is your interpretation regarding the data based on your analyses in questions 5 and 6?

8) Transform the dataset you have computed the Hoteling's T^2 on as follows: Divide *sshist* by 100 (call it *hist*), multiply verbal by 1.5 and add 10 (call it *verb*), and subtract 5 from *science* and divide it by 5 (call it *scien*). What are the new measures of central tendency?

9) Rerun the Hoteling's T^2 on your newly transformed dataset. ***(Remember you must also transform the corresponding hypothesized means accordingly).***

10) What conclusions can you draw regarding the multivariate generalization of the t-test?