Psychology 204b (Causal Modeling) Simonton Winter 2008

Exam I Your Name _____.

The statistics reported on the last page of this test are taken from a database regarding award-winning narrative films. The dependent variable LGROSS is a measure of the US box office earnings (in millions of dollars, log transformed). The independent variables are as follows: BUDGET = cost to make (in millions of dollars); SCREENS = the number of screens on which it opened; METACRIT = a composite of the critical reviews the film received on first theatrical release (on a 1-100 point scale); BEST = a measure of the number of best picture nominations and awards received from 7 major organizations; DRAMATIC, VISUAL, TECHNIC, and MUSIC similar tabulations for awards in the corresponding categories (e.g., DRAMATIC = directing, writing, acting, and editing). All variables are in raw-score form.

1) Use the results given in the printout to construct two separate regression equations, one for the original scores (3 points) and the other for the standardized scores (2 points). (total: 5 points)

2) (a) What proportion of variance in the dependent variable is explained by the variables in the equation? (2 points)
(b) Is there any difference between sample and population estimates in this respect? What does this mean? (1 point)
(c) What is the correlation between the predicted value and the actual value? (2 points) (total: 5 points)

3) (a) If one subtracted the predicted score from the observed score, squared the difference, summed up all these squared residuals, and then divided by the appropriate degrees of freedom, what number would result? (2 points) (b) The square root of this last number is what? (1 point) What does this last quantity tell us? (1 point) (c) Why are no calculations required for answering these questions? (1 point) (total: 5 points)

4) (a) What is the purpose of the Analysis of Variance? (2 points) (b) What does the *F* ratio mean? (1 point) (c) How are the degrees of freedom arrived at? (1 point) (d) What decision do these results lead to? (1 point) (total: 5 points)

5) (a) Looking at the actual regression coefficients, when are the unstandardized slopes ("Coefficient") more informative? (2 points) (b) Under what circumstances are the standardized slopes ("Std Coef") more useful? (1 point) (c) Why is the standardized coefficient for the constant term zero? (1 point) (d) What does the constant term mean? (1 point) (total: 5 points)

6) (a) What do the values in the column marked “Std Error” tell us? (2 points) (b) If these values were not given, how might they be calculated from other columns on the printout? (3 points) (total: 5 points)

7) (a) If the significance levels for the predictors were omitted from the printout, how could these be determined from the values under column “t”? (2 points) (b) If we had no table of *t* values, but instead had a table of *F*s, what could we do then? (1 point) (c) If columns marked “t” and “P(2 Tail)” were both missing, how could we still determine that a given regression coefficient was statistically significant at the .05 level according to a 2-tailed test? (2 points) (total: 5 points)

8) (a) Using the available information, calculate the semi-partial (or part) correlation between LGROSS and BUDGET removing from the latter variable all the variance that it shares with the other independent variables in the equation. (2 points) (b) What does the square of this semi-partial tell you? (1 point) (c) How is the size of this semi-partial related to that of the corresponding standardized coefficient? Which coefficient is smaller than the other? Under what situations would these be the same? (1 point) (d) Which of these must be statistically significant given the results on the printout? (1 point) (total: 5 points)

9) (a) Now calculate the squared semi-partial correlations for all the independent variables in the equation. Add them all up. (2 points) (b) How does this sum compare with the squared multiple correlation? Smaller or the same? (1 point) (c) How does this sum compare with the sum of the squared zero-order correlations between the dependent variable and each independent variable? (1 point) (d) What do these comparisons signify? (1 point) (5 points)

10) (a) What information on the printout enables you to discern the presence of suppression? (3 points) (b) Are there any suppression effects? If so, provide a substantive interpretation. (2 points) (total: 5 points)

11) (a) What information on the printout enables you to determine whether multicollinearity is a major problem? (2 points) (b) What does the column labeled "Tolerance" tell us? (2 points) (c) What is the reciprocal of a variable's tolerance sometimes called? Why? (1 point) (total: 5 points)

12) Suppose that the column marked “Coefficient” was totally obliterated. Show how you could reconstruct all the missing values from the information still provided. (3 points) Please do the actual calculations for the CONSTANT and for METACRIT. (2 points) (total: 5 points)

13) (a) Now a prediction problem using raw scores. Suppose that a particular film scores two standard deviations above the mean in METACRIT and BEST picture awards, one standard deviation below the mean on SCREENS and BUDGET, and exactly at the mean on all of the remaining variables. (Remember: none are in mean-deviation form). Reconstruct the prediction equation with the proper values plugged in, and obtain the point estimate for LGROSS. (7 points) (b) What happens to the predicted score if you ignore the predictors that are not statistically significant? (3 points) (total: 10 points)

14) Now another prediction problem, this time in standard (*z*) scores. Suppose the film scored two standard deviations lower than average on BUDGET and SCREENs, one standard deviation above the mean on METACRIT and DRAMATIC, and exactly at the mean on everything else. What is the predicted *z*-score for LGROSS? (total: 10 points)

15) Suppose that you decided to perform a stepwise regression on these data. What would you expect to be the most likely result? (3 points) What causal inferences might you draw from this expected outcome? (2 points) (total: 5 points)

16) (a) Contemplate the following statistics of the printout: the means, standard deviations, correlations, *R*, adjusted-R^2, standard error of estimate, coefficient, standard error, standardized coefficient, tolerance, *t*, *F*-ratio. Identify which are "descriptive" and which are "inferential." That is, which numbers describe sample characteristics, and which are involved in making inferences about population attributes? (3 points) (b) Do these two sets of statistics make different basic assumptions to justify their calculation and interpretation? If so, what are those differences? (2 points) (total: 5 points)

17) (a) Suppose that you do not like this equation and want to estimate your own. Specifically, you want to predict LGROSS using just two independent variables, BUDGET and SCREENS (i.e., the only things the distributor knows prior to opening weekend). In your enthusiasm for hand-done regression, you calculate the following: the standardized and unstandardized regression coefficients, the constant term, and the squared multiple correlation. (8 points) (b) Is this equation better or worse than the one given? Any surprises? If so, what do they mean? Does this change your answer to #15? (2 points) (total: 10 points)

	LGROSS	BUDGET	SCREENS	METACRIT	BEST
N of cases	329	329	329	329	329
Minimum	-0.6931	0.0000	0.0000	9.0000	0.0000
Maximum	6.2136	200.0000	4163.0000	98.0000	1.8750
Mean	3.5563	49.6474	1764.7903	59.1611	0.1292
Standard Dev	1.4629	40.1095	1351.6160	20.3127	0.3004

	DRAMATIC	VISUAL	TECHNIC	MUSIC
N of cases	329	329	329	329
Minimum	0.0000	0.0000	0.0000	0.0000
Maximum	7.5833	5.2500	6.0000	3.1250
Mean	0.6859	0.3505	0.3921	0.1755
Standard Dev	1.3304	0.8747	0.8363	0.4032

Pearson correlation matrix

	LGROSS	BUDGET	SCREENS	METACRIT	BEST
LGROSS	1.0000				
BUDGET	0.6269	1.0000			
SCREENS	0.6294	0.6925	1.0000		
METACRIT	0.0933	-0.1341	-0.3815	1.0000	
BEST	0.1396	-0.0446	-0.2646	0.4686	1.0000
DRAMATIC	0.1502	-0.1010	-0.3443	0.5016	0.7900
VISUAL	0.2573	0.1949	-0.0429	0.2726	0.5772
TECHNIC	0.4215	0.4330	0.1996	0.2578	0.4045
MUSIC	0.3040	0.2002	0.0494	0.3301	0.5362

	DRAMATIC	VISUAL	TECHNIC	MUSIC
DRAMATIC	1.0000			
VISUAL	0.4970	1.0000		
TECHNIC	0.3525	0.5660	1.0000	
MUSIC	0.4206	0.5818	0.4541	1.0000

Number of observations: 329

Dep Var: LGROSS N: 329 Multiple R: 0.7912 Squared multiple R: 0.6259

Adjusted squared multiple R: 0.6166 Standard error of estimate: 0.9058

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	0.7475	0.2158	0.0000	.	3.4641	0.0006
BUDGET	0.0075	0.0019	0.2060	0.4238	3.9224	0.0001
SCREENS	0.0007	0.0001	0.6477	0.3999	11.9787	0.0000
METACRIT	0.0162	0.0031	0.2244	0.6363	5.2344	0.0000
BEST	-0.3337	0.3015	-0.0685	0.3049	-1.1066	0.2693
DRAMATIC	0.3183	0.0647	0.2895	0.3372	4.9169	0.0000
VISUAL	0.0710	0.0832	0.0424	0.4719	0.8526	0.3945
TECHNIC	0.0562	0.0829	0.0321	0.5210	0.6783	0.4981
MUSIC	0.1175	0.1666	0.0324	0.5543	0.7051	0.4813

Analysis of Variance

Source	Sum-of-Squares	df	Mean-Square	F-ratio	P
Regression	439.3930	8	54.9241	66.9351	0.0000
Residual	262.5783	320	0.8206		