STA 6208 – Spring 2002 – Exam 2

Print Name: **SSN:**

1) A company with a new product has no direct competition and is interested in the effects of two new potential advertisements on sales. They choose $n = 9$ test markets of comparable size, and vary the amounts of expenditures on each ad type among the markets. The response Y_i is sales of the product in the i^{th} test market, X_{i1} is the expenditures for ad type 1 in the i^{th} test market, and X_{i2} is the expenditures for ad type 2 in the i^{th} test market. They fit the model:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon \qquad \varepsilon \sim NID(0, \sigma^2)$$

They obtain the following data on expeditures and sales in the 9 test markets:

i	Y_i	X_{i1}	X_{i2}
1	10.28	0	0
2	17.73	0	5
3	25.28	0	10
4	12.90	5	0
5	20.16	5	5
6	27.01	5	10
7	15.37	10	0
8	22.50	10	5
9	29.81	10	10

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 0.4444 & -0.0333 & -0.0333 \\ -0.0333 & 0.0067 & 0.0000 \\ -0.0333 & 0.0000 & 0.0067 \end{bmatrix} \qquad \hat{\boldsymbol{\beta}} = \begin{bmatrix} 10.459 \\ 0.479 \\ 1.452 \end{bmatrix} \qquad SS(\text{Residual}) = 0.273$$

a) Give the null and alternative hypotheses for testing whether the effect of increasing expenditures for ad type 1 by one unit, controlling for ad type 2 expenditure is the same at the effect of increasing expenditures for ad type 2 by one unit, controlling for ad type 1 expenditure. Write it in the form $\mathbf{K}'\,\boldsymbol{\beta} = \mathbf{m}$

b) Give the estimate $K'\,\hat{\boldsymbol{\beta}}$, its estimated variance, and sum of squares.

c) Give the test statistic and rejection region for your test in part a) where $\alpha = 0.05$.

d) What would you conclude regarding the marginal effects of the ad types? Which would you use in a national advertising campaign?

2) For the linear regression model:

$$\mathbf{Y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

the response vector $\mathbf{Y}$ can be written as $\mathbf{Y} = \hat{\mathbf{Y}} + \mathbf{e}$.

a) Write the total uncorrected sum of squares in terms of the model and residual sums of squares. Show that the predicted vector $\hat{\mathbf{Y}}$ is orthogonal to the residual vector $\mathbf{e}$.

b) Give the distribution of $SS(\text{M}odel)/\sigma^2)$.

c) Give the distribution of $SS(\text{R}esidual)/\sigma^2)$.

d) Show that these two quantities are independent.

e) We wish to test $H_0: \boldsymbol{\beta} = \mathbf{0}$ vs $H_A: \boldsymbol{\beta} \neq \mathbf{0}$. Give the test statistic, as well as its distribution under H_0 and under H_A.

3) A regression model was fit relating Y (log sales among stores in a retail chain) to X_1 (log area in square feet of the store) and X_2 (log inventory), as well as an intercept term.

The sequential (Type I) and partial (Type II) sums of squares are given below for the model where $SS(\text{Total Corrected}) = 412.56$:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \quad \varepsilon \sim N(0, \sigma^2)$$

Term	Type I SS	Type II SS
Intercept	271091	19.40
X_1 (log area)	15.3826	0.0959
X_2 (log inventory)	211.32	211.3219

a) Give $R(\beta_2|\beta_0)$

b) Give the coefficient of partial determination between sales and area, controlling for inventory: $R^2_{Y1\cdot 2}$.

c) Give the coefficient of partial determination between sales and inventory, controlling for area: $R^2_{Y2\cdot 1}$.

4) A market researcher is interested in the relationship between the amount of time a consumer is exposed to an ad (X in seconds) and the relative importance a consumer places on a product attribute mentioned in the ad (Y on a scale of 0 to 10).

The experiment was conducted with 11 subjects being exposed to a 60 second ad, 10 subjects being exposed to a 105 second ad, and 9 subjects being exposed to the 150 second ad. You are given the following summary statistics for the importance ratings for each group, which can be used to obtain the least squares simple regression equation below.

Statistic	$X = 60$ Group	$X = 105$ group	$X = 150$ group
Mean	4.3	6.8	7.1
Std Dev	1.8	1.7	1.5
Sample Size	11	10	9

$$\hat{Y} = 2.7195 + 0.0319X$$

a) Give the fitted values for each group based on the simple linear regression model.

b) Give the Pure Error sums of squares and degrees of freedom.

c) Give the Lack of Fit sums of squares and degrees of freedom.

d) Test whether the simple linear regression model gives an adequate fit to this data ($\alpha = 0.05$).

5) Based on all possible regressions, consider the following four potential measures to base model selection on: R^2, Adjusted-R^2, $SS(\text{Residual})$, $MS(\text{Residual})$. Demonstrate any relations among using these 4 measures as model selection criteria.

6) 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=0.50.

b) Give the sequence of models (in terms of independent variables included), based on backward elimination with SLS=0.15.

c) Give the sequence of models (in terms of independent variables included), based on stepwise regression with SLE=0.50 and SLS=0.40.

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