STA 6208 – Spring 2003 – Exam 3

Print Name: **SSN:**

1) A study of the sensations of the alimentary canal investigated the relationship between the location of electrical stimulation (X) and the distance between the perceived and actual location by the subject (Y). The experiment was conducted 25 times at each of 5 locations (X=20 to 40 by 5). Each time the subject pointed to where they perceived the stimulation to be and the distance (in absolute value was obtained). A plot of the data is given at bottom of page.

The fit appears to be nonlinear, so the following models were fit: (i) linear, (ii) quadratic, and (iii) cubic. The following table gives model statistics and estimated regression coefficients (P-values), for X values centered around their mean ($\overline{X} = 30$). The total corrected sum of squares is 1477.3.

Model (p)	R^2	$MS(Residual)$	$\hat{\beta}_0$ (P-val)	$\hat{\beta}_1$ (P-val)	$\hat{\beta}_2$ (P-val)	$\hat{\beta}_3$ (P-val)
Linear (1)	.113	10.657	6.260 (.000)	-0.163 (.000)	—	—
Quadratic (2)	.230	9.327	7.666 (.000)	-0.163 (.000)	-0.028 (.000)	—
Cubic (3)	.299	8.553	7.666 (.000)	-0.527 (.000)	-0.028 (.000)	0.004 (.001)

a) Test whether given a cubic relationship, that the quadratic relationship is the true relationship at $\alpha = 0.05$ significance level. Clearly state the null and alternative hypotheses and decision.

b) For each model (Linear, Quadratic, Cubic), give the predicted value for each of the sampled X levels:

X	Centered X	$\hat{Y}_L$	$\hat{Y}_Q$	$\hat{Y}_C$
20	-10			
25	-5			
30	0			
35	5			
40	10			

c) Sketch these curves on the plot below.

d) At what location is the quadratic function maximized (be careful of units)?

2) A regression model is fit relating heating oil consumption (Y) to atmospheric temperature (X_1), amount of attic insulation (X_2) and an indicator of whether the house is a ranch (long, one story) style home ($X_3 = 1$ if ranch style, 0 otherwise). The following models are fit:

Model 1 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_3 + \beta_5 X_2 X_3 + \varepsilon$

Model 2 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

Model 3 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

Model 4 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ (for ranch style homes only)

Model 5 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ (for non-ranch style homes only)

The analysis is based on 15 homes, 7 ranch style, and 8 non-ranch style. The residual sum of squares for each model is given in the following table.

Model	SS(Residual)
1	2374
2	2728
3	8121
4	674
5	1699

a) Test whether the (linear) effects of atmospheric temperature and amount of attic insulation (on oil consumption) are the same for ranch and non-ranch style homes.

i) H_0 :

ii) H_A :

iii) Test Statistic:

iv) Rejection Region ($\alpha = 0.05$):

v) Conclusion:

b) Regardless of your previous part, test whether the regression parameters (β's) of the linear regression relating atmospheric temperature and amount of attic insulation to oil consumption are the same for ranch and non-ranch style homes.

i) H_0 :

ii) H_A :

iii) Test Statistic:

iv) Rejection Region ($\alpha = 0.05$):

v) Conclusion:

c) Give ν_1, ν_2, S_1^2, and S_2^2 which would be used to test for equal variances (where ν_i is the Residual degrees of freedom for the regression for group i.

d) Compute Adjusted-R^2 and C_p for models 1, 2, and 3, where $SS(TotalCorr) = 236135$. Note:

$$C_p = \frac{SS(Res)_p}{s^2} + 2p' - n$$

where p' is the number of parameters in the candidate model and s^2 is the estimate of σ^2 for the "full" model.

3) A regression model is fit, relating the carbon monoxide content (Y, in mg) to **Tar** (in mg), **Nicotine** (in mg), and **Weight** (in g). All regression models are given below.

a) Give the sequence of models (in terms of independent variables included), based on forward selection with SLE=0.20.

b) Give the sequence of models (in terms of independent variables included), based on backward elimination with SLS=0.40.

c) Give the sequence of models (in terms of independent variables included), based on stepwise regression with SLE=0.30 and SLS=0.30.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.74328	0.67521	4.06	0.0005
tar	1	0.80098	0.05032	15.92	<.0001

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.66467	0.99360	1.68	0.1074
nicotine	1	12.39541	1.05415	11.76	<.0001

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-11.79527	9.72163	-1.21	0.2373
weight	1	25.06820	9.98028	2.51	0.0195

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3.08961	0.84377	3.66	0.0014
tar	1	0.96247	0.23666	4.07	0.0005
nicotine	1	-2.64627	3.78720	-0.70	0.4920

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3.11433	3.41620	0.91	0.3718
tar	1	0.80419	0.05904	13.62	<.0001
weight	1	-0.42287	3.81299	-0.11	0.9127

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.61398	4.44663	0.36	0.7201
nicotine	1	12.38812	1.24473	9.95	<.0001
weight	1	0.05883	5.02395	0.01	0.9908

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3.20219	3.46175	0.93	0.3655
tar	1	0.96257	0.24224	3.97	0.0007
nicotine	1	-2.63166	3.90056	-0.67	0.5072
weight	1	-0.13048	3.88534	-0.03	0.9735

4) Which of the following four criteria will always give you the same model in model selection? Why?

i) Choose model with highest R^2.
ii) Choose model with highest Adjusted-R^2.
iii) Choose model with smallest $SS(Residual)$.
iv) Choose model with smallest $MS(Residual)$