

# TEST 1 STAT 572

## Spring 1999

1. A realtor wants to investigate the effect of different factors on the price of houses in a certain area. For that purpose she collects data on 108 properties recently sold. After testing different models, the following is fitted. The variables involved are:

Price: Selling price (in thousands)

Sq Ft: Square footage

Style: Contemporary(A), Bilevel (B), and Ranch (C)

School district: Beard or Kent

The results of the fitting appear below.

Response: Price

Summary of Fit	
RSquare	0.783414
RSquare Adj	0.770548
Root Mean Square Error	12.66005
Mean of Response	97.99226
Observations (or Sum Wgts)	108

Lack of Fit

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	6.2120074	5.870312	1.06	0.2925
Sq Ft	0.0510864	0.003403	15.01	<.0001
Style[A-C]	-5.446871	8.17393	-0.67	0.5067
Style[B-C]	28.521112	7.82025	3.65	0.0004
School[Beard-Kent]	4.4489625	1.360616	3.27	0.0015
Sq Ft*Style[A-C]	0.0006926	0.004704	0.15	0.8833
Sq Ft*Style[B-C]	-0.012981	0.005009	-2.59	0.0110

Effect Test					
Source	Nparm	DF	Sum of Squares	F Ratio	Prob>F
Sq Ft	1	1	36110.733	225.3022	<.0001
Style	2	2	2212.109	6.9009	0.0016
School	1	1	1713.631	10.6917	0.0015
Sq Ft*Style	2	2	1395.500	4.3534	0.0154

a) Obtain the F-ratio and test the utility of the model

b) Comment on the contribution (p-values) of each of the variables in the model.

c) Interpret the value of the coefficients for Style. Order the styles in terms of their average price.

d) Explain the meaning of the interaction term in this model. Draw a plot to illustrate your point.

2. In the table below describe the effect of deleting one regressor from a MLR model.

Note: Mark and X in the corresponding column (D: decrease, I: increase, CT: can't tell) and then explain your answer.

Aspect	D	I	CT	Reason
$R^2$				
MSE				
The p-value for the model				
The p-value for one of the remaining regressors				
The estimated coefficient for one of the remaining regressors				

3. The ANOVA table for a MLR model with 48 observations is given below

<b>Source</b>	<b>Df</b>	<b>SS</b>	<b>MS</b>
Model	3	2346	
Error			
Total		4312	

a) Complete the missing terms of the table and compute  $R^2$

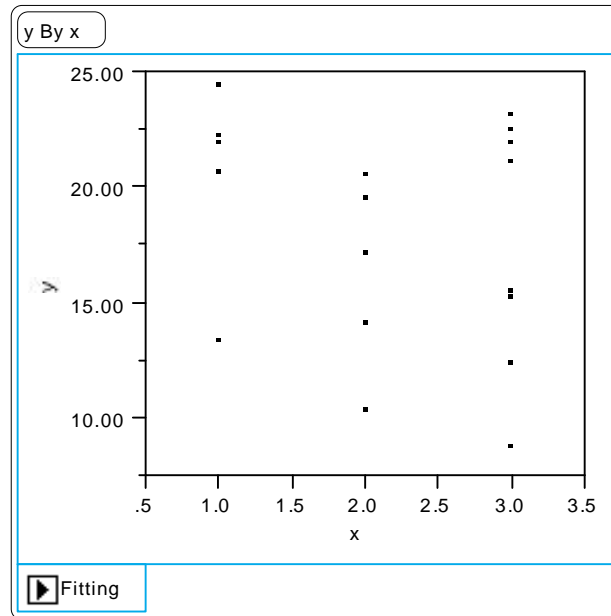
b) Test the utility of the model and give an approximate p-value.

c) A new regressor is added to the model and  $R^2$  increases by 0.09. Is the contribution of this regressor significant? Should it be kept in the model? [Hint: Do not guess]

d) Determine if the test in d) is a partial or a sequential test.

4. Consider the following data and the corresponding plot:

x	y
1	22.26
1	24.50
1	20.74
1	21.98
1	13.44
2	17.17
2	20.62
2	14.15
2	10.41
2	19.57
3	21.97
3	21.15
3	21.17
3	15.59
3	23.19
3	22.57
3	12.42
3	15.32
3	8.87



a) Determine the highest degree of a polynomial that can be fitted to these data and sketch the corresponding ANOVA table (with degrees of freedom only).

b) Determine the degrees of freedom for pure error and lack of fit and the critical value for the F-test for lack of fit at the 5% level (from tables).