Psychology 407

Assignment B

A social psychologist was interested in problem solving carried out co-operatively by small groups of individuals. The theory upon which these experiments were based suggested that within a particular range of possible group sizes, the relationship between group size and average time to solution for a particular kind of problem would be linear and negative: the larger the group, the less time on the average should it take for the problem to be solved. To check on this theory, the psychologist decided to form a set of experimental groups, ranging in size from groups consisting of 1 individual each, five groups consisting of 2 individuals, five groups each of 3, and so on until there were six different and nonoverlapping sets of five groups each, the last consisting of five groups each of size 6. Each individual subject participated in one and only one group in the study.

Furthermore, five different problems were available for use with these groups. These problems had been scaled in difficulty, from 5 for the most difficult to 1 for the least difficult. To see if scaling correlated with the time it took the groups to solve these problems, and also as a way to reduce error variance, the experimenter decided to give all five problems within each set of groups of the same size, with each group receiving one problem assigned at random. Thus, every combination of problem difficulty and size was represented by exactly one group.

Each experimental group then solved its assigned problem, and the time taken for them to do so was noted; this time to completion was used as the dependent variable.

The data that resulted from this experiment are as follows, where Y = time to completion, X_1 = group size, X_2 = difficulty:

Y	X_1	X_2
26	1	1
28	1	2
30	1	3
29	1	4
32	1	5
22	2	1
23	2	2
24	2	3
22	2	4
22	2	5
23	3	1
24	3	2
24	3	3
26	3	4
27	3	5
18	4	1
20	4	2
20	4	3
21	4	4
19	4	5
21	5	1
25	5	2
23	5	3
22	5	4
25	5	5
16	6	1
18	6	2
16	6	3
18	6	4
20	6	5

In answering the question to follow, you should find the following summary information helpful: (all summations are assumed to be from i = 1 to 30, where n = 30):

$\sum Y_i = 684$;
$\sum Y_i^2 = 16058$;
$\sum X_{i1} = 105$;
$\sum X_{i1}^2 = 455$;
$\sum X_{i2} = 90$;
$\sum X_{i2}^2 = 330$;
$\sum Y_i X_{i1} = 2243$;
$\sum Y_i X_{i2} = 2090$;
$\sum X_{i1} X_{i2} = 315$;

In computing the various expression you will need in answering the questions, make sue you carry a lot of decimal places until you construct the final results.

a) Find **b** and the necessary intermediate results, i.e., $(\mathbf{X}'\mathbf{X})^{-1}$ and $\mathbf{X}'\mathbf{Y}$.

b) Construct the ANOVA table and test $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ simultaneously.

c) Construct 95% confidence intervals for β_0, β_1, and β_2, and test the hypotheses that $H_0 : \beta_0 = 0$; $H_0 : \beta_1 = 0$; $H_0 : \beta_2 = 0$.

d) Construct a 95% confidence interval for the true mean value of Y when $X_1 = 4$ and $X_2 = 3$.

e) Construct a 95% prediction interval on Y for a new observation when $X_1 = 4$ and $X_2 = 3$.

f) Obtain the standardized regression coefficients corresponding to β_1 and β_2.

g) Find the correlations between Y and X_1; Y and X_2; and X_1 and X_2.

h) Find the correlations of X_1 with $\mathbf{Xb}$ and of X_2 with $\mathbf{Xb}$. Is there a nice relationship to what you found in (g)? Any explanation?

i) Calculate R^2 and adjusted R^2.

j) What is the relation between R^2 and the sum of the squared correlations between Y and X_1 and between Y and X_2? Any explanation?

k) If one included a third term in the regression equation of the form, $X_3 = X_1X_2$, what particular substantive concerns would be addressed?

l) Would it be possible to find a Sum of Squares for Pure Error given the way the data were collected? If not, how could you redesign the experiment so a Pure Error term could be constructed?