# STAT 337 PRACTICE LAB EXAM

**Instructions**

**1.** This exam consists of 3 problems. For each problem, carry out the appropriate analysis using SPSS and give the answer in the space provided. When a text box is provided below the question, give brief, concise answers in the format provided by the box. All numerical values should be rounded to four digits. The exam is out of 91. The exam consists of 6 pages.

**2.** For each of the three problems you will have to download the appropriate data from the STAT 337 Labs web site (*Lab Exam Data* in *Exams and Tests* panel). Once you have saved the exam data files to your desktop, close all programs including your authentication. At this point you are only allowed to use SPSS.

**3.** You are allowed to use the *Statistical Sleuth* text in the exam. You are not to communicate with any other individual, in any manner, with the exception of the proctor.

**4.** Complete the following (please print):

Lab Section Number: _____ Name _____

I.D. Number: _____

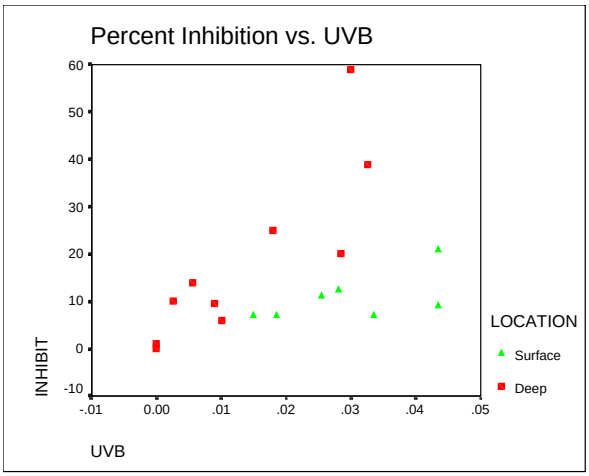
**Problems**

1. In order to estimate the degree to which oceanic phytoplankton production is inhibited by exposure to UVB, both near the ocean surface and below the surface, the researchers sampled from the ocean column at various depths at 17 locations around Antarctica during the austral spring of 1990. The related data is saved in the file *exam21.sav* which can be downloaded by clicking on the *exam21* link. The following is a description of the variables contained in the data file:

| **Variable Name** | **Description of Variable** |
|---|---|
| *inhibit* | Percent Inhibition, |
| *UVB* | UVB Exposure, |
| *deep* | 0 (at the ocean surface), 1 (in deep water). |

Use the data to answer the following questions:

(a) Obtain a scatterplot of percent inhibition vs. UVB exposure with different marking symbol for each location (ocean surface or deep water). Sketch the plot below (2 points):




(b) Use the plot to compare the percent inhibition at the surface and in deep water. Moreover, compare the strength of the relationship between percent inhibition and UVB at the surface and in deep water. (4 points)

Percent inhibition at the surface is smaller than the one in deep water. The relationship between percent inhibition and UVB is stronger for deep water than the one for the surface.

Consider the following regression model:

$$\mu\{inhibition \mid UVB\} = \beta_0 + \beta_1 * UVB + \beta_2 * deep + \beta_3 * deep * UVB.$$

Use the *Regression* tool in SPSS to perform the above regression. Then answer the following questions:

(c) What is the estimated regression line? (3 points)

$$\hat{\mu}\{inhibition \mid UVB\} = 2.967 + 258.936 * UVB - 1.467 * deep + 980.039 * deep * UVB.$$

(d) What percent of the variation in inhibition is explained by the three explanatory variables? (3 points)

72.9%

(e) Is linear regression significant? State this question as null and alternative hypotheses about the regression coefficients. Report the value of the test statistic and its P-value from the output, and give your conclusion. (5 points)

Hypotheses: $\beta_1 = \beta_2 = \beta_3 = 0.$

Value of the test statistic: F= 11.649, DF (numer)= 3, DF(denom)=13.

P-value: 0.001

Conclusion: At least one predictor is useful

(f) Which explanatory variable is significant given the other variables in the model? State the question as null and alternative hypotheses about the appropriate regression coefficient, report the value of the test statistic, the distribution of the test statistic under the null hypothesis, the p-value of the test, and give your conclusion. (3 points)

$H_0 : \beta_3 = 0,$ versus $H_a : \beta_3 \neq 0.$

t = 2.569, DF = 15, p-value = 0.023

the interaction term deep*UVB is significant given the other variables in the model.

(g) Use the fit obtained in part (a) to compare the percentage inhibition at the surface and in the deep. Comment briefly. (3 points)

At the surface: $\hat{\mu}\{inhibition \mid UVB\} = 2.967 + 258.936 * UVB.$
In the deep water: $\hat{\mu}\{inhibition \mid UVB\} = 1.5 + 1238.975 * UVB.$

(h) Examine carefully the normal probability plot of standardized residuals. Does it look that the assumption of normality may be violated? (3 points)

normality assumption slightly violated.

(i) Examine the plot of standardized residuals versus standardized predicted values. Describe the pattern of the residuals. Do the residuals appear to be randomly scattered about a horizontal line at zero? Is there any indication that the assumption of equal standard deviations may be violated? (3 points)

There is some tendency toward larger spread at higher fitted values, the assumption of equal variance may be violated. Given the relatively moderate sample size, the pattern is not prominent enough to make strong claims in this case.

(j) Obtain a 95% confidence interval for the mean percent inhibition at the surface when UVB exposure is 0.03. Moreover, obtain also a 95% prediction interval for the percent inhibition in this case. (3 points)

CI for the mean: (-5.16258, 18.86516)
PI : (-15.1299, 28.83245)

2. A consumer product evaluation group is interested in comparing the mean life (in minutes) of four types of batteries commonly used with children's toys. A random sample of each of the four battery types is selected and subjected to a continuous test. For each battery, a measurement is made of the time required for the energy output to fall below a predetermined acceptable level. The tests results are given in the file *exam12.sav*.

The following is a description of the variables in the data file:

| **Column** | **Name of Variable** | **Description of Variable** |
|---|---|---|
| 1 | TIME | Time required for the energy output to fall below a level, |
| 2 | TYPE | Battery type (an integer 1 to 4). |

Use SPSS to test the null hypothesis of no difference in time among the four groups. Then answer the following questions:

(i) What is the value of the appropriate test statistic, the p-value of the test and what the decision should be made about the null hypothesis? (5 points)

Value of the test statistic: F=11.516

P-value of the test: 0

Conclusion: Strong evidence of differences in mean time among the four groups.

(ii) Use the SPSS output to estimate the standard deviation σ, assuming this is the same for all groups (3 points)

$$\hat{\sigma} = \sqrt{10.103} = 3.17852$$

(iii) Do the data provide evidence that the average battery life for batteries 1 and 2 is significantly greater than the mean life for battery 3? Answer the question by carrying a test about an appropriate contrast. Estimate the contrast and report the p-value of the test. (6 points)

Hypotheses: $H_0 : \gamma = 0, H_a : \gamma > 0.$

Contrast γ: $\gamma = 0.5\mu_1 + 0.5\mu_2 + (-1)\mu_3 + 0\mu_4$

Estimate g of the contrast γ is g= 3.9333

p-value of the test: 0.004/2 = 0.002.

Conclusion: Strong evidence that the average battery life for batteries 1 and 2 is greater than the mean life for battery 3.

(iv) Compute a 95% confidence interval for the contrast specified in part (iii). (3 points)

From SPSS output, g = 3.9333, SE = 1.2675, DF=36-4=32. Thus

$3.9333 \pm t_{32}(0.975)\ SE = 3.9333 \pm 2.042\ (1.2675) = (1.345065, 6.52153)$

(v) Consider a three-mean model (the batteries from groups 1 and 2 have the same mean, possibly different from the mean of group 3 and the mean of group 4). What is the residual sum of squares for the model? (3 points)

Sum of squares for the three-mean model = 330.813 (DF=33)

(vi) Does the four-mean model provide a significantly better fit than the three-mean model considered in part (vii)? Answer the question, by calculating the value of the appropriate statistic and estimating its p-value. (5 points)

Value of the test statistic: Sum of squares for the 4-mean model = 323.289 (DF=32)

$$F = \frac{(330.813 - 323.289)/(33-32)}{323.289/32} = 0.7452$$

F distribution with DF(numerator)=1, DF(denominator)=32.

P-value: larger than 0.10

Conclusion: It does not provide a significantly better fit.

3. In order to evaluate the influence of a chamber on electrical resistance of printed wiring boards, the electrical resistance of each of 24 randomly selected boards is measured inside and outside the test chamber (in teraohms). The data are available in the file *exam13.sav* in two different formats. The choice of the format depends on the type of a test you are going to apply to the data.

The following is a description of the variables in the data file:

| Column | Name of Variable | Description of Variable |
|---|---|---|
| 1 | OUTSIDE | Electrical resistance outside test chamber |
| 2 | INSIDE | Electrical resistance inside test chamber |

Use the data to answer the following questions:

(a) Does the data provide any evidence that the test chamber influences the electrical resistance of the boards? Answer the question by carrying out an appropriate t-test. Providing the p-value of the test and state your conclusion. (5 points)

Name of the t-test in SPSS you used: Paired-Samples T Test

Hypotheses: $H_0 : \mu_D = 0, H_a : \mu_D \neq 0.$ Here D = OUTSIDE-INSIDE

P-value of the test: 0.111

Conclusion: No evidence that the test chamber influences electrical resistance

(b) Give a 95% confidence interval for the mean difference in the electrical resistance of boards inside and outside the test chamber. Is the interval consistent with the outcome of the test in (a)? Explain briefly. (4 points)

[-0.9737, 0.1071], as the interval contains 0, the result is consistent with the outcome of the test.

(c) What are the assumptions necessary for the test to be applied? Are the assumptions satisfied in this case? Answer the question by referring to the appropriate plot in the output. (5 points)

Assumptions: D follows a normal distribution, observations in D are independent

Which diagnostic plots in SPSS you used: Q-Q Plot applied to the differences

Which assumptions violated (if any): Normality clearly violated, there are systematic and substantial departures of the points in the Q-Q plot from a straight-line pattern.

(d) Consider appropriate transformation of the data. From your plot obtained in (c), which transformations appear to be worthy of pursuing? Apply the transformation to the data and describe the pattern in the new plot. Does it look that the transformation was useful? (5 points)

Type of transformation (log, square,..etc): Log transformation

Transformation applied to (specify variable/variables): each of OUTSIDE, INSIDE

Which diagnostic plot used: Q-Q plot applied to the differences

Pattern in the plot: All points very close to a straight-line pattern

(e) Carry out the test for the transformed data. Specify the null and alternative hypotheses. What is the p-value of the test that the test chamber affects the electrical resistance of the boards, on the average? (6 points)

Hypotheses: Median(INSIDE)=Median(OUTSIDE)

P-value of the test: t= -4.136, p-value =0.

Conclusion: Strong evidence that the test chamber affects electrical resistance.

(f) What is the estimate of a multiplicative effect of the test chamber on electrical resistance of the boards? Use to the output obtained in (e) to obtain the estimate. (4 points)

As the average difference LN(OUTSIDE)-LN(INSIDE) is estimated as –0.5195, thus

$$LN\left[\frac{OUSIDE}{INSIDE}\right] = -0.5195. \text{ Hence } \frac{OUTSIDE}{INSIDE} = \exp(-0.5195).$$

$$INSIDE = 1.68 * OUTSIDE.$$

(g) Carry the appropriate non-parametric test (you can choose any of the two discussed in the class) to test the hypothesis that the chamber affects the electrical resistance of the boards. State the p-value of the test. (5 points)

Name of the t-test in SPSS you used: 2-Related Samples

P-value of the test: Two-sided: Wilcoxon Signed Ranks (.014), Sign Test (.007).

Conclusion: Strong evidence that the test chamber affects electrical resistance.