

TEST 1 STAT 572

Spring 2001

1. An article in the Journal of Socio-Economics reports a study done using data from a 45-year period. The variables measured were:

- y = Suicide rate
- x1 = Unemployment rate
- x2 = Percentage of females in labor force
- x3 = Divorce rate
- x4 = Logarithm of Gross National Product
- x5 = Annual percent change in GNP

The model containing only main effects yielded the following results:

Variable	Estimate	Std error
Intercept	.002	
x1	.0204	.002
x2	.0231	.02
x3	.0765	>.10
x4	.2760	>.20
x5	.0018	>.10

$$R^2 = .45$$

a) Test the overall utility of the model and interpret.(10 p)

b) Informally (without tables) test the contribution of each regressor and interpret the effect of the significant ones. (5 p)

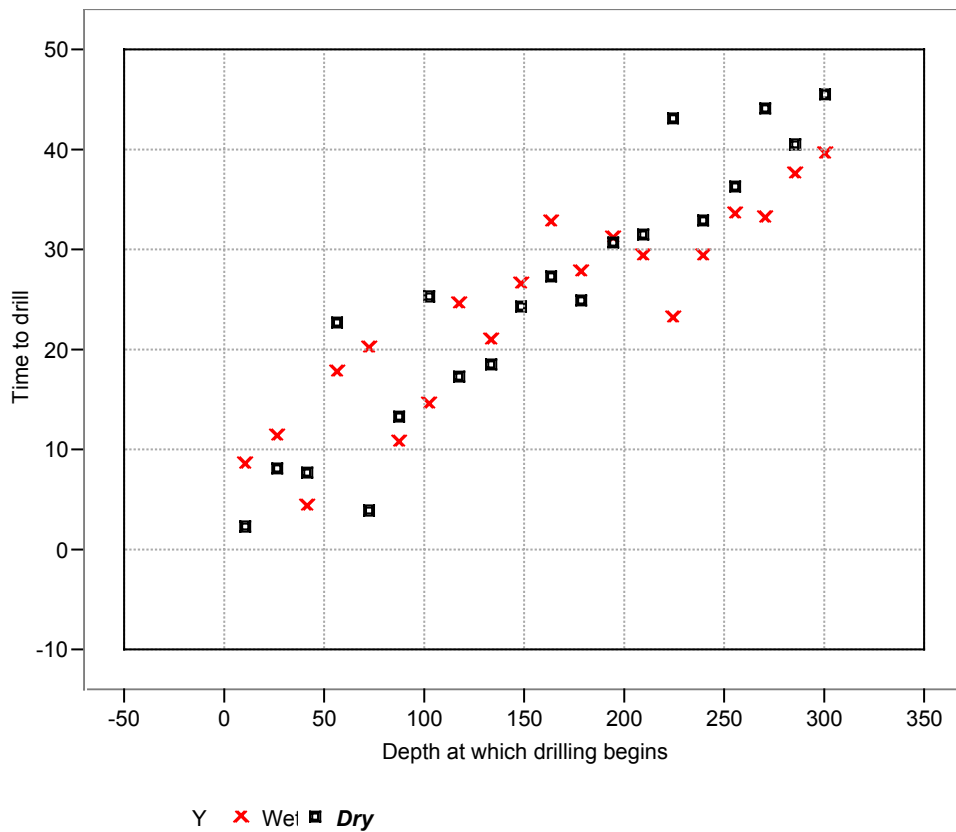
c) Someone says that one of the most surprising findings of the study is that the divorce rate has no effect on the suicide rate. Decide whether you agree or not and justify your answer. (5 p)

d) When a new regressor is added, R^2 increases to .56. Would you keep this term in the model? [Hint: Carry out a test] (5 p)

e) Another potential regressor has a simple correlation with suicide rate of .89 with a corresponding p-value of .0001. Is it clear that this regressor should be added to the model? Explain.

2. Two methods for drilling rock are being tested, one a dry method and the other a wet method. The response is the time (in minutes) needed to drill a hole of 5 feet of depth starting at different depths in the rock. The data appear below along with a scatterplot.

Depth at which drilling began	Wet (time in min)	Dry (time in min)
10	8.7	2.4
26	11.5	8.2
41	4.6	7.8
56	17.9	22.8
72	20.3	3.9
87	10.9	13.4
102	14.8	25.3
117	24.7	17.3
133	21.1	18.6
148	26.7	24.3
163	33	27.3
178	27.9	25
194	31.4	30.8
209	29.5	31.5
224	23.4	43.1
239	29.6	32.9
255	33.8	36.3
270	33.4	44.1
285	37.8	40.5
300	39.8	45.6



You are given the task of analyzing the data and writing a report for the mining company. After trying several models, you select one, whose results appear below.

Response Time to drill (in minutes)

RSquare	0.854384
RSquare Adj	0.842249
Root Mean Square Error	4.64766
Mean of Response	24.5475
Observations (or Sum Wgts)	40

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	4562.6329	1520.88	70.4086
Error	36	777.6269	21.60	Prob > F
C. Total	39	5340.2598		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	5.9173783	1.492138	3.97	0.0003
Depth at which drilling begins	0.1198464	0.008354	14.35	<.0001
Method[Dry]	-2.379993	1.492138	-1.60	0.1195
Depth at which drilling begins*Method[Dry]	0.0185751	0.008354	2.22	0.0325

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Depth at which drilling begins	1	1	4445.5395	205.8049	<.0001
Method	1	1	54.9543	2.5441	0.1195
Depth at which drilling begins*Method	1	1	106.7911	4.9439	0.0325

a) Superimpose a graph of the model fitted on the scatterplot above. Show details. (10 p)

b) Write a short report explaining to the mining company executives the most important aspects of your analysis. Assume no statistical expertise on the part of the executives. (10 p) [Hint: More than three paragraphs is too long!]

c) There are 40 observations in the data set, but they tell you that they only made two holes, one for each method and the measurements were taken as the drill went deeper into the rock. Comment on the consequences of this fact on the data and the results. [Hint: Think on the assumptions of linear regression]