Solutions for 2003 Statistics Waiver Exam

(1) C

The slope in this model has units of $/cup, so the expected spending per cup sold is $3.30.

(2) D

A confidence interval gives a plausible range for the population characteristic (here, the intercept) that is consistent with the observed sample.

(3) D

From the fitted model, the prediction is 79.9 + 3.3(500) = 1729.9

(4) A

The RMSE of the fitted model is $309, so being $300 low for one day is not unusual.

(5) E

The SE of the slope gets smaller with more variation in the predictor.

(6) C

With more specialty sales, we would expect a steeper (or larger) slope because it measures the typical spending per cup.

(7) C

The change expected for an increase in sales of 100 more cups is 100 times the slope of the fitted model. Taking 100 times the 95% interval 3.3 ± .6 gives the answer.

(8) B

These data are a time series, and the residuals should be investigated for time trends and autocorrelation.

(9) D

Since the square root of the sample size appears in the denominator of the SE for the slope, the CI for the slope would be shorter with more data.

(10) E

This plot shows no anomalous features.

(11) E

The normal quantile plot of the residuals is consistent with normally distributed errors, but it does not prove that the underlying errors are normal.

(12) E

Both of these approaches add the same information to the model. For “D” notice that once you know the total number of cups and the number of small and medium cups, you also know the number of large cups.

(13) B

The slope indicates the typical growth per year, so in five years we expect to see a gain of about $30,000,000 (recall the data are in thousands). The SE of this estimate also increases by a factor of 5 to 1000. So, a 95% interval has width ± 2(1000). The RMSE is not directly germane to this question.

(14) B

This is the difference of the two categorical terms, 28363+15486 = 43849 (thousand).

(15) C

The sales for each quarter increases by 6031 (thousand), so the total growth is 4 times this increment.

(16) E

Consider the plots of the data and residuals. A linear model for the trend such as that used by this model under-predicts the early and later data.

(17) E

To estimate a different slope for each group of quarters (i.e., one slope for Q1, one for Q2, etc) would require the addition of the interaction.

(18) D

The plots of the original data and residual plot indicate the presence of curvature missed by the model as well as increasing variation. Logs handle this mixture of problems well.

(19) B

The correlation matrix shows the correlation between log price and torque to be 0.757. Thus the R^2 of the simple regression of log price on torque would be 0.573. From this, the F ratio for the simple regression would be 144 (0.573)/(1-0.573) = 193

(20) E

There‘s a lot more in the plots shown in the scatterplot matrix than found in the correlations alone.

(21) A

The square root of the R^2 is the correlation between predicted values and the actual response.

(22) E

Converted to the log scale 37000 becomes 10.5187. Now add 2 RMSE to obtain the upper end of the 95% prediction interval, 10.9447 on the log scale. Exponentiating gives the answer on the dollar scale.

(23) D

The partial F for cylinders is not significant. You cannot look directly at the raw p-values and t-stats when the categorical variable has many levels like this one.

(24) B

Because the model is non linear, the effects of changes in HP are greater for larger values of HP. The simple approach is to observe that each addition of 1HP increases the price by about .45% (the slope for HP) on average. Hence, a 20 HP change would produce about a 9% change. Since cars with 160 HP are more expensive than those with only 120, its 9% of a larger value. Worked out in detail, assume that the price for some set of torque, cylinders, fuel efficiency, and HP=120 is $37,000. Increasing HP by 20 increases the expected price by *$3,484* to $40,484. If the HP is moved up to 160, then the expected price jumps to $44,297. If the HP goes further up to 180, the price jumps by more than *$4,000* to $48,469.

(25) A

These standard errors are similar to those for an average, for which the SE is inversely proportional to the square root of the sample size.

(26) B

The data is collinear, as shown in the correlations and scatterplot matrix.

(27) E

This is the definition of a p-value.

(28) E

Such imaginary cars are quite far from the observed data. As a prediction, the intercept represents a huge extrapolation from the rest of the data.

(29) B

The leverage plot shows this point pulling up the partial slope for HP.

(30) D

This compression of the range of data is a sign of collinearity (which in effect reduces the variation in correlated predictors).

(31) A

Yes, the effect test for *Efficiency* is significant in the two-way anova that controls for the variation due to style preference (which otherwise obscures the effect of efficiency, as in the *FlexTime* example covered in the casebook and class).

(32) C

The easiest way to get the fit for this combination is to look at the table of means supplied in the output, giving 1021.8. Now add ± 1 RMSE = sqrt(25983) = 161 (see the anova summary table)

(33) D

An interaction would imply such an effect.

(34) D

From the sums of squares in the ANOVA table, R^2 = 1395846/3032791

(35) D

Since the design of this experiment is balanced, the effects are uncorrelated with each other. So, the variation associated with the interaction would be returned to the error terms along with its DF. The F for efficiency would then be

$$F = (234236/2)/ ((3032791+187485)/(63+4))=2.44$$

(36) D

The columns are located at the predicted values, and there appear to be fewer than 9 distinct groups.

(37) B

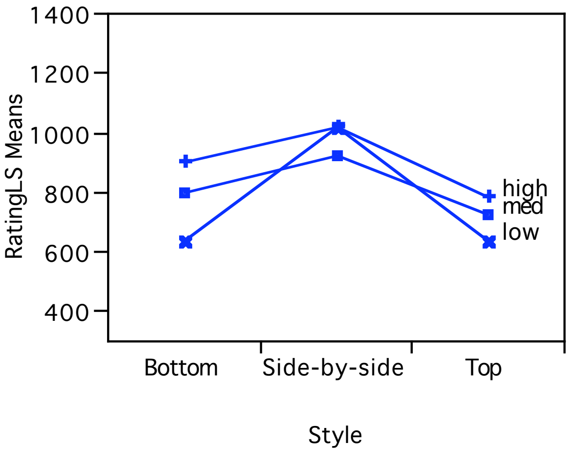
A basic diagnostic.

(38) B

In order to find the largest of the 9 averages, we need the Hsu comparisons.

(39) D

The lines would cross very slightly as shown below. The effect is not significant.



(40) A

As in a paired t-test, such a sampling design with repeated responses from the same person requires a rather different analysis to deal with the dependence.

(41) C

This question requires the partial elasticity (the coefficient of log capital in model c), multiplied by 2.

(42) B

The net increase (in percent) for increasing both by 1/2 of a percent is the average of the two coefficients in the multiple regression (0.38). Hence this model suggests putting all of the increase into capital (the term with the larger slope). Of course, one could not continue to pour all assets into capital alone in this fashion indefinitely, barring some huge innovation in productivity.

(43) B

Though the predictors are collinear, the addition of log labor to the model with log capital results in a significant improvement (as given by the t-ratio).

(44) B

The estimated coefficient for log labor is less than 0.5 (at 0.34), but with standard error 0.14, it is not significantly less than 0.5. For example, the 95% confidence interval for the slope is 0.34 ± 2(0.14), a range that includes 0.5.