Nopane Advertising Strategy Analysis

Questions:

1. What does Regression 1 in the case say about the merits of "emotional" vs. "rational" copy? What does Regression 3 say about the two types of copy? What is the interpretation of the coefficient on copy in Regression 1? Regression 3?

In Regression 1, the estimated coefficient on the "Dum Copy" variable is 2.134. Since Dum Copy = 1 if the "emotional" copy is used, and = 0 if the "rational copy is used, **a positive coefficient would indicate that the "emotional" copy results in higher average sales than the "rational" copy**. However, we should note that the t-value associated with the hypothesis test with null hypothesis that the true coefficient on Dum Copy = 0 is 1.1. The degrees of freedom associated with this t-value are n - # of coefficients estimated = 19. Therefore we can determine the corresponding p-value to be 0.285. (Note that if we actually calculated the t-value by dividing 2.134 by 2.027, we get a number closer to 1.05 than to 1.1. The 1.1 in the case just comes from rounding up. So really, a more exact p-value is the one corresponding to a t of 1.05, or p-value = 0.306.) **Thus, even though the estimated coefficient is positive, our hypothesis test tells us that the evidence that it is different from zero is not very strong.**

Similarly, in Regression 3, the estimated coefficient on Dum Copy is -3.00. **This indicates that the "rational" copy seems to give higher average sales than the "emotional" copy.** The t-value given here is -1.7 and the number of degrees of freedom = 20. Therefore the associated p-value for the hypothesis test with null that the true coefficient equals 0 is 0.1046. **Thus, while still not overwhelmingly strong, we do have moderately strong evidence that the true coefficient is different from zero**. In particular, if we wanted to prove that the true coefficient on Dum Copy in Regression 3 was less than zero, the corresponding p-value would be 0.0523, indicating only about a 5% chance of making a mistake if we conclude that the true coefficient is less than zero.

These two coefficients, then, seem quite different. How can this be? The key to understanding is to interpret the meaning of the coefficient on Dum Copy in the two regressions and see the difference in interpretation. **In Regression 1, the coefficient on Dum Copy tells us the average additional sales per 100 prospects that we would expect if we switched from using the "rational" copy to using the "emotional" copy, holding fixed the amount of Nopane ad spending, the amount of our competitors' ad spending, and the Segment where the sales territory is located. In Regression 3, the coefficient on Dum Copy tells us the same thing *except* our competitors' ad spending is not held fixed.** (We can see this because the only difference between the two regressions is that the Competition variable is not included in Regression 3.) Thus Regression 1 predicts 2.134 additional sales per 100 territories, on average, holding the three other variables fixed, for the "emotional" copy as compared to the "rational" copy.

Regression 2 predicts an average loss of 3 sales per 100 territories, holding Segment and our ad spending fixed, for the "emotional" as compared to the "rational" copy.

(Note: As we have seen, it is important to check the plot of residuals vs. fitted values and run a BP test when you look at a regression. Doing so for Regressions 1 and 3 turns up essentially no evidence of non-constant variance (very high p-values for the BP test) and no distinct curvature, so things look pretty good as far as our regression assumptions go.)

2. Assuming Alison Silk's hypothesis is correct, which of the regressions is most relevant for choosing an advertising strategy? Why?

Ms. Silk suggested (see p. 3 of the case) that Nopane's competitors might react to a national strategy by Nopane in a way very similar to their reaction in the advertising experiment. **If this is true then Regression 3, the regression leaving out the "competitors' advertising" variable, is the most relevant one.** Why? Think about what it means to set a national advertising strategy for Nopane -- choosing both ad copy and an advertising spending level. To help in choosing a strategy, the question that needs to be answered for each strategy is: what will happen to sales if this strategy is implemented? We can use Regression 3 to provide an answer (for each Segment, A and B) by plugging in the values corresponding to the copy and the Nopane advertising spending under the given strategy. For example, this method predicts that running the "emotional" copy and spending $4.75 per 100 prospects will yield an average of 28.66 unit sales per 100 prospects in Segment A (on the coasts) and 28.83 units in Segment B (the middle of the country). The prediction for "rational" copy and $2.50 per 100 prospects is 29.42 (region A) and 29.59 (region B). Regression 1 cannot be used directly for this prediction, since we do not know the values of competitors' advertising to plug-in under the various strategies.

Q: But what about omitted variable bias? We can see from regression 2 that competitors' advertising spending seems related to our copy and ad spending. Doesn't this mean that the estimated coefficients on copy and ad spending in regression 3 are biased?

A: Yes, we would expect the coefficient estimates to be biased. However, recall the nature of the bias: the coefficient estimates reflect not only the effect of the variable included in the regression (holding other included variables fixed) but *also* include the effect of any variables left out of the regression, to the extent that they are related to the included variable. Specifically, our coefficient estimate of -3.00 on Dum Copy in Regression 3 *includes* the effect of the expected change in competitors' advertising spending when Nopane's copy changes. Since we are assuming that competitors will behave in the real campaign as they did in the experiment, this "combined effect" is exactly what we want to estimate, so the "bias" works in our favor.

Sidelight: How can we relate this combined effect to the other regressions? In Regression 1, the estimated coefficient on Dum Copy is 2.134. This says that if we use

2

the "emotional" copy rather than the "rational" copy, and all the other variables (including competitors' behavior) are held fixed, we would expect 2.134 more sales per 100 prospects. But what does Regression 2 say is the effect on competitors' behavior of using "emotional" copy rather than "rational" copy? It says that on average we would expect them to spend 9.083 dollars more per 100 prospects when we use the "emotional" copy. What effect does this have on sales? From Regression 1, we expect this to change sales by -(0.5652)*9.083 = -5.134 units. So, the expected net effect of using "emotional" rather than "rational" copy and having our competitors react is 2.134 - 5.134 = -3.00 !

(Hard bonus sidelight to think about: What is wrong with using Regression 2 to predict competitors' advertising spending in response to our policy and then plug that value into Regression 1 to get a prediction for sales? Does this give the same predicted value? The same prediction interval?)

(Answer: Yes, same predicted value. Smaller prediction interval. Prediction interval is not the correct one since the only reason it is smaller is that Regression 1 incorporates the movement of competitors' ad spending which is unrelated to the other variables (we can see there is some, since the R-squared in Regression 2 is only .571). But this is precisely the component of our knowledge of competitors' behavior which is lacking when we use Regression 2 to predict it. Thus we need to take into account the extra variability that comes from knowing the competitors' behavior imperfectly. Regression 3 does this.)

3. Answer all the questions in 2 assuming that Stanley Skamarycz's hypothesis is correct.

Here Regression 1 is the most appropriate choice, since Mr. Skamarycz believes (see p. 3 of the case) that he can accurately specify competitors' behavior ($19 per 100 prospects always) and this behavior differs substantially from that in the experiment. This is where the argument above about omitted variable bias does have bite. Once we know competitors' ad spending, we want to estimate the effects of our strategy holding the competitors' ad spending fixed at the known level. The proper way to do this is by including competitors' ad spending in the multiple regression. Regression 1 does this.

4. Given the data from the case (in nopane.xls), what national advertising strategy (i.e., copy and one of the three levels of ad spending) would you advocate? Assume an additional unit sold per 100 prospects over a six-month period yields a profit (net of production and delivery costs, but <u>not</u> net of advertising costs) of 10 dollars per 100 prospects over a six-month period. Provide support for your position.

As noted above, Regression 1 and Regression 3 have opposite (or indeterminate) implications for which ad copy to use. In both of these regressions, the coefficient on Nopane ad dollars is significantly greater than 0.10 (the break-even level of 1 extra unit expected for each 10 dollars in additional advertising) at a 5% level, thus both suggest

setting ad spending as high as possible -- $8.00 per 100 prospects. It seems that the regressions don't give an unambiguous recommendation for copy, though they do for spending.

However, in thinking further about these regression specifications, it seems we may have left out some obvious effects. In particular, these regressions assume that the effect of advertising expenditure is the same no matter what ad copy we run. While this may be true, there is certainly no reason to think it must be. They also assume that advertising expenditure is equally effective across regions and that copy is equally effective across regions as well. We can modify our analysis to check for these effects by adding slope dummy variables to the regressions: specifically we add (Nopane ad dollars)*(Dum Copy), (Nopane ad dollars)*(Dum Region), and (Dum Copy)*(Dum Region). The modified Regression 1 is shown below:

Regression: Unit Sales

	constant	Ad Dollars	Competition	Dum Seg	Dum Copy
coefficient	23.0526814	1.93656481	-0.10347303	-1.4099363	10.8363976
std error of coef	4.68728679	0.47803475	0.23921527	3.29164829	3.43227586
t-ratio	4.9181	4.0511	-0.4326	-0.4283	3.1572
p-value	0.0154%	0.0927%	67.1114%	67.4116%	0.6102%
beta-weight		0.9236	-0.1433	-0.1489	1.1446
variance inflation		3.2307503	6.8180242	7.5130753	8.1687435

	Nop. Ad*Dum Copy	Nop. Ad*Dum Seg	Dum Copy*Dum Seg
coefficient	-2.252673154	0.547583423	-2.890858226
std error of coef	0.900098374	0.535654043	2.457074897
t-ratio	-2.5027	1.0223	-1.1765
p-value	2.3546%	32.1862%	25.6586%
beta-weight	-1.4283	0.3472	-0.2644
variance inflation	20.243799	7.1693675	3.1397057

standard error of regression	2.941576725
R-squared	74.26%
adjusted R-squared	63.00%
number of observations	24
residual degrees of freedom	16
t-statistic for computing 95%-confidence intervals	2.1199

Note that the coefficient on Nopane Ad*dum copy is highly significant (p-value = 0.0235) despite the fact that multicollinearity has greatly reduced the precision of the estimate (VIF = 20.24). It seems that increased ad spending is really only helpful when running the "rational" copy and doesn't do much (or even hurts somewhat) when running the "emotional" copy. Below are the predictions for different policies (in Segment A) that are calculated using the above regression.

Policy: $8.00 ad spending, "rational" copy
predicted value of Unit Sales   39.54994

Policy: $4.75 ad spending, "rational" copy
predicted value of Unit Sales   31.47646

Policy: $2.50 ad spending, "rational" copy
predicted value of Unit Sales   25.88713

Policy: $8.00 ad spending, "emotional" copy
predicted value of Unit Sales   29.4741

Policy: $4.75 ad spending, "emotional' copy
predicted value of Unit Sales   28.7218

Policy: $2.50 ad spending, "emotional" copy
predicted value of Unit Sales   28.20098

These results suggest that the "rational" copy and the highest level of ad spending would lead to the highest sales. Doing the profit calculations with these sales numbers shows that this policy is also best in terms of profit: it is predicted to yield a profit of 39.55 * ($10) - $8.00 = $385.50 per 100 prospects in Region A.

Note that we could have also added slope dummies involving the competitors' ad spending. These do not come out significant and the regression with them leads to similar conclusions to the above.

We can also add slope dummy variables to the specification in Regression 3 (appropriate if competitors' behavior is as in the experiment). The regression results are below:

Regression: Unit Sales

	constant	Ad Dollars	Dum Seg	Dum Copy	Nop. Ad*Dum Copy
coefficient	21.3748865	1.99182561	-1.7559037	11.3810173	-2.566757493
std error of coef	2.56791571	0.44949874	3.11572013	3.11572013	0.519036442
t-ratio	8.3238	4.4312	-0.5636	3.6528	-4.9452
p-value	0.0000%	0.0366%	58.0410%	0.1970%	0.0123%
beta-weight		0.9499	-0.1855	1.2021	-1.6274
variance inflation		3	7.0694823	7.0694823	7.0694823

	Nop. Ad*Dum Seg	Dum Copy*Dum Seg
coefficient	0.57493188	-2.666666667
std error of coef	0.519036442	2.343660125
t-ratio	1.1077	-1.1378
p-value	28.3422%	27.0978%
beta-weight	0.3645	-0.2439
variance inflation	7.0694823	3

standard error of regression	2.870385719
R-squared	73.96%
adjusted R-squared	64.77%
number of observations	24
residual degrees of freedom	17
t-statistic for computing	
95%-confidence intervals	2.1098

Observe that the coefficient estimates are roughly similar to those in the regression we used for prediction above. In fact, if we calculate predictions for the various policies we again find that the policy which gives the highest predicted profit in Segment A is to use the "rational" copy and spend $8.00 per 100 prospects on advertising.

Policy: $8.00 ad spending, "rational" copy
predicted value of Unit Sales   40.15304

Profit = 40.153*($10) - $8.00 = $401.53 - $8.00 = $393.53 per 100 prospects per six-months in Segment A.

A similar analysis for Segment B territories also suggests that "rational" copy and $8.00 ad spending is best. Observe that with the more complete specification, we have also rendered moot the question of whether Ms. Silk's or Mr. Skamarycz's hypothesis about competitors' behavior is correct. Under either assumption, the data suggest that using the "rational" copy and a high level of ad spending is best.

5. Instead of a single national campaign, Ms. Silk knows that it would be possible (though more costly) to have one campaign for the East and West Coast states and another for the middle of the country. Comment on the desirability of splitting up the campaign.

It only makes sense to split up the campaign if the best strategy for the coasts is different from the best strategy for the middle of the country. Does this seem to be the case here? The analysis and calculations done in question 4 above suggest that it doesn't make any difference: in both regions you would want to run the "rational" copy and spend $8.00 per 100 prospects.

Note: If we consider only the regressions given in the case write-up, regressions 1 and 3, since the segment dummy variable does not interact with any of the policy variables, the segment information cannot change the ordering of predicted performance of the different strategies. Why? Because the segment dummy is an intercept dummy variable, and its only effect is to move the regression up or down by a constant. If $8.00 spending and "rational" copy performs 0.35 units better on the coasts than in the middle, then the regression would also predict that $2.50 spending and "emotional" copy would perform 0.35 units better on the coasts than in the middle. Therefore, the regressions in the case write-up do not allow us to answer this question.

The specifications used in question 3 that include variables that interact the region with the strategy: Nopane Ad spending * Dum Segment and Dum Copy * Dum Segment did let us address this issue. If the policy that is best is going to change from region to region the change must be captured by these slope dummy/interactive terms.

Additional class question: Do the estimates of the coefficients on the interactive segment terms given above indicate a significant regional difference in relative strategy effectiveness?

Looking at the two regressions, the preliminary answer seems to be no. The t-statistics and p-values for these two variables indicate that neither can be shown to differ from zero at a reasonable level of significance. Since we are considering the joint significance of two variables, we should calculate an F-test as well. Kstat does this for us and gives the following output for the regression model with competitors' ad spending included:

Analysis of variance

	base model sum of squares	df	extended model sum of squares	df	difference sum of squares	Df
regression	377.8353678	5	399.3873552	7	21.55198744	2
residual	159.9979656	18	138.4459781	16	138.4459781	16
total	537.8333333	23	537.8333333	23	159.9979656	18
F-ratio	8.5014		6.5938		1.2454	
degrees of freedom	(5, 18)		(7, 16)		(2, 16)	
p-value	0.02822%		0.08960%		31.42875%	

Similarly, for the model without competitors' ad spending we get the following output:

Analysis of variance

	base model sum of squares	df	extended model sum of squares	df	difference sum of squares	df
regression	376.9925068	4	397.7683924	6	20.77588556	2
residual	160.8408265	19	140.064941	17	140.064941	17
total	537.8333333	23	537.8333333	23	160.8408265	19
F-ratio	11.1335		8.0463		1.2608	
degrees of freedom	(4, 19)		(6, 17)		(2, 17)	
p-value	0.00801%		0.03171%		30.86255%	

The p-value of the F-test in the two models are 0.3142 and 0.3086 respectively. So the F-tests are not significant at a reasonable level.

Thus, we don't see any strong evidence that the region of the country makes a statistically significant difference in the relative merits of different advertising strategies (copy and/or ad spending)

 (**Additional comment: In what sense did the competitors' interfere with the experiment?**

According to Regression 1, the estimated coefficient on Dum Copy is 2.134 with a standard error of 2.027 (p-value about 0.306). Thus the regression provides only weak evidence in favor of the "emotional" copy. Why such an insignificant coefficient estimate? It could be that the true coefficient is close to zero. However part of the explanation may well lie in the "interference" run by the competitors. How did they interfere with the experiment? By varying their advertising as a function of the Nopane ad copy and spending levels, they created correlation among the X variables in the regression. This can be seen in Regression 2, which shows that about 57% of the variation in competitors' ad spending can be explained by the variables manipulated in Ms. Silk's experiment. It can also be seen in the correlation matri**x,**

Correlations

	Ad Dollars	Competition	Dum Seg	Dum Copy
Ad Dollars	1.00000	0.29332	0.00000	0.00000
Competition	0.29332	1.00000	0.06993	0.69295
Dum Seg	0.00000	0.06993	1.00000	0.00000
Dum Copy	0.00000	0.69295	0.00000	1.00000

or by examining the variance inflation factors (VIFs) calculated by Kstat for the regression. Recall from the course packet that a VIF for a regression coefficient tells you how many times smaller the estimated variance of the coefficient would have been if there were no correlation in the X variables.

Regression: Unit Sales

	constant	Ad Dollars	Competition	Dum Seg	Dum Copy
coefficient	32.5937779	1.477158685	-0.56517516	0.35141056	2.13367437
std error of coef	2.52888749	0.33793489	0.162221831	1.40053429	2.02746886
t-ratio	12.8886	4.3711	-3.4840	0.2509	1.0524
p-value	0.0000%	0.0329%	0.2484%	80.4576%	30.5825%
beta-weight		0.7045	-0.7825	0.0371	0.2254
variance inflation		1.2005933	2.3315501	1.0114019	2.1195549

standard error of regression	3.411202373
R-squared	58.89%
adjusted R-squared	50.24%

number of observations	24
residual degrees of freedom	19

t-statistic for computing

95%-confidence intervals   2.0930

We see that Dum Copy has a VIF of 2.1, meaning that the variance of the coefficient estimate is more than twice as large (standard error is about $\sqrt{2}$ or 1.4 times as large) as it would have been with uncorrelated X's. So the standard error of Dum Copy has been inflated from 1.43 to 2.027 by multicollinearity. This is the sense in which the competitors' behavior "interfered" with the experiment: the correlation it created makes the coefficient estimates less accurate (i.e., have larger standard deviations) than they otherwise would have been.)