THIS IS NOT REPRESNTATIVE OF CURRENT CLASS MATERIAL

STOR 455 Midterm 2 November 2, 2010

INSTRUCTIONS:

BOTH THE EXAM AND THE BUBBLE SHEET WILL BE COLLECTED. YOU MUST PRINT YOUR NAME AND SIGN THE HONOR PLEDGE ON THE BUBBLE SHEET. YOU MUST BUBBLE-IN YOUR NAME & YOUR STUDENT IDENTIFICATION NUMBER.

EACH QUESTION HAS ONLY ONE CORRECT CHOICE (decimals may need rounding).

USE "NUMBER 2" PENCIL ONLY - DO NOT USE INK - FILL BUBBLE COMPLETELY.

NO NOTES OR REMARKS ARE ACCEPTED - DO NOT TEAR OR FOLD THE BUBBLE SHEET.

A GRADE OF ZERO WILL BE ASSIGNED FOR THE ENTIRE EXAM IF THE BUBBLE SHEET IS NOT FILLED OUT ACCORDING TO THE ABOVE INSTRUCTIONS.

QUESTIONS are worth **1 point** each.

Consider the following SAS print out for questions 1-3:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	252378	252378	???	<.0001
Error	23	54825	2383.71562		
Corrected Total	24	307203			

Root MSE	???	R-Square	???
Dependent Mean	312.28000	Adj R-Sq	0.8138
Coeff Var	15.63447		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	62.36586	26.17743	2.38	0.0259
size	1	3.57020	0.34697	10.29	<.0001

1. The value of the F statistic is

A) 0.8215 B) 0.8138 C) 10.29 **D) 105.88** E) None of the above

2. The value of R-Square is
A) 0.8215 B) 0.8138 C) 10.29 D) 105.88 E) None of the above

3. The estimator of the residual σ is:
A) 0.8215 B) 0.8138 C) 10.29 D) 105.88 **E) None of the above**
(48.8)

Consider the following SAS code for questions 4-7:

```
proc reg data=blah;
model y=x1 x2 / ???;
run;
```

4. If we desire confidence intervals for unknown parameters we use
A) clm B) covb C) cli D) ss1 **E) None of the above**
(clb)

5. If we desire prediction for Y_h with intervals that express uncertainty in our prediction we use
A) clm B) covb **C) cli** D) ss1 E) None of the above

6. If we desire point estimate and a confidence interval for the subpopulation mean we use
A) clm B) covb C) cli D) ss1 E) None of the above

7. If we desire type I sum of squares we use
A) clm B) covb C) cli **D) ss1** E) None of the above

Consider the following SAS output for questions 8-11:

Number of Observations Used 18

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	91645	91645	315.38	<.0001
Error	16	4649.47024	290.59189		
Corrected Total	17	96295			

Root MSE	17.04676	R-Square	0.9517
Dependent Mean	630.05556	Adj R-Sq	0.9487
Coeff Var	2.70560		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	925.75267	17.12866	???	???
temperature	1	-13.75335	0.77445	???	???

8. What is the value of the t-statistic for testing $H_0: \beta_1=0$ vs. $H_1: \beta_1 \neq 0$?
A)-13.75 **B) -17.76** C) 17.76 D) 54.05 E) None of the above

9. Is the test $H_0: \beta_1=0$ vs. $H_1: \beta_1 \neq 0$ statistically significant?
A) yes B) no C) Cannot tell from the info provided

10. What is the value of the t-statistic for testing $H_0:\beta_0=0$ vs. $H_1:\beta_0 \neq 0$?
A)-13.75 **B) 54.05** C) 17.76 D) -17.76 E) None of the above

11. What is the 99% confidence interval for β_1?
A) (875.72, 975.78) **B) (-16.03, -11.49)**
C) (-15.98, -11.52) D) (-15.40, -12.11) E) None of the above

Consider the model $Y=\beta_0+\beta_1 X_1+ \beta_2 X_2+\varepsilon$ **and the following info for questions 12-15**
n=30, MSE=0.09689,

$$X'X = \begin{pmatrix} 40 & -4.394974844 & -4.340158967 \\ -4.394974844 & 32.444165213 & -0.522920052 \\ -4.340158967 & -0.522920052 & 44.611345727 \end{pmatrix} \qquad X'Y = \begin{pmatrix} 124.52133375 \\ -77.81767379 \\ 32.129310345 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 40 & -4.394974844 & -4.340158967 \\ -4.394974844 & 32.444165213 & -0.522920052 \\ -4.340158967 & -0.522920052 & 44.611345727 \end{pmatrix}$$

12. Find the value of b_1
A) -77.8 B) 32.4 **C) -1.98** D) 3.00 E) None of the above

13. Find the standard error of b_1
A) 0.055 B) 0.003 C) 0.097 D) 0.311 E) None of the above

14. When the data were generated we used the value of $\beta_1=-2$. Compute the 80% confidence interval for β_1. Does it contain the true value?
A) no B) not enough information C) it does not because 80% is too low.
D) there is an 80% chance it does **E) yes**

15. What are the degrees of freedom for the ANOVA F test F=MSR/MSE?
A) $df_R=2$, $df_E=27$ B) $df_R=1$, $df_E=28$ C) $df_R=3$, $df_E=27$ D) Not enough info
E) None of the above

For questions 16-18 use the extra sum of squares $SSR(X_1, X_2, X_3|X_4, X_5)$
16. $SSR(X_1, X_2, X_3|X_4, X_5)$ is equal to
A) $SSR(X_1, X_2, X_3, X_4, X_5) - SSR(X_4, X_5)$
B) $SSE(X_4, X_5) - SSE(X_1, X_2, X_3, X_4, X_5)$
C) $SSR(X_1, X_2, X_3, X_4| X_5) - SSR(X_4 | X_5)$
D) all three A, B and C are correct
E) both A and B are correct but C is not correct

17. The number of degrees of freedom associated with $SSR(X_1, X_2, X_3|X_4, X_5)$ is
A) 3 B) 2 C) 5 D) n-3 E) None of the above

18. The expression $SSR(X_1, X_2, X_3 \mid X_4, X_5)/SSE(X_4, X_5)$ gives
A) Nothing of significance
B) The F test statistic testing if adding X_1, X_2, X_3 to a model containing X_4, X_5 results in a statistically significant reduction in sum of squares.
C) Partial correlation measuring the reduction in the sum of square error resulting from adding X_1, X_2, X_3 to a model containing X_4, X_5.
D) Type I sum of squares
E) None of the above

19. The expression $MSR(X_1, X_2, X_3 \mid X_4, X_5)/MSE(X_1, X_2, X_3, X_4, X_5)$ gives
A) Nothing of significance
B) The F test statistic testing if adding X_1, X_2, X_3 to a model containing X_4, X_5 results in a statistically significant reduction in sum of squares.
C) Partial correlation measuring the reduction in the sum of square error resulting from adding X_1, X_2, X_3 to a model containing X_4, X_5.
D) Type I sum of squares
E) None of the above

Questions 20-21 relate to the following SAS statement model y= x2 x3 x1;
20. The type I sum of squares corresponding to this model statement are
A) SSR(x2|x1, x3), SSR(x3|x1, x2), SSR(x1|x2, x3)
B) SSE(x1), SSE(x2, x1), SSE(x3 x1, x2)
C) SSR(x1), SSR(x2|x1), SSR(x3|x1, x2)
D) SSR(x2), SSR(x3|x2), SSR(x1|x2, x3)
E) None of the above

21. The type II sum of squares corresponding to this model statement are
A) SSR(x2|x1, x3), SSR(x3|x1, x2), SSR(x1|x2, x3)
B) SSE(x1), SSE(x2, x1), SSE(x3 x1, x2)
C) SSR(x1), SSR(x2|x1), SSR(x3|x1, x2)
D) SSR(x2), SSR(x3|x2), SSR(x1|x2, x3)
E) None of the above

22. When working on residual analysis the plot of response variable Y vs. the studentized residuals r should
A) never be examined, examine the plot of predicted values $\hat{Y}$ vs. r instead;
B) always be examined;
C) studentized residuals should never be used, use regular residuals e instead;
D) should be examined only if the number of predictors is large.
E) None of the above

23. To fit quadratic regression $Y=\beta_0+\beta_1 x+\beta_2 x^2+\varepsilon$ in SAS we
A) create variable x2=x*x; in the data step first and than use it in proc reg writing model y=x x2;
B) in proc reg write **model y=x x*x;**
C) both A and B are correct
D) proc reg is for linear regression only, it cannot handle quadratic regression
E) None of the above

Consider the following SAS output for questions 24 - 25

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Type I SS	Type II SS	Variance Inflation
Intercept	1	1.41233	0.41564	3.40	0.0022	8192.95154	14.20146	0
x1	1	-0.05413	1.49021	-0.04	0.9713	2137.70546	0.00162	4069.04361
x2	1	1.03129	1.49076	0.69	0.4952	0.65472	0.58864	4068.96105
x3	1	-0.53272	0.20240	-2.63	0.0141	8.52098	8.52098	1.00216

24. Based on the SAS output, is there a problem with multicolinearity?
A) not enough info B) no **C) yes**

25. Notice that the Type I and Type II SS for x3 are equal. Does this give us any useful information?
A) Yes, x3 must be an important predictor.
B) No, the type I and type II SS are always equal for the last variable in the model.
C) No, type II SS are never useful
D) Yes, but we would need to assess the p-value first
E) None of the above

Questions 26-28 are based on the following
Consider a simple linear regression $Y=\beta_0+\beta_1 X_1+\varepsilon$. In matrix notation we write this model as $\mathbf{Y=X\beta+\varepsilon}$. The data was given in the following table

Y	5	11	4	9
X	1	-1	1	-1

26. The matrix X is

$$A)\ X=\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}\quad B)\ X=\begin{pmatrix}5&1\\11&-1\\4&1\\9&-1\end{pmatrix}\quad C)\ X=\begin{pmatrix}1&1\\1&-1\\1&1\\1&-1\end{pmatrix}\quad D)\ X=\begin{pmatrix}1&5\\1&11\\1&4\\1&9\end{pmatrix}$$

E) None of the above C is correct

27. The hat matrix H is

$$A)\ H=\begin{pmatrix}.25&0\\0&.25\end{pmatrix}\quad B)\ H=\begin{pmatrix}.25&-.25&.25&-.25\\-.25&.25&-.25&.25\\.25&-.25&.25&-.25\\-.25&.25&-.25&.25\end{pmatrix}\quad C)\ H=\begin{pmatrix}.5&0&.5&0\\0&.5&0&.5\\.5&0&.5&0\\0&.5&0&.5\end{pmatrix}$$

$$D)\ H=\begin{pmatrix}0.5323&0.0505&0.4959&-0.0223\\0.0505&0.5699&0.0118&0.4924\\0.4959&0.0118&0.4642&-0.0517\\-0.0223&0.4924&-0.0517&0.4336\end{pmatrix}$$

E) None of the above C is correct

28. The mean square error is
A) 70.91 **B) 1.25** C) 0.833 D) 2.5 E) None of the above