Psychology 407
Assignment D

Suppose that an experimenter is interested in "level of aspiration" as the dependent variable in an experiment. An experimental task has been developed consisting of a difficult game apparently involving motor skill, yielding a numerical score that can be attached to a person's performance. But this appearance is deceptive: unknown to the subject, the game is actually under the control of the experimenter, so that each subject is made to obtain exactly the same score. After a fixed number of trials, during which the subject unknowingly receives the preassigned score, the individual is asked to predict what the score will be on the next group of trials. However, before this prediction, the subject is given "information" about how the score compares with some fictitious norm group. In one experimental condition, the subject is told that the first performance is above average for the norm group; in the second that it is average; and in the third that it is below average. There are thus three possible experimental "standings" that might be given to any subject. (Of course, after the experiment, each subject is full informed of this little ruse by the experimenter.)

The dependent score value Y is based on the report the subject makes about anticipated performance in the next group of trials. Because each subject has obtained the same score, this anticipated score on the next set of trials is treated as equivalent to a level of aspiration that the subject has set. Each subject is tested privately, and no communication is allowed between subjects until the entire experiment is completed. Each of the three groups contains 20 randomly assigned subjects.

In addition to the dependent measure, Y, prior to the experiment each subject had been tested on a game very similar to that used in the experiment proper, and a "skill score", X_1, obtained for each. The data that resulted from this experiment can be represented in the following form:

above average		average		below average	
Y	X_1	Y	X_1	Y	X_1
52	44	28	38	15	23
48	47	35	26	14	17
43	30	34	36	23	31
50	38	32	30	21	25
43	40	34	36	14	27
44	45	27	23	20	35
46	36	31	45	21	25
46	41	27	28	16	28
43	40	29	34	20	30
49	43	25	37	14	37
38	48	43	40	23	32
42	24	34	36	25	32
42	39	33	41	18	34
35	36	42	29	26	48
33	46	41	39	18	39
38	33	37	37	26	38
39	38	37	47	20	30
34	26	40	34	19	24
33	41	36	47	22	31
34	36	35	31	17	19

In addition to Y and X_1, define three "dummy" variables, X_2, X_3, and X_4: $X_j = 1$, if the subject belongs to group $j-1$; $= 0$, otherwise. The SYSTAT output that is attached gives the basic statistics in addition to information on fitting a variety of models. (In giving the basic statistics, Groups 1, 2, and 3 are above average, average, and below average, respectively.) The models that are fitted are given as (a) through (f) below:

Model (a):

$$Y = \beta_0 + \beta_1(X_1X_2) + \beta_2(X_1X_3) + \beta_3(X_1X_4) + \epsilon$$

Model (b):

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

Model (c):

$$Y = \beta_0 + \beta_1 X_2 + \beta_2 X_3 + \epsilon$$

Model (d):

$$Y = \beta_0 + \beta_1(X_1X_2) + \beta_2(X_1X_3) + \beta_3(X_1X_4) + \beta_4 X_2 + \beta_5 X_3 + \epsilon$$

Model (e):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

Model (f):

$$X_1 = \beta_0 + \beta_1 X_2 + \beta_2 X_3 + \epsilon$$

Questions:

i) Model (c) performs a "one-way analysis of variance" on the dependent measure Y in relation to the 3 groups. Show explicitly the relation between the means on Y within each of the three groups (given in the basic statistics) and the estimated means on Y for all the various combinations of values that X_2 and X_3 can take on. What does the analysis-of-variance table say about the "effectiveness" of the 3 treatments? Why isn't an X_4 term included in model (c)?

ii) Looking at model (f), carry out a similar interpretation as in (c). Are the results surprising? Why?

iii) Plot the regression lines of Y on X_1 implied by model (d) for the three separate groups. Superimpose on this plot the regressions of Y on X_1 implied by model (e) for the three separate groups. Carry out a test of model (d) versus model (e) and interpret.

iv) Plot the regression lines of Y on X_1 implied by model (e) and model (b).

Carry out a test of model (e) versus model (b) and interpret. This is called "analysis-of-covariance", and supposedly is a way of assessing the effectiveness of the three treatments. How is it different than what was done in (i)?

(It may help to interpret what was done in (i) as a comparison of model (c) against a restricted model, $Y = \beta_0 + \epsilon$.)

Analysis of covariance is based on an assumption that model (e) is the "Full Model". How does this relate to what was done in (iii)?

v) Suppose I have some given value on X_1, say P. Using model (e), what are the expected values on Y for the three separate groups. Suppose I have a second given value on X_1, say Q. What are the expected values on Y for the three separate groups, again using model (e), and what are the relationships between the two sets of expected values.

Now, do the same for model (d) and comment on the differences from using model (e).

Using these interpretations, why is it argued that one cannot compare the effectiveness of treatments merely by comparing model (d) and (a) (when model (e) cannot be assumed correct)?

Also, why is it argued that we can actually "control" for the effect of X_1 in assessing treatment effectiveness when model (e) is "true" but not if model (d) is "true"?

vi) Carry out a test of model (d) versus (b). What is this a test of any how does it differ from a comparison of model (d) versus model (a) and of model (e) versus model (b)?

	Y	X1	X2	X3	X4
N of cases	60	60	60	60	60
Minimum	14.000	17.000	0.0	0.0	0.0
Maximum	52.000	48.000	1.000	1.000	1.000
Mean	31.733	34.833	0.333	0.333	0.333
Standard Dev	10.457	7.549	0.475	0.475	0.475

The following results are for:
GROUP = 1.000

	Y	X1	X2	X3	X4
N of cases	20	20	20	20	20
Minimum	33.000	24.000	1.000	0.0	0.0
Maximum	52.000	48.000	1.000	0.0	0.0
Mean	41.600	38.550	1.000	0.0	0.0
Standard Dev	5.915	6.557	0.0	0.0	0.0

The following results are for:
GROUP = 2.000

	Y	X1	X2	X3	X4
N of cases	20	20	20	20	20
Minimum	25.000	23.000	0.0	1.000	0.0
Maximum	43.000	47.000	0.0	1.000	0.0
Mean	34.000	35.700	0.0	1.000	0.0
Standard Dev	5.171	6.602	0.0	0.0	0.0

The following results are for:
GROUP = 3.000

	Y	X1	X2	X3	X4
N of cases	20	20	20	20	20
Minimum	14.000	17.000	0.0	0.0	1.000
Maximum	26.000	48.000	0.0	0.0	1.000
Mean	19.600	30.250	0.0	0.0	1.000
Standard Dev	3.872	7.276	0.0	0.0	0.0

Dep Var: Y N: 60 Multiple R: 0.874 Squared multiple R: 0.764

Adjusted squared multiple R: 0.751 Standard error of estimate: 5.220

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	21.265	3.514	0.0	.	6.051	0.000
X1*X2	0.519	0.094	0.928	0.151	5.545	0.000
X1*X3	0.351	0.101	0.584	0.151	3.490	0.001
X1*X4	-0.038	0.116	-0.054	0.152	-0.324	0.747

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	4926.017	3	1642.006	60.268	0.000
Residual	1525.716	56	27.245		

Dep Var: Y N: 60 Multiple R: 0.534 Squared multiple R: 0.285

Adjusted squared multiple R: 0.273 Standard error of estimate: 8.918

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	5.975	5.480	0.0	.	1.090	0.280
X1	0.739	0.154	0.534	1.000	4.808	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	1838.561	1	1838.561	23.116	0.000
Residual	4613.173	58	79.537		

Dep Var: Y N: 60 Multiple R: 0.880 Squared multiple R: 0.774

Adjusted squared multiple R: 0.766 Standard error of estimate: 5.057

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	19.600	1.131	0.0	.	17.334	0.000
X2	22.000	1.599	1.000	0.750	13.758	0.000
X3	14.400	1.599	0.655	0.750	9.005	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	4994.133	2	2497.067	97.649	0.000
Residual	1457.600	57	25.572		

Dep Var: Y N: 60 Multiple R: 0.892 Squared multiple R: 0.796

Adjusted squared multiple R: 0.777 Standard error of estimate: 4.938

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	11.209	4.837	0.0	.	2.317	0.024
X2	22.162	8.305	1.007	0.027	2.669	0.010
X3	16.412	7.883	0.746	0.029	2.082	0.042
X1*X2	0.213	0.173	0.382	0.040	1.236	0.222
X1*X3	0.179	0.172	0.297	0.046	1.042	0.302
X1*X4	0.277	0.156	0.397	0.076	1.782	0.080

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	5135.207	5	1027.041	42.126	0.000
Residual	1316.526	54	24.380		

Dep Var: Y N: 60 Multiple R: 0.892 Squared multiple R: 0.795

Adjusted squared multiple R: 0.784 Standard error of estimate: 4.857

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	12.737	3.053	0.0	.	4.171	0.000
X1	0.227	0.094	0.164	0.788	2.405	0.020
X2	20.117	1.724	0.915	0.595	11.669	0.000
X3	13.164	1.620	0.598	0.674	8.127	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	5130.571	3	1710.190	72.490	0.000
Residual	1321.163	56	23.592		

Dep Var: Y N: 60 Multiple R: 0.880 Squared multiple R: 0.774

Adjusted squared multiple R: 0.766 Standard error of estimate: 5.057

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	19.600	1.131	0.0	.	17.334	0.000
X2	22.000	1.599	1.000	0.750	13.758	0.000
X3	14.400	1.599	0.655	0.750	9.005	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	4994.133	2	2497.067	97.649	0.000
Residual	1457.600	57	25.572		

```
Variables in the SYSTAT Rectangular file are:
 Y            X1            X2            X3            X4

Dep Var: X1 N: 60 Multiple R:  0.460 Squared multiple R:  0.212

Adjusted squared multiple R:  0.184    Standard error of estimate:        6.820

Effect          Coefficient    Std Error     Std Coef Tolerance     t    P(2 Tail)

CONSTANT             30.250        1.525        0.0         .       19.837     0.000
X2                    8.300        2.157        0.523     0.750      3.849     0.000
X3                    5.450        2.157        0.343     0.750      2.527     0.014

                              Analysis of Variance

Source              Sum-of-Squares   DF  Mean-Square     F-Ratio        P

Regression                711.433     2      355.717       7.649       0.001
Residual                 2650.900    57       46.507
--------------------------------------------------------------------------------
```