DATA TRANSFORMATIONS, T-TOOLS AND ALTERNATIVES

As long as the data are approximately normal, the t-tools can be used to make approximate but valid and reliable inferences from the data. However, if the assumption of normality is seriously violated, the t-tools may lead to wrong conclusions based on the data.

In this assignment you will learn how to transform the original measurements when the normality assumption is violated, and how to get exact inferences about the data. You will use both t-tools and distribution-free methods available in SPSS to make the inferences about the data.

Fetal Lung-to Liver Signal Intensity at MRI

The recent developments in neonatal medicine has made it possible to preserve the lives of many newborns who, had they been born several decades ago, would undoubtedly have perished. For example, infants whose lungs are not sufficiently developed to permit them to breathe on their own can now be saved with new therapies. One of the serious conditions suffered by many newborns is pulmonary hypoplasia characterized by incomplete development of lung tissue. So far autopsy has been used as a standard and reliable diagnostic tool to identify fetuses with hypoplastic lungs. Unfortunately, an efficient method of accurately diagnosing the condition during gestational age has yet to be determined.

In the recent paper by L.J. Brewerton etc. "Fetal lung-to-liver signal intensity ratio at MR imaging: Development of a normal scale and possible role in predicting pulmonary hypoplasia in utero" (Radiology. 2005 Jun; 235(3): 1005-1010) a new method based on magnetic resonance imaging (MRI) was proposed to diagnose patients with pulmonary hypoplasia. The researchers followed 157 pregnant patients from the Northern Alberta Perinatal Program and recorded the lung-to-liver signal intensity ratio (LLSIR) in fetuses by using half-Fourier single-shot turbo spin-echo MRI. One of the goals of the study was to determine whether there are significant differences between LLSIR readings for normal and hypoplastic lungs.

In this assignment, you will use a subset of the data for 50 patients, including 25 fetuses with normal lung function and the other 25 fetuses with hypoplastic lungs. You will use the data to evaluate the usefulness of MRI in the diagnosis of pulmonary hypoplasia. The data are available in the SPSS file lab1.sav located on the STAT 337 Laboratories web site at http://www.stat.ualberta.ca/statslabs/index.htm (click Stat 337 link). Some of the data have been modified for the needs of the assignment. The data are not to be printed in your submission.

Answer the following questions using the data:

1. First you will analyze the study design.

    (a) What was the purpose of the study?

    (b) How any possible misdiagnosis of hypoplasia among the 50 fetuses might have affected the outcome of the study? Explain briefly.

    (c) Is the study an example of an observational study or a randomized experiment? Can you generalize the statistical results to all fetuses with pulmonary hypoplasia? Provide brief explanations.

In Questions 2 to 4, you will compare the mean LLSIR for left normal lungs with that for right normal lungs using the first two columns in the data file. Notice that each row contains the LLSIR values for the same patient.

Column	Variable Name	Description of Variable
1	LEFT	LLSIR values for the left normal lungs,
2	RIGHT	LLSIR values for the right normal lungs.

2. Use the *Explore* procedure to obtain the descriptive statistics of LLSIR values separately for left and right normal lungs. Obtain and paste the descriptive statistics into your report. Compare the means and standard deviations of the two distributions.

3. Use the *Compute* command in the *Transform* menu to form a new variable *DIFF* defined as follows: *DIFF = LEFT-RIGHT*. Then obtain a boxplot of the differences in LLSIR between left and right normal lungs. Paste the plot into your report. Comment about the shape (symmetric, skewed) of the distribution. Are there any outliers?

4. Do the data allow us to conclude that the mean LLSIR for left normal lungs is different from the mean LLSIR for right normal lungs? Carry out the appropriate test in SPSS to answer the question. In particular, specify the null and alternative hypotheses, report the value of the test statistic, the distribution of the test statistic under the null hypothesis, and the p-value of the test. State your conclusions.

5. The test used in the previous question assumes that the differences come from a normal population. Use the *Q-Q Plot* procedure in the *Graph* menu to test the assumption. Obtain and paste the plot into your report. What do you conclude?

In Questions 6 to 9, you will use the data in columns 3 and 4 in the data file to compare the mean LLSIR of normal lungs with the mean LLSIR of hypoplastic lungs. The following is a description of the corresponding variables in the data file:

Column	Variable Name	Description of Variable
3	LLSIR	average LLSIR value of the left and right lungs for each patient,
4	TYPE	1 = normal lungs, 2 = hypoplastic lungs.

6. Use the *Explore* procedure to obtain the descriptive statistics, the side-by-side boxplots, and the normality plots of LLSIR separately for normal lungs and hypoplastic lungs. In particular:

    (a) Obtain and paste the descriptive statistics for each group into your report. Compare the means and standard deviations of the two distributions.

    (b) Obtain and paste the side-by-side boxplots into your report. Comment about the center, spread, and shape (symmetric, skewed) of each distribution. Are there are any outliers?

    (c) Obtain and paste the normality plots for each distribution into your report. Is there any evidence that the assumption of normality necessary to apply the t-tools may be seriously violated in either case? Explain briefly.

7. Use the *Compute* command in the *Transform* menu to create a new variable *LNLLSIR* defined as follows: *LNLLSIR = LN(LLSIR)*, that is, apply the natural logarithm transformation to the *LLSIR* variable. Then use the *Explore* procedure to carry out the analysis in Question 6 for the log-transformed data. In particular:

    (a) Obtain and paste the descriptive statistics for each group for the log-transformed data into your report. Compare the means and standard deviations of the two distributions.

    (b) Obtain and paste the side-by-side boxplots of the log-transformed data into your report. Comment about the shape (symmetric, skewed) of each distribution. Compare the centers and spreads of the two distributions. Are there are any outliers?

    (c) Obtain and paste the normality plots for each distribution into your report. Is there any evidence that the assumption of normality necessary to apply the t-tools to the log-transformed data may be seriously violated in either case? Explain briefly.

8. Now you will use SPSS and hand calculations to make inferences about the differences in LLSIR between normal and hypoplastic lungs

    (a) Use the SPSS output in part (a) of Question 7 to estimate the differences in LNLLSIR (LLSIR on the log scale) between normal and hypoplastic lungs. Moreover, obtain an estimate of the ratio of the medians of LLSIR of normal lungs to that of hypoplastic lungs.

    (b) Do the data provide sufficient evidence that the mean LLSIR value for normal lungs is different from that for hypoplastic lungs? Answer the question by carrying out an appropriate parametric test in SPSS for the log-transformed data. In particular, define the null and alternative hypotheses; report the value of the test statistic, and the p-value of the test. State your conclusions. Moreover, report a 95 % confidence interval for the differences in LNLLSIR between normal and hypoplastic lungs.

    (c) Use the 95% confidence interval obtained in part (b) to compute by hand a 95% confidence interval for the ratio of LLSIR between normal and hypoplastic lungs (take the antilogarithms of the endpoints). Summarize the interval in one sentence.

LAB ASSIGNMENT 1 MARKING SCHEMA

Question 1

(a)   Purpose of the study: 2 points
(b)   Misdiagnosis effect: 2 points
(c)   The study design: 1 point
        Statistical inferences: 1 point

Question 2

Descriptive Statistics for each group: 2 points each (4 total)
Comparison of means, and standard deviations: 2 points

Question 3

Boxplot of differences: 4 points
Shape of the distribution: 1 point
Outliers: 1 point

Question 4

Null and alternative hypotheses: 3 points
The test statistic: 2 points
Null distribution: 2 points
P-value: 2 points
Conclusions: 2 points

Question 5

Q-Q plot: 4 points
Conclusions: 2 points

Question 6

(a)   Descriptive Statistics for each group: 2 points each (4 total)
          Comparison of means, and standard deviations: 2 points
(b)   Side-by-side boxplots: 4 points
          Center, spread, and shape: 3 points
          Outliers: 1 point
(c)   Normality plots: 2 points each (4 total)
          Comments about normality: 2 points

Question 7

(a)   Descriptive Statistics for each group: 2 points each (4 total)
          Comparison of means, and standard deviations: 2 points
(b)   Side-by-side boxplots: 4 points
          Center, spread, and shape: 3 points
          Outliers: 1 point
(c)   Normality plots: 2 points each (4 total)
          Comments about normality: 2 points

Question 8

(a)   Estimate of the difference on the log scale: 2 points
          Estimate of the ratio: 2 points
(b)   Null and alternative hypotheses: 3 points
          The test statistic: 2 points
          P-value: 2 points
          95% confidence interval for the difference on the log scale: 2 points
(c)   The 95% confidence interval on the original scale: 2 points
          Summary: 2 points

TOTAL= 92