Using SAS® to Determine the Factors that Label College Students as "Computer Hackers" at two Los Angeles Universities

Annaliza Dialino, Loyola Marymount University, Los Angeles CA

Abstract

The purpose of this paper is to discover, with the use of SAS®, the factors that best determine if a college student is a "computer hacker"; a person who spends the majority of their leisure time using a computer, and takes pride in being able to successfully "hack" into a computer software. The methodology used was a personal characteristic survey distributed at two Los Angeles based universities, Loyola Marymount University and the University of California, Los Angeles. An econometric model was then developed specifying the number of hours spent on the computer as the dependent variable. A sample of the independent variables used (i.e., regressors), includes grade point average (GPA), gender, and a question requesting the number of hours spent on the computer per week. These hours were then allocated to correspond to specific programs such as the Internet, WordPerfect, or a programming language. A Likert-scale was also used to rank the importance of several college activities (e.g., clubs, sports, and the Greek system).

Introduction and Methodology

First, it should be noted that the primary focus of this paper is to determine the factor(s) that make one person more likely to "hack" on a computer than another. The term "hacker" is not intended to represent a derogatory statement about individuals who break-in or illegally "hack" into hardware, software, or other people's computers. Rather, in this paper, it is used simply to label those students or individuals who enjoy spending a considerable amount of time using a computer to program, word process, or simply play games. It does not necessarily imply that an individual who "hacks" is performing an immoral or illegal act. In this paper, the term "hacker" is inspired by the "heroes of the computer revolution" at MIT, documented in Steven Levy's book called "Hackers".

Ninety-eight observations were obtained from the campuses of Loyola Marymount University and the University of California, Los Angeles to be regressed with the twenty-three variables initially constructed in a personal characteristic and Likert-scale survey. A multiple regression was conducted to determine which variables most contribute to an individual's computer usage. The PROC Stepwise procedure was implemented to represent the order of the variables importance and influence to the dependent variable. The econometric model was formulated using the number of hours spent on the computer as another method to distinguish the "hacker". Simply stated, the following model was used to identify the factors that most influenced an individual's computer usage.

Number of Hours spent on a computer=
***f* (GPA, sex, age, SAT score, your major, sleep required, etc.)**

The variables used in the model were: sex (or gender); age; grade point average (GPA); Standardized Aptitude Test (SAT) score; hours spent using the computer each week; whether those hours are spent on word processing, programming, the Internet, or Bitnet; if you work, and if so, if it is in a computer lab; the number hours of sleep you require daily; whether your major is in the College of Liberal Arts, Communications and

Arts, Science and Engineering, or Business Administration; Math grade point average (GPA); English grade point average (GPA); whether you are a Computer Science major; and finally, the Likert-scale rating the importance of the following factors in your life (1 being most important to 5 being least important); having a boyfriend or girlfriend, being in a fraternity or sorority, being in a club or organization, or having a part-time job.

It should also be noted that dummy variables were used in several of the questions, most of which were yes or no questions (e.g., 1 for NO and 0 for YES, while 0=MALE and 1=FEMALE) and a sequence of dummy variables were used to identify which college your major belongs to (i.e., Liberal Arts, Business Administration, etc.). More specifically, the data for this category was formulated using Business Administration as the default major. In this case, all zeros for each corresponding college defaults the individual as having a major in the College of Business Administration.

Analysis and Conclusion

After running the PROC Reg and the PROC Stepwise to determine the most influential factors that identify an individual as a "hacker", the following results were observed in Model 1:

Model 1

Model: MODEL 1
Dependent Variable: HRSCOMP

Analysis of Variance

	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	21	6234.23101	296.8681	3.355	0.0011
Error	31	2742.95767	88.48251		
C Total	52	8977.18868			

Root MSE	9.40651	R-square	0.6945
Dep Mean	10.8584	Adj R-sq	0.4875
C.V.	86.62819		

With the dependent variable being the number of hours spent on the computer, and the independent variables being the items previously listed in the introduction, the R-Square denoted that sixty-nine and forty-five hundredths percent (69.45%) of the data in the observations accounts for and corresponds to the econometric model. The F-value in Model 1 further supports a significance in the econometric model used. The F-value of three and three hundred and fifty-five thousandths (3.355) had a P-value attached of eleven ten thousandths (0.0011). Since the P-value is less than five hundredths of a point (0.05), this indicates that the econometric model as a whole is statistically significant. Furthermore, the parameter estimates illustrated in Model 2 provide further evidence that several of the independent variables used in the model were significant.

Model 2

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO	Prob>|T|
INTRCPT	1	-31.5718	25.4443	-1.241	0.2240
SEX	1	-6.25786	3.46497	-1.806	0.0806
AGE	1	1.246322	0.92063	1.354	0.1856
GPA	1	10.74436	5.01232	2.144	0.0400
SAT	1	0.000901	0.01039	0.087	0.9315
WP	1	0.061349	0.07300	0.840	0.4072
PROG	1	0.178240	0.10431	1.709	0.0975
INTRNET	1	-0.001901	0.10859	-0.018	0.9861
BITNET	1	0.143713	0.10166	1.414	0.1674
WATLAB	1	21.99317	6.44867	3.410	0.0018
SLEPHRS	1	0.367294	1.01256	0.363	0.7193
COLLIB	1	9.977709	4.81237	2.073	0.0465
COLSCI	1	5.607820	5.75183	0.975	0.3371
COLCA	1	9.221133	9.01640	1.023	0.3144
MTHGPA	1	-2.861564	3.95392	-0.724	0.4747
ENGGPA	1	-4.943506	4.04451	-1.222	0.2308
COMSCIM	1	-3.794748	7.39394	-0.513	0.6114
BOYGIRL	1	-0.412781	1.38426	-0.298	0.7675
FRATSOR	1	0.348178	1.27069	0.2745	0.7859
CLUBORG	1	0.694418	1.42391	0.4881	0.6292
JOB	1	-2.302271	1.35638	-1.326	0.1944
COLLEGE	1	-4.679876	3.57909	-1.308	0.2006

These parameter estimates display which regressors correspond to the number of hours an individual spends using a computer. With the T-values, it can be determined which of the variables were significant simply by examining which of them had a T-value of greater than or equal to the absolute value of

two (>=2). In this case, the variables GPA, whether you work at a computer lab, and whether your major is categorized under the College of Liberal Arts, were most significant in affecting the number of hours you "hack" or spend using a computer. Therefore, these are the three most influential variables that determine whether an individual is more likely to "hack" on computers than others.

Number of hours spent on a computer =
$B_0 + B_1$ **(GPA)** $- B_2$ **(sex)** $+ B_3$ **(age)...etc.**

As an example with the corresponding values, this econometric equation can now be written as:

Number of hours spent on a computer =
-31.57+ 10.74(GPA) -6.25(sex) + 1.24(age)...etc.

Furthermore, by corresponding the parameter estimates to each variable, it may also be assumed that the variable WATLAB is significant to the model. More specifically, the variable WATLAB has an intercept of approximately twenty-one and ninety-nine hundredths of a point (21.99). This indicates, ceteris paribus, that if you work at a computer lab, you will tend to spend approximately twenty-one point ninety-nine hundredths (21.99) more hours "hacking" on a computer than a person who does not work at a computer lab.

References

Gujarati, Damodar. *Basic Econometrics*. Second edition, McGraw-Hill, 1988.

Levy, Steven. *HACKERS*. Bantam Doubleday Dell Publishing, 1984.

SAS Institute Inc. *SAS System for Elementary Statistical Analysis 1987.*

SAS Institute Inc., *SAS/ETS User's Guide*. 1988 Version 6, First Edition, Cary, NC: SAS Institute Inc., 1988.

SAS Institute Inc., *SAS/ETS Software: Applications Guide 1, Version 6*, First Edition, 1991.

SAS Institute Inc., *SAS/STAT User's Guide*, Vols I and II, 1990 Version 6, Fourth Edition. Cary, N.C.: SAS Institute Inc.

Acknowledgments

I would like to thank, Dr. Joseph Earley, for his motivation, patience, and dedication. I would also like to thank, Scott Matthew Miller, for his love and support.