ASSIGNMENT 1

USING THE T-TOOLS

In this assignment you will apply descriptive and graphical tools in SPSS to examine data about brain sizes for two groups of mammals. You will also learn how to transform the original measurements when the distribution assumptions are violated, and how to apply the t-tools to make inferences about the data. Moreover, you will compare the inferences produced by the t-tolls with those obtained with some distribution-free methods. Your statistical conclusions will be made within the limit of the study design.

Brain Size of Mammals

Biologists have keen interest in the characteristics that enable a species to withstand the selective mechanisms of evolution. In particular, they drive evolutionary conclusions from investigating which physiological characteristics are related to brain size. In an attempt to study these characteristics a group of evolution biologists collected data on average values of brain weight, body weight, gestation lengths (length of pregnancy), and litter size for 96 species of mammals. The data were obtained from zoo records and literature. The data are available in the paper by G.A. Sacher and E.F. Staffeldt, "Relation of Gestation Time to Brain Weight for Placental Mammals; Implications for the Theory of Vertebrate Growth," American Naturalist 108 (1974): 593-613.

In order to determine if mammals with large litter sizes have bigger brains than those with smaller litter sizes, a cut off point was chosen using the information collected on average litter size and the mammals were thus categorized into small or large litter size groups. Two variables derived from the original data set are:

Column	**Variable Name**	**Description of Variable**
1	BRAINSIZ	Relative brain weights (brain weight divided by body weight),
2	LITTERSI	Litter size = SMALL or LARGE.

The above data are available in the SPSS file *lab1.sav* that can be downloaded to your local station from the Statistical Laboratories web site at http://www.stat.ualberta.ca/statslabs/index.htm (click *Stat 252* link and *Data* for *Lab 1*). The data are not to be printed in your submission.

Answer the following questions:

1. Comment on the study design. In particular, is this an observational study or experimental study? What kind of inference or interpretation can you make? Is it possible to establish a casual link between brain size and litter size using this data? Can the finding be generalized to the population of mammals?

2. Use the *Explore* procedure to obtain the descriptive statistics, the side-by-side boxplots, and the normality plots of relative brain weights. In particular:

   (a) Obtain and paste the descriptive statistics for the relative brain weight for each litter size into your report. Compare the means and standard deviations of the two distributions. Moreover, report the standard errors of the means for two litter sizes. What does the standard error measure?

   (b) What is the 95% confidence interval for the mean relative brain weight in each group (confidence interval can be obtained from the descriptive statistics in part (a)). Which confidence interval is wider? Explain why the wider one is wider. In particular, tie your explanation back to standard errors reported in part (a). Are the two confidence intervals overlapping?

(c) Obtain and paste the side-by-side boxplots of relative brain weight for the two groups into your report. Compare the medians and the interquartile ranges of the two distributions. Comment about the shape (symmetric, skewed or neither) of each distribution. What would you recommend for a measure of central tendency for both distributions – the mean or median? Justify your choice. Are there any outliers?

(d) Obtain and paste the normality plots for each distribution into your report. Comment briefly about each plot. In particular, is there any evidence that the assumption of normality may be seriously violated in either case?

(e) What are the assumptions necessary for the application of t-tools to compare the mean relative brain weights for the two groups? Using your observations in questions (a) to (c), is there any evidence that some of these assumptions necessary may be seriously violated?

3. Now, obtain a log-transformation of relative brain weight and use the *Explore* procedure to obtain the descriptive statistics, the side-by-side boxplots, and the normality plots of the log-transformed relative brain weights.

(a) Obtain and paste the descriptive statistics for each litter size group into your report. Compare the means and standard deviations of the two distributions. Moreover, compare the standard errors of the means for the two litter sizes.

(b) Obtain and paste the side-by-side boxplots into your report. Compare the medians. Comment about the shape (symmetric, skewed or neither) and spread of each distribution.

(c) Obtain and paste the normality plots for each distribution into your report. Is there any evidence that the assumption of normality may be seriously violated in either case?

(d) Comment on the overall effect of the log-transformation on the distribution of relative brain weight.

4. Now you will use the appropriate t-tools to analyze the data on the log-transformed scale.

(a) Obtain the difference between the average of log-relative brain weight for large litter size group and small litter size group (large minus small). Then obtain the antilogarithm of this difference. Interpret this antilogarithm of the difference in the context of this study.

(b) Is there any evidence that relative brain sizes tend to be different for the two groups? Answer the question by carrying out an appropriate test in SPSS. In particular, state the null and alternative hypotheses, specify the distribution of the test statistic under the null hypothesis, report the value of the test statistic, and the p-value of the test. State your conclusion.

(c) Using the SPSS output of part (b) and hand calculations report the 95% confidence interval for the difference in average relative brain weight (original scale, not log scale) between the two groups of mammals. What would you conclude about the difference in relative brain weight between the two groups of mammals? Is your conclusion consistent with the outcome of the test in part (b)?

5. Verify the SPSS output in part (b) of Question 4. In particular, use hand calculations to obtain the value of the standard error of the difference in mean relative brain weight for the two groups (for the log-transformed data) and the value of the test statistic (show your calculations). You may use the values of the mean log-relative brain weight and the standard deviation for each group from the SPSS output in part (a) of Question 3. Is your hand calculation consistent with the SPSS output?

6. In Question 4 you applied the t-tools to the log-transformed data because the model assumptions of the t-test were grossly violated on the original scale. It is also possible to use some distribution-free methods available in SPSS such as the Mann-Whitney test (rank-sum test) to test for the differences between the two litter size groups.

   (a) Use the Mann-Whitney test in SPSS (*Analyze*, *Nonparametric Tests*, *2 Independent Samples*) to see whether there is evidence that the relative brain weight is different for the two groups. In particular, define the null and alternative hypotheses and report the value of the test statistic.

   (b) What do you conclude about the association between litter size and relative brain weight? Is the outcome of the test consistent with the test in Question 4?

7. Suggest other factors (other than litter size) that might have been considered in the study.

LAB ASSIGNMENT 1 MARKING SCHEMA

Proper Header and Appearance**:** 10 points

Question 1

Description of study: 2 points
Study design: 1 point
Type of inference: 1 point
Scope of inference: 1 point

Question 2

(a) Descriptive statistics: 4 points (2 points for each group)
    Comparison of mean and standard deviations: 2 points
    Standard errors and meaning: 4 points
(b) Confidence intervals: 4 points (2 points each)
    Wider interval and explanation: 3 points
    Overlap: 1 point
(c) Boxplots: 4 points
    Comparisons of median and interquartile range: 2 points
    Shape of distribution: 2 points
    Measure of central tendency and justification: 3 points
    Outliers: 1
(d) Normal probability plot: 4 points
    Description of pattern: 2 points
(e) T-tool assumptions: 2 points
    Violation: 1 point

Question 3

(a) Descriptive statistics: 4 points (2 points per group)
    Comparison of mean and standard deviations: 2 points
    Comparison of standard errors: 2 points
(b) Boxplots: 4 points
    Comparison of median: 2 points
    Shape and spread: 2 points
(c) Normal probability plots: 2 points
    Description of pattern: 2 points
(d) Effect of log-transformation: 2 points

Question 4

(a) Difference of averages: 2 points
Antilog of difference: 2 points
Interpretation of antilog of difference: 4 points
(b) T-test: 6 points itemized as follows:
Hypotheses: 2 points
Test-statistic: 1 point
P-value: 1 point
Null distribution: 1 point
Conclusion: 1 point
(c) Confidence intervals: 4 points
Conclusion: 2 points
Consistency: 1 point

Question 5

Standard error calculation: 4 points
Test-statistic: 2 points
Consistency: 1 point

Question 6

(a) Mann-Whitney test: 5 points itemized as follows:
Hypotheses: 2 points
P-value: 1 point
(d) Conclusion and consistency: 2 points

Question 7

Suggested factors: 2 points

TOTAL = 111