

Stat 328 Final Exam Solution Summer 2000

1. (a) **(3 points)** The variable with the largest absolute value of its correlation with y is variable x_5 . It would produce the largest R^2 in a SLR since R^2 will be the square of that correlation. The sign of that correlation is $+$, which is the same as the sign one gets for the fitted regression coefficient, b_5 in a SLR.

(b) **(2 points)** R^2 for the single SLR on the printout is .836. The regression with dummy and cross product allows for 2 completely different lines (2 SLRs) to be fit to the early and then the late data. It has

$$R^2 = \frac{SSR}{SSTot} = \frac{SSTot - SSE}{SSTot} = \frac{.02372439 - .00238427}{.02372439} = .900$$

This is a reasonable increase in R^2 and indicates that perhaps two SLRs should be considered.

(c) **(3 points)** $.0013324 \pm (2.074)(.000126)$

(d) **(2 points)** This is roughly .265 to .322 reading from the "outer" limits on the JMP plot on page 3 of the printout.

(e) **(3 points)** A p -value less than .0001 comes from either the t test of $H_0: \beta_1 = 0$ summarized on page 4 or the equivalent F test of the same hypothesis in the "effect test" for x_1 on page 6 of the printout. Even after accounting for x_5 , the explanatory power of x_1 is an important addition to the model.

(f) **(2 points)** $x_1 = 270$ and $x_2 = 300$ imply $x_5 = 81,000$. Similarly, $x_1 = 273$ and $x_2 = 310$ imply $x_5 = 84,630$. Then

$$L = (\beta_0 + 273\beta_1 + 84,630\beta_5) - (\beta_0 + 270\beta_1 + 81,000\beta_5) = 3\beta_1 + 3,630\beta_5$$

(g) **(2 points)** For $x_1 = 270$, $\mu_{y|x_2} = \beta_0 + 270\beta_1 + 270\beta_5x_2$. That is, the rate of increase of mean y wrt x_2 is $270\beta_5$. So multiply confidence limits for β_5 by 270. These are then $270(.000009) \pm 270(2.080)(.000002)$.

(h) **(2 points)** I see a downward trend in the residuals in the pre-embargo data points and another one in the post-embargo data points. This model 1) underpredicts small pre-embargo y' and overpredicts large pre-embargo y' , while it also 2) underpredicts small post-embargo y' and overpredicts large post-embargo y' . This is probably because I have fit a single model to 2 basically different regimes/data sets.

(i) **(2 points)** Data point #9 (the next-to-last pre-embargo data point) probably has the "largest combination of "leverage" h_{ii} and "Cook's influence" D_i . (It also has the largest x_5 and y' values in the data set, and is in some sense then an "extreme" point.)

(j) **(2 points)** The e_i and e_{i-1} appear to be correlated. That says that the ϵ_i are probably not sensibly modeled as independent. This calls into question the relevance of the MLR nominal confidence and significance levels. The fact that the "current" residual is positive suggests (since residuals should "average to" 0 and they seem to be positively correlated) that the next one might also be slightly positive, just using the regression model. This suggests that I increase my prediction from the regression model slightly (to try to cancel this effect).

2. (a) **(2 points)** There is nothing in this plot that causes worry about the constant variance assumption. If anything, the plot looks "too good to be true," with essentially the same vertical spread for each combination. $s_{\text{Pooled}} = .222$. σ is the standard deviation of sales for any fixed short term sales strategy (combination of "Display" and "Price").

(b) **(3 points)** A p -value less than .0001 comes from the overall F test that all β 's for dummies are 0. JMP lists this as the "effect test" for "Combination" on page 3. There are clearly differences in mean sales somewhere in the 9 combinations.

(c) **(2 points)** This is combination 9. On page 4 the "least squares mean" is $\hat{y} = 18.297$, with standard error of .1284. So use $\hat{y} \pm t\sqrt{s^2 + (\text{std error for } \hat{y})^2}$. Here that is $18.297 \pm 2.101\sqrt{(.222)^2 + (.1284)^2}$.

(d) **(2 points)** We probably DO need to include the interactions. 1) Going from the reduced to the full model, R^2 goes from .90 to .99. 2) The F test for lack of fit in the reduced mode is equivalent to the F test for dropping the interactions terms from the full model and this has $F = 258$ with p -value less than .0001.

(e) **(3 points)** Confidence limits for β_{A1} are $-2.854 \pm 2.101(.0605)$. This is the difference between the average of the three "Display 1" mean sales figures and the average of all 9 mean sales figures.

(f) **(2 points)** The upper left 2×2 block of the table comes straight off the printout as β 's for products of A and B dummies. Then the fact that rows and columns must sum to 0 lets you finish out the table. The whole table is

1.56	-.47	-1.09
-2.53	.39	2.14
.97	.08	-1.05

(g) **(3 points)** Under this scenario, the mean sales are $10.146 \times (\$100) = 1014.60$ and the standard deviation of sales is $.222 \times (\$100) = \22.2 . The model further says that observations are normal. So, $z = (1000 - 1014.6)/22.2 = -.66$. So the desired fraction is $.5 + .2454 = .7454$.