TEST 1
STAT 572
Spring 1995

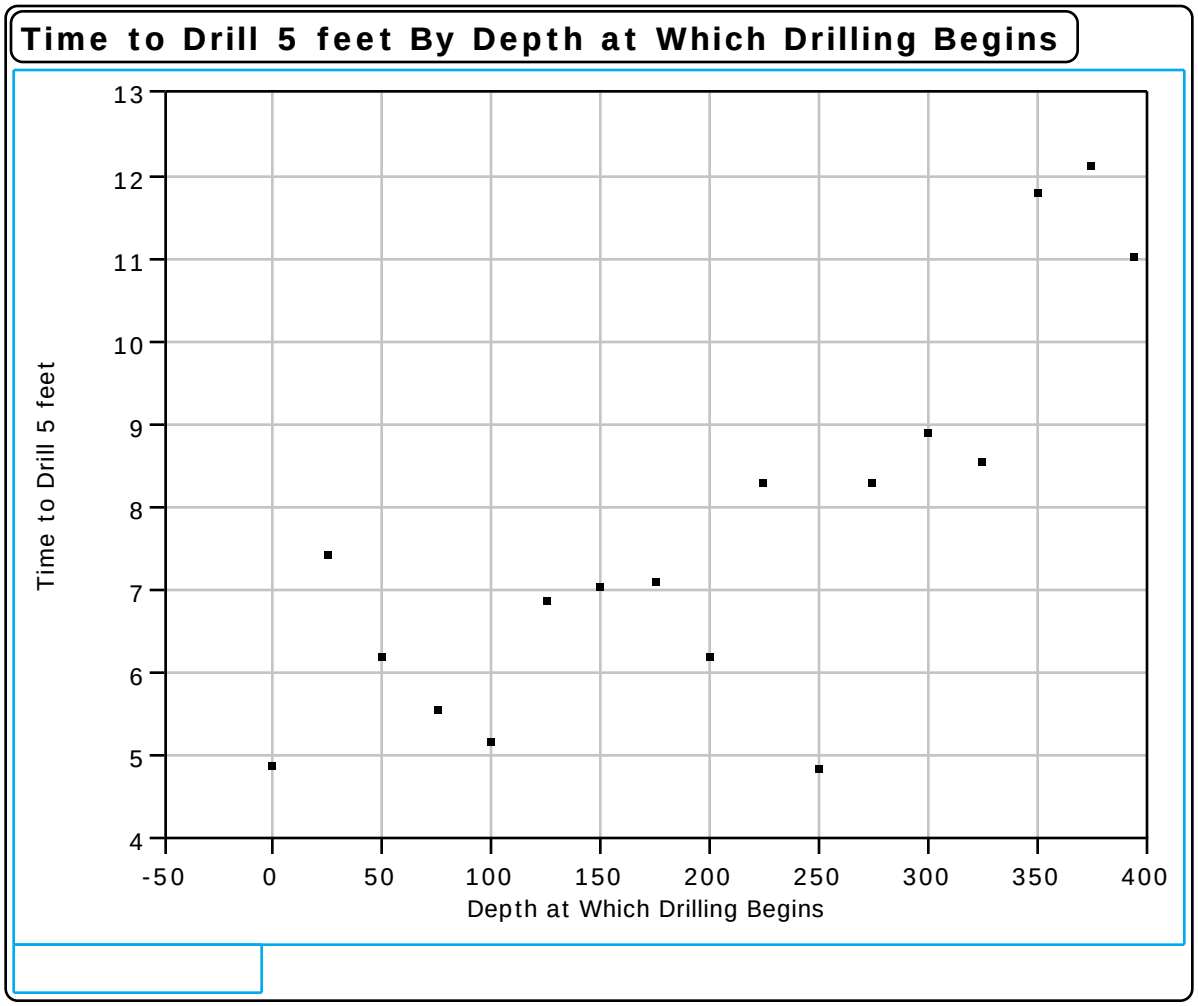
1. An experiment was conducted to investigate the relationship between the time it takes to drill a distance of 5 feet in rock and the depth at which the drilling takes place. The data is shown below.

Depth (ft)	Time (minutes)
7.	8.
9.	10.
11.	12.
13.	14.
15.	16.
17.	18.
19.	20.
21.	22.
23.	24.
25.	26.
27.	28.
29.	30.
31.	32.
33.	34.
35.	36.
37.	38.
39.	40.
Mean	
St. dev	

A scatterplot of the data appears in the following page. The correlation coefficient is r = 0.794.

a) Obtain the LS line and draw it on the scatterplot. (10)

b) Interpret **in the context of the problem** the value of the slope. (5)



c). Estimate the average time that it would take to drill 5 ft at a depth of 400 ft (5)

d). Construct and interpret an appropriate 90% interval for your estimate in c) (10)

e). Determine if the observation at 250 ft can be considered an outlier. (5)

[Hint: $(n-p-1)s^2_{(i)} = (n-p)s^2 - \frac{e_i^2}{1-h_i}$]

f) You are asked to estimate the average time that it would take to drill 20 ft at a depth of 350 ft. Give such an estimate and discuss your assumptions. (5)

g) Suppose that we have two observations per depth and that the residual sum of squares is SSE = 85.36, and the pure error sum of squares is SSPE = 34.07. Test for lack of fit in the model.

2. Real estate appraisers rely heavily on MLR to determine the value of properties. Typically, the sale price of a house is modeled as a function of variables like size, condition, and location. The following is real data obtained from 10 recent home sales.

Sale Price (thousands)	Home size (100's of sq.ft.)	Condition (1=poor, 10=best)
64.	65.	66.
67.	68.	69.
70.	71.	72.
73.	74.	75.
76.	77.	78.
79.	80.	81.
82.	83.	84.
85.	86.	87.
88.	89.	90.
91.	92.	93.

The result of fitting a MLR model appears below

Response: Sale Price

Summary of Fit

RSquare	0.990123
RSquare Adj	0.987301
Root Mean Square Error	1.080545
Mean of Response	51.73
Observations (or Sum Wgts)	10

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	9.7822706	1.630481	6.00	0.0005
Home Size	1.8709353	0.076174	24.56	0.0000
Condition	1.2781408	0.144400	8.85	0.0000

Effect Test

Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Home Size	1	1	704.36011	603.2660	0.0000
Condition	1	1	91.47595	78.3468	0.0000

Additional results from the analysis are:

Correlations

Variable	Sale Price	Home Size	Condition
Sale Price	1.0000	0.9379	0.3727
Home Size	0.9379	1.0000	0.0433
Condition	0.3727	0.0433	1.0000

Partial Corr

Variable	Sale Price	Home Size	Condition
Sale Price	•	0.9942	0.9581
Home Size	0.9942	•	-0.9513
Condition	0.9581	-0.9513	•

partialled with respect to all other variables

a) Comment on the results of the fit for the full model. (10)

b) Determine the p-value of the simple correlation between sale price and condition. Compare this p-value with the one from the full model and explain the difference **in practical terms.** (10)

c) Test the significance of the partial correlation of sale price and condition adjusted by home size and interpret the result. (5)

d) Explain **as detailed as possible** how to construct a prediction interval for the sale price of a house with 2000 sq. ft. and a condition rating of 5. (5)

e) Explain what the interaction between size and condition would mean **in the context of this problem.** [Hint: Pictures can help] (5)

3. Determine the truthfulness or falseness of the following propositions. **Justify your answers.** (5 each)

T F A high correlation is an indication of a causal association

T F Weighted least squares is useful when the assumption of homoscedasticity is violated

T F If two variables are interacting, then they must be correlated.