TEST 1
STAT 572
Spring 1996

1. The management of a chain of package delivery stores would like to develop a model for predicting the weekly sales (in thousands of dollars) for individual stores. A sample of 20 stores was selected and the number of customers and the amount of sales were recorded. The data appear below.

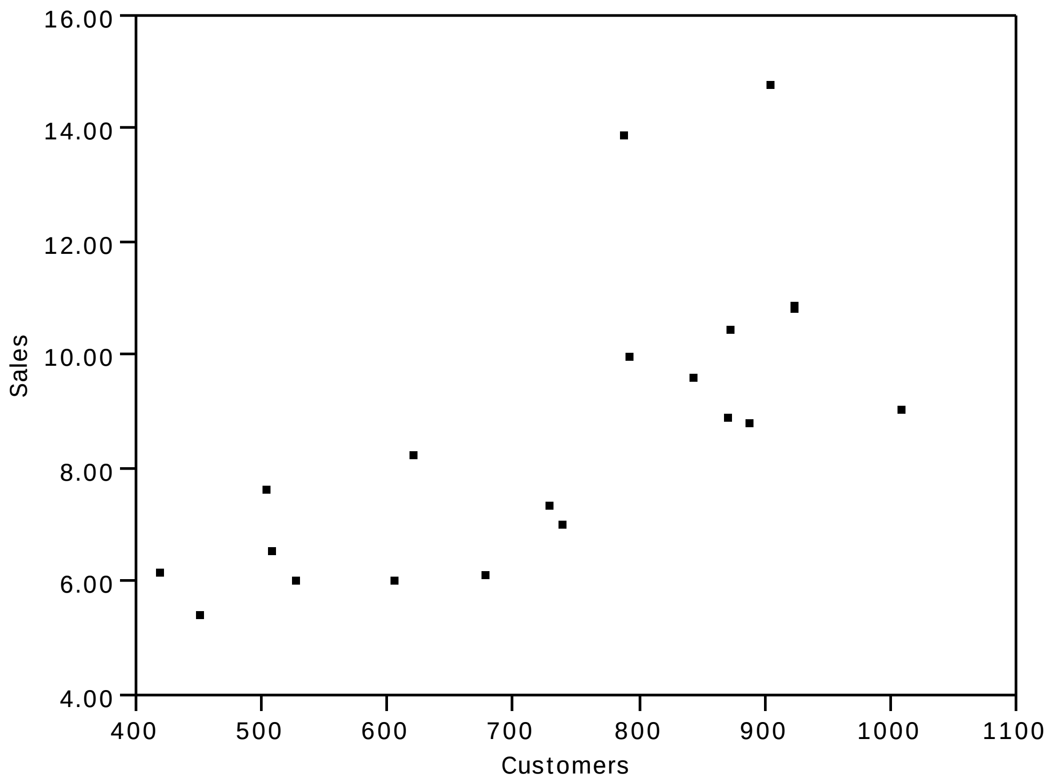
Store	Customers	Sales ($000)
1	907	14.81
2	926	10.89
3	506	7.63
4	741	7.04
5	789	13.9
6	889	8.83
7	874	10.48
8	510	6.56
9	529	6.05
10	420	6.18
11	679	6.15
12	872	8.93
13	924	10.83
14	607	6.03
15	452	5.43
16	729	7.36
17	794	10.01
18	844	9.63
19	1010	9.07
20	621	8.28

A scatterplot of the data along with the results of a SLR analysis appear in the following page.

a) Plot the SLR model on the graph. Show work. (5 p)

b) Is the model useful to explain sales? Be as thorough as possible. (10 p)

Sales By Customers



Linear Fit

Linear Fit

Sales = 1.2669 + 0.01017 Customers

Summary of Fit

RSquare	0.498083
RSquare Adj	0.470199
Root Mean Square Error	1.886997
Mean of Response	8.704817
Observations (or Sum Wgts)	20

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	1	63.60402	63.6040	17.8625
Error	18	64.09367	3.5608	**Prob>F**
C Total	19	127.69770		0.0005

Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	1.2668957	1.809746	0.70	0.4929
Customers	0.0101729	0.002407	4.23	0.0005

c) To investigate the predictive power of the model, you are asked to construct a confidence interval for the average amount of sales when 900 customers are serviced. Give the interval and interpret. [Hint: $\overline{X}$ = 731.15] (10 p)

d) One of the managers notices that the p-value for the intercept is fairly large and proposes to fit a no-intercept model. The slope of this model is 0.0118. Plot the line in the scatterplot. (5 p)

e) Which of the two models would you favor and why? (5 p)

2. The manager who gathered the data pointed out that the two observations with the highest sales were the only ones taken during december. It was decided to create an indicator variable to account for this effect (one for these two observations and zero otherwise). The result of the analysis appear below.

Response: Sales

Summary of Fit

RSquare	0.862151
RSquare Adj	0.845934
Root Mean Square Error	1.017579
Mean of Response	8.704817
Observations (or Sum Wgts)	20

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.1947963	0.985697	2.23	0.0398
Customers	0.0081909	0.001331	6.15	<.0001
Dummy	5.2124368	0.777904	6.70	<.0001

Effect Test

Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Customers	1	1	39.198541	37.8559	<.0001
Dummy	1	1	46.490721	44.8983	<.0001

a) Compare this model with the previous one. Explain why this model looks much better than the previous? (10 p)

b) Interpret the value of the coefficient for the dummy variable and its significance. (5 p)

c) Explain how this model could be used to predict sales for any month of the year.
(5 p)

d) Describe the method to obtain prediction confidence intervals for this model. Be as thorough as possible. (5 p)

3. An experiment was run to investigate the effect of particle size on the strength of particleboard. Six different particle sizes were tested and the strength of the board recorded. Each particle size was observed three times, yielding a total of 18 points. Partial results of a SLR analysis were:

Source	**df**	**SS**	**MS**
Model		150	
Error			
Total		540	

a) Complete the ANOVA table and test the usefulness of the model. (5 p)

b) Using the fact that the pure error sum of squares is 125, carry out a test for lack of fit and interpret the result. (5 p)

4. Determine the truthfulness or falseness of the following statements. Justify your answers. (5 p each)

T F The mean square error is a good way to compare a no-intercept against an intercept model

Suspose that $r_{y,\, x_i}$ is the correlation between y and x_i, then

T F $R^2_{1,2} \le r^2_{y,\, x_1} + r^2_{y,\, x_2}$. Where $R^2_{1,2}$ is the coefficient of multiple determination of y on x_1 and x_2.

T F Every time a new regressor is added to a model, the mean square error decreases.