Solutions for 2002 Statistics Waiver Exam

(1) C

Consider the size of the intercept. If you consider the size of the prediction for a store with, say \$30,000 in sales in 2000, the sales are larger on average in 2001. The scales shown for the two variables in the scatterplot of the data also reinforce the increase.

(2) C

The interval is $0.61 \pm 2(.12)$.

(3) D

From the fitted model, the predicted sales are $14.8 + .61(30) = 33{,}100$

(4) C

The expected difference is $0.61(2000) = 1220$.

(5) B or C

The predicted level of sales for this new store is $14.8 + 0.61(29) = \$32490$. Thus the actual sales at \$33,000 are slightly higher than predicted when one notes the RMSE is \$844. i.e., the actual sales are about 1 RMSE larger than predicted.

(6) D

Under the assumptions of normality and independence, the difference in sales is the difference between two random errors, say $e_1 - e_2$. Both are normal with mean zero and SD = RMSE. Hence, the expected difference is zero (both are predicted to be the same) and the variance of the difference is the sum of the variances, or 2 $RMSE^2$. The SD of the difference is sqrt(2) RMSE, or $1.414(0.844) = \$1{,}200$. Thus, the chance of the difference being more than one SD from its mean is about 1/3.

(7) A

Such a store would add more variability to the predictor and hence improve the accuracy of the slope estimate by reducing its standard error.

(8) E

No problems here.

(9) D

Any of the steps would be useful.

(10) E

The data are not time series. The analyst must have sorted the data into some ordering to produce such an anomalous DW statistic.

(11) D

A paired t-test would take into account other similarity of the stores and allow a direct assessment of the change over the year.

(12) D

The p-value for the intercept is the probability of an estimated intercept so large as the one observed by chance if the null hypothesis that the population intercept equals zero is true.

(13) C

Removing either of the terms would cause the R^2 to decrease significantly.

(14) A

The overall F-ratio is very significant.

(15) A

For one cart on a rainy day, the slope is $124 - 78 = 46$, which is about 50.

(16) B

The 95% CI for the effect of temperature is 21.1 ± 26.4 = [-5.3, 47.5]. Multiply the upper bound by 10.

(17) D

The effect is twice the shown coefficient for the categorical term. The rest of the values are imply that nothing else changes so that the multiple regression coefficient is appropriate.

(18) A

The predicted value indicates that the vendor has a cushion of 400 units, which is about 1 RMSE.

(19) E

The slight clustering is due to the lower sales on rainy days.

(20) C

The interaction would capture the difference in slopes.

(21) E

Nothing unusual, no collinearity (narrowing) or outliers. All is well, but just not significant.

(22) C

The intercept is the asymptote that controls the size of the fit for large values of the predictor.

(23) B

One can work directly with the change in the predicted awareness, or just take the derivative of the fit. The derivative is 0.9/day^2. At day 7, the slope is about 0.018, implying a sales gain of $18,000 obtained at a cost of $20,000.

(24) B

The awareness expected on day 10 is .88-.9/10 = .88-.09 = .79. Multiply times 100.

(25) C

Plot the data using the transformed coordinates.

(26) D

Add the factor to the model, not analyze it separately (which would introduce confounding).

(27) B

Small proportions near 0 and large proportions near 1 are less variable than proportions near 1/2.

(28) E

Just the average for this combination from the table.

(29) E

This is just the Bonferroni method. The fact that the interaction is significant does not imply that it specifically affects Technology2.

(30) D

Draw the profile plot of the averages and you will see that the others are not true. Averaging the shown group means gives 43.

(31) A

This is the interaction.

(32) D

Was it the rain or was it the Farm location?

(33) D

From the correlations, $F = (n-2) R^2/(1- R^2)=280 (.75^2)/1-.75^2) = 360$

(34) C

The confidence interval for the elasticity of sampling in the multiple regression is -.05 ± 2(0.15) = [-0.35, .25]. So a 10% increase in sampling could get at most a 2.5% gain in scripts. You need the multiple regression since the other factors are held constant.

(35) E

The analysis does not have an interaction to answer this question.

(36) C

The sales are "perfectly elastic" in the number of visits. Increase the number of visits, and you get a proportional gain in scripts.

(37) C

Again, an elasticity.

(38) B

Tukey-Kramer is designed for these pair-wise comparisons of categories.

(39) E

One cannot directly compare R^2 statistics when the response variable has changed.

(40) D

Collinearity narrows the leverage plot.

(41) A

The interaction would allow them to see the size of the differences.

(42) C

This is just one component of a categorical factor that was added as a group, not individually.