Math 484 Quiz 8 Fall 2016 Name____________________

	L1	L2	L3	L4
# of predictors	6	4	3	2
# with $0.01 \le$ p-value ≤ 0.05	0	0	0	0
# with p-value > 0.05	4	2	1	1
R^2_I	0.941	0.941	0.939	0.932
$\text{corr}(\hat{Y}, \hat{Y}_I)$	1.0	0.9998	0.9992	0.9953
$C_p(I)$	6.0	2.61	2.83	13.84
$\sqrt{MSE}$	16.700	16.592	16.687	17.584
p-value for partial F test	1.0	0.742	0.424	0.005

1) The above table gives summary statistics for 4 MLR models considered as final submodels after performing variable selection. The data set had 111 cases and the response plot and residual plot for the full model L1 was good. Model L2 was the minimum C_p model found.

a) For the partial F test, if the $0.07 \le$ p-value < 0.10, there is a small amount of evidence that Ho should be rejected. If $0.01 \le$ p-value < 0.07 then there is moderate evidence that Ho should be rejected. If p-value < 0.01 then there is strong evidence that Ho should be rejected. For which models, if any, is there strong evidence that "Ho: reduced model is good" should be rejected.

L4

b) Which model should be used as the final submodel? Explain briefly why each of the other 3 submodels should not be used.

L3 = I_I

L1 and L2 have too many predictors

L4 has $C_p > 2k$ and pval = .005 is too small

7) Find shorth(5) for the following data set. Show work.

6 76 90 90 94 94 95 97 97 1008

88 = 94-6

18 = 94-76

5 = 95-90

7 = 97-90

3 = 97-94 ← shortest

914 = 1008-94

$\boxed{\text{shorth}(5) = [94, 97]}$

Base terms: (log[TeachSal] log[TeachTax])

	df	RSS	k	C_I	
Add: log[Service]	41	3.34509	4	6.966	← model to look at
Add: log[Bread]	41	3.37904	4	7.413	
Add: log[EngSal]	41	3.44498	4	8.279	
Add: log[EngTax]	41	3.57509	4	9.989	
Add: log[BusFare]	41	3.5824	4	10.085	
Add: log[VacDays]	41	3.62314	4	10.621	
Add: log[WorkHrs]	41	3.63334	4	10.755	

Base terms: (log[TeachSal] log[TeachTax] log[Service])

	df	RSS	k	C_I	
Add: log[EngSal]	40	3.1498	5	6.400	← model to look at
Add: log[Bread]	40	3.18237	5	6.828	
Add: log[BusFare]	40	3.21972	5	7.319	
Add: log[EngTax]	40	3.32914	5	8.757	
Add: log[WorkHrs]	40	3.33908	5	8.887	
Add: log[VacDays]	40	3.34442	5	8.958	

Base terms: (log[TeachSal] log[TeachTax] log[Service] log[EngSal])

	df	RSS	k	C_I	
Add: log[BusFare]	39	2.88385	6	4.904	← $I_I = I_{min}$
Add: log[Bread]	39	3.07875	6	7.466	model I with
Add: log[EngTax]	39	3.11603	6	7.956	
Add: log[VacDays]	39	3.14563	6	8.345	$C_p(I) \le C_p(I_{min}) + 1 = 5.9$
Add: log[WorkHrs]	39	3.14787	6	8.374	

and # predictors ≤ # predictors for I_{min}

3) The above output is for the Big Mac data. Do not forget the constant.

a) List the k=4 model.

log(teach sal), log teach tax, log(service), constant

model to look at

b) List the k=5 model.

log(teachsal), log(teach tax), log(service), log EngSal), constant

model to look at

c) List the k=6 model.

log (teach sal), log(teachtax), log(service), log(EngSal) log(BusFare), constant

I_I

d) What is the value of k for model I_I?

k = 6

Other models to look at have $C_p(I) \le C_p(I_{min}) + 4 = 8.904$ and fewer predictors than I_I