CS555 - March 23rd, 2023

Zuowen Tang

# Assignment 4 Report

**Question 1/2.** The data in this document is from 3 groups of students (math, chemistry, and physics) on an IQ related test. Save the data, and read the data into R. Use this data to address the following questions (14 points)

(1) How many students are in each group? Summarize the data relating to both test score and age

by the student group (separately). Use appropriate numerical and/or graphical summaries. (3

points)

Text

Description automatically generated

Chart, histogram

Description automatically generated

Chart, histogram

Description automatically generated

(2) Do the test scores vary by student group? Perform a one way ANOVA using the aov or Anova

function in R to assess. Use a significance level of α=0.05. Summarize the results using the 5-

step procedure. If the results of the overall model are significant, perform the appropriate

pairwise comparisons using Tukey’s procedure to adjust for multiple comparisons and

summarize these results. (7 points)

Text, letter

Description automatically generated

Text

Description automatically generated

**5-step procedure:**

Step 1.

Н0: The means scores of all groups are equal.

Ha: The mean scores of all groups are different.

Step 2.

Shown as the above results of ANOVA test that p-value is 3.5e-08.

Step 3 & Step 4

P-value is less than 0.05, so we want to reject the H0.

Step 5.

There is significant evidence show that the mean scores of all groups are different.

(3) Create an appropriate number of dummy variables for student group and re-run the one-way

ANOVA using the lm function with the newly created dummy variables. Set chemistry students

as the reference group. Confirm the results are the same (specifically point out test statistics, p-

values, etc. that show the results are equivalent). What is the interpretation of the beta

estimates from the regression model? (4 points)

Text

Description automatically generated

By suing lm() function, we can see the **p-value is 3.496e-08** and the **F-statistic is 26.57**, which are the same as the ANOVA function (**p-value = 3.5e-08, F-statistic = 26.57**.)

**Interpretation of the beta estimates:**

The estimate std. of math students is -8.667, which means that their mean of IQ is 8.667 less than chemistry students. While the std of physics students is -12.133, means their mean of IQ is 12.133 less than chemistry students.

**Question 2/2.** In the United States, there is a strong relationship between education and smoking:

well-educated people are less likely to smoke. Does a similar relationship hold in France? To find out, researchers recorded the level of education and smoking status of a random sample of 459 French men aged 20 to 60 years. The two-way table below displays the data. Is there convincing evidence of an association between smoking status and educational level among French men aged 20 to 60 years? (Identify the right type of chi-square test, explain why, and check the conditions. State an appropriate pair of hypotheses to test in this setting and carry out the test by calculating the test statistic, degree of freedom, and P-value. What conclusion would you draw at α= 0.05?) (6 points)

H0: There is an association between smoking status and educational level.

Ha: There isn’t an association between smoking status and educational level.

Based on the table, the degree of freedom is (r-1)(c-1) = (3-1)(3-1) = 4

While degree of freedom = 4 and α= 0.05, the critical value is 9.488.

So reject H0 if the X^2 > 9.488.

Text

Description automatically generated

In the test we can see that X^2 < 9.488, which means we cannot reject the H0. Hence there is an association between smoking status and educational level.

**Extra Credit.** The endorphins released by the brain act as natural painkillers. For example, a study monitored endorphin activity and pain thresholds in pregnant rats during the days before they gave birth. The data showed an increase in pain threshold as the pregnancy progressed. The change was gradual until 1 or 2 days before birth, at which point there was an abrupt increase in pain threshold. Apparently, a natural painkilling mechanism was preparing the animals for the stress of giving birth. The following data represent pain-threshold scores. Do these data indicate a significant change in pain threshold? Use a repeated-measures ANOVA with α =.01. (4 points)

H0: Data does not indicate a significant change in pain threshold.

Ha: Data does indicate a significant change in pain threshold.

n=20, total df = (20-1) = 19, between treatment df = (4-1) = 3, within treatment df = sum(4) = 16

While df = (3,16), and α =.01, F-value is 5.292.

So we reject H0 if F-value is greater than 5.292.

Text, letter

Description automatically generated

We can see that F-value is 152.82 which is greater than 5.292, hence we have to reject H0, which means that the data does indicate a significant change in pain threshold.