Developments In Business Simulation & Experiential Exercises, Volume 20, 1993
A MULTIPLE REGRESSION CASE IN EXPERIENTIAL LEARNING

Craig G. Harms, University of North Florida

ABSTRACT

This paper presents the development of a vertically integrated case in experiential learning using multiple regression to analyze a large database. Rather than giving students a set of data and having them perform a statistical analysis, this paper presents a vertically integrated approach to a marketing research case where the students are the sole participant. The case begins with inception of the model and continues through statistical analysis of the results and evaluation of the model.

INTRODUCTION

The following questions were recently posed to a class of students taking a quantitative methods business course: What independent variables are related to how much television a person watches during an average week? Can we hypothesize a quantitative model? Can we test this model with empirical data? Once the model is validated, can we use it to predict the number of hours of television a person might watch in a given week? Can this type of model be used in other modes of marketing research? Instead of just discussing possible answers to the above questions, this class proceeded through an entire analytical process from model development to statistical analysis and finally to model evaluation. This process is summarized in Table One.

TABLE ONE
"VERTICAL INTEGRATION" STEPS IN THE MARKET RESEARCH CASE

1. General Overview of Causal Model
      We will use multiple regression analysis. Although factor analysis may be a more appropriate technique, the course parameters required multiple regression to be taught.
2. Develop a set of independent variables and hypothesize the quantitative model
3. Develop "rules of the game" to execute the research model
4. Perform the survey
5. Collect the data
6. Construct the database
7. Analyze data
8. Draw conclusions
9. Compare results with models developed by previous classes

RESEARCH SETTING

In order for the students to be fully involved and the case self contained, the subjects of the experiment were the students enrolled in the class. The class realized that the causal model they developed would not replicate the general population in the city. Recognizing this limitation, the causal model was, What independent variables may be used to predict the number of hours of television that a local undergraduate student will watch during an average week?'

The students developed the model, participated in the survey by keeping a television log for twelve weeks, and performed a multiple regression analysis on the empirical database. After each student developed a multiple regression equation using a subset of the database, results were compared with multiple regression models developed in previous semesters. The case lasted the entire term, with classroom time being used during the first and last week of the term. The students spent the middle twelve weeks keeping their logs, i.e., collecting data for computation of the dependent variable.

DEVELOPING THE CAUSAL MODEL

The specific independent variables that would affect television viewing were the subject of classroom discussion. Some of the obvious variables are presented in Table Two. Some of the not so obvious variables were discussed at length. The January through May, 1991 semester included an identity (0/1) variable for "military family, a potentially Important factor given the proximity of the campus to five military bases. Obviously the military families had great concern during the Gulf War and indeed the class included a wife and a father of two army officers. The military family had CNN turned on eight to ten hours a day.

TABLE TWO
OBVIOUS CAUSAL VARIABLES INCLUDED IN TELEVISION MODEL

1. Number of paid working hours per week
2. Number of class credit hours taken during term
3. Number of televisions in home
4. Cable hookup (yes/no = 1/0)
5. Female/Male (1/0)
6. A "significant other in the house (yes/no 1/0)
7. VCR (yes/no = 1/0)
8. Subscribe to local newspaper (yes/no = 1/0)

Another much discussed causal variable was 'the number of people younger than 17 years in the house." Some student researchers believed that with young children in the home more television would be watched because of the many hours children have television occupying their time. Others disagreed saying that children in the house would result in less television watching because the family would organize other activities. The outcome of this discussion was to include this variable in the model with the understanding that analysis would be performed after database development to determine if there was a positive, negative, or insignificant correlation between the hours of television watched and the number of young children.

Thus, some of the causal variables were logical, and the student's hypotheses concerning the correlation being direct or inverse were logical. Other variables were more mysterious and students were genuinely interested in the analysis to be performed later in the term. The class agreed on a total of 15 independent variables. (This was the limit of the software database.)

RULES OF THE GAME

Several rules were established to qualify recording television time in the log. Although several seemed trivial, much discussion preceded agreement by the class. For example, if you taped a television show on your VCR during the day and watched the tape at night, did that count as television time? The class agreed that this counted in the television log. The logic was that even though you may "zap through the commercials, your mind would still see the corporation advertising, although every word might not be understood.

If the television was just "on' and you were sweeping the house, this did not count because you probably could not hear the television and probably were not even looking at it during the majority of the time. If you were sitting in a chair reading a magazine, this did count as television time because once again, your mind was seeing the television. Renting a movie and watching it did not count, even if there were advertisements at the beginning of the movie. The reasoning was that this would be a random decision, while television advertisers planned time blocks when their ads were shown.

The list of "rules" was extensive, but important to eliminate as much 'noise" from the survey as possible.

PERFORMING THE SURVEY

For twelve weeks each student recorded the time they watched television in minimum fifteen-minute blocks. No show titles were recorded, only the time when the television was on and they were in the room, close to the television, and watching." At the end of each week students added up the total time watched during the week and kept a running total at the bottom of each page.

COLLECTING THE DATA

Two weeks before the end of the term students were given a handout listing the fifteen independent variables and were asked to fill it out. At that time they turned in the television log. At the bottom of the last page was the total television time during the twelve weeks.

CONSTRUCTION OF THE DATABASE

The hours of television watched on a weekly basis--the value of the dependent variable--was computed by dividing the total hours by the twelve-week logging period. The faculty member developed the database using an editor program. The number of observations--the number of students participating in the survey--ranged from. 25 to 52 during each of the last five semesters. A data file containing the database was put on reserve in the computer lab for students to copy to their own personal disks.

DATA ANALYSIS

Students individually built a multiple regression model using a subset of the fifteen independent variables. Some of the variables had absolutely no correlation with the dependent variable. Some had weak correlation. There was no multicollinearity in the correlation matrix. Unlike textbook problems where the coefficient of determination may be as high as 0.95, students were informed that models they developed may have coefficients of determination of only 0.35 (at best)'

During the most recent semester 31 students participated in the survey. They were specifically told not to work together as they analyzed the data and built their models. In addition, they were given the following parameters to guide them in model building:

(1) Your model should contain three, four, or five independent variables from the original database of 15 variables. Thus among the 31 students a large variety of different models would be developed.

(2) The F-statistic for the model should be 3.0 or larger. Thus the E will be significant at the .05 level.

(3) The t-statistics of the independent variables would probably be marginal at best. After some preliminary work students were told to accept t-values of less than 2.0. At the .10 level the critical t-value is 1.3. Unlike most textbook problems, where t-statistics are usually three or higher, students' models will have several variables with t-statistics of less than 2.0.

(4) The adjusted R-square must be at least 0.25. Again, unlike textbook problems where the coefficient of determination is very high this problem evaluates individual human responses and thus a lower adjusted R-squared must be accepted.

(5) Finally, the level and sign of the beta coefficients of the independent variables in the model must be logical.

With these guidelines, students started on their model building process. The multiple regression software accessed the database of 15 independent variables. Students could build a multiple regression model of up to seven independent variables. The program allowed for quick and easy iterations allowing the students to quickly try and evaluate many multiple regression models.

ANALYZING THE RESULTS

Multiple regression models from the Spring 1992 term contained the best' statistical results. An example of one of the best models is presented in Table Three. Although some of the expected variables, as presented in Table Two, were not present in this model, the model certainly shows that the more people under seventeen years of age in a home, the more television the average student watches. Of the other three independent variables in this model, the number of outside organization memberships and the subscription to the local newspaper have explainable beta coefficients.

TABLE THREE
MULTIPLE REGRESSION FROM SPRING, 1992 TERM

Anova Table

Source	Sum of Squares	df	Mean Square	F
Model	665.786	4	166.446	5.285
Error	629.882	20	31.494	
Total	1295.668	24		

Variable Name	Estimated Coefficient	T-statistic
Intercept	21.039	
# of TVs in home	-2.3239	-2.3812
# of outside organization memberships	-3.4952	-2.5616
Subscription to local newspaper (1 = yes)	-2.4883	-1.0217
# people in home less than 17 years old	5.4661	3.0004

R-squared = 0.515   r = 0.717
Standard error of estimate = 5.612 (hours)
Adjusted R-squared = 0.386

Although the t-statistic on the newspaper variable is very weak, it is a "crack filler." The students were very surprised to see a negative beta coefficient for The number of televisions in the home" variable, a result that sparked much classroom discussion. There were more than 20 different multiple regression models developed by the students using the same database. None of the models contained multicollinearity among the independent variables, and many, many of the models included the "number of televisions in the home" variable as a statistically significant variable. No one, including the faculty member, could logically explain the coefficient's negative beta coefficient.

THE NEED TO DRAW CONCLUSIONS

From this case students were able to see how their future corporation might use multiple regression to help predict the hours of television watched during a given week, and thus plan advertising expenses. Obviously, the class was a very narrow sample with very special characteristics. Therefore the causal model that each student developed was skewed toward undergraduate students and not the general public. But the reason for using the students as researchers subjects was to help the students understand the total 'vertical integration" of the process.

COMPARISON OF RESULTS FROM PREVIOUS TERMS

Of particular interest was the comparison of models developed for a Fall term versus a Spring term. Table Four summarizes the comparison.

TABLE FOUR
COMPARISON OF FALL AND SPRING TERM MODELS

Fall Term Model--Independent Variables:

Variable	Beta Coefficient	t-statistic
Intercept	6.786	
Age	0.322	1.638
Paid hours of work	-0.136	-1.866
Female = 1; Male = 0	-5.335	-2.601
Cable hookup? Yes = 1	3.306	1.379
Game system? Yes = 1	-3.012	-1.345

Spring Term Model--Independent Variables:

Variable	Beta Coefficient	t-statistic
Intercept	11.266	
Paid hours of work	-0.186	-2.847
Credit hours taken	0.062	0.242
Female = 1; Male = 0	3.227	1.449

These two models were developed independently by different students in the respective classes. Therefore some of the variables are different. Students were asked to explain why the Female/Male identity variable had a negative beta coefficient for the fall term model and a positive beta coefficient for the spring term model. All other things being equal, men seem to watch about 8.5 more hours of television per week in the fall than females. The men in the class explained this phenomenon immediately. Males watch college and professional football in the Fall! (The women concurred and rolled their eyes.)

FACULTY EVALUATION

This football phenomenon seemed to validate the use of multiple regression in the minds of the students. Everyone, especially the wives and the female "significant others," knew football to be a "time-sink" from September through January. They could see that the quantitative model did confirm this suspicion. This faculty member could see the "light come on" in the students' minds.

After using this case for the last five semesters, I am convinced that the time and effort was worthwhile. Students were able to understand the marketing research function from the viewpoint of the researcher and the participant. By spending considerable time on the project, meaningful learning did take place and students will be able to replicate this process on their own later in their job setting.

CONCLUSION

Experiential learning is a key to keep the United States the world leader in business and industry. "Hands-on" and "real-world" cases such as the television survey analysis case are excellent and self-contained methods of meaningful learning for students.

REFERENCES

Coleman, B. J. (1990) The Analysis of Statistical Relationships, (93 page booklet)

Neter, J., & Wasserman, W. (1974) Applied linear statistical models. Richard D. Irwin