IJCTA, 9(17) 2016, pp. 8415-8422
© International Science Press

Forecasting and Hypothesis Testing for Cyberloafing in Organizations Using Multiple Linear Regression

Soham Banerjee*[*] and Sanjeev Thakur[*]

ABSTRACT

Business is an arena where organizations strive to rule the market. Employees use company resources to ensure maximum productivity. With the recent uprisings in cybercrimes, organizations have heavily invested on employee surveillance systems to monitor employee usage behavior of valuable resources like data, internet, application software's etc. Cyberloafing basically deals with employees who use company resources for personal use and slack work during office hours. In this paper we have developed a forecast model based on multiple linear regression that will be applied over a set of factors generated by user responses. Based on the data set and p-values from regression analysis we will try to approve or disapprove certain hypothesizes based on cyberloafing.

Keywords: Forecast, Cyberloafing, Multiple Linear Regression, Hypothesis testing

I. INTRODUCTION

Employees form the competitive foundation for any organization and play a vital role to crown the organization they work in. Productivity has a direct dependency on employee performance and in the globalized network, organizations continuously monitor their employees and look forward to process improvement. But in the due process of hectic market dominance, employees often avoid or don't follow company policies, especially in IT industries where the core asset is information. Among many employees, some try to adopt malpractices for personal benefit. For e.g. some employees send spam emails or try to access unauthorized data. Cyberloafing has now become an increased menace in the organization which is a threat to organizations with respect to cost and productivity.

Cyberloafing or goldbricking is a term used to define actions of employees who use organization resources especially, internet to complete their own personal errands during organization work hours. Thus it is important for organizations to identify employees who misuse internet usage and create opportunities to tamper with confidential data or harass other employees. Cyberloafing is based on human behavior which changes over time depending how the various work environment variables change over time. Psychological factors like personality, age etc. also have a great emphasis on cyberloafing [11]. However it has been observed that most forecasting techniques especially multi linear regression generate low R-squared values when predicting human behavior and almost in most cases this value is less than 50%. [10].

Therefore there is now a dire need of applications that can predict whether the employees of an organization are satisfied with their job or will they misuse resources. Employee surveillance systems just monitor the current state of usage and provide simple statistical support to analyze data. Also an organization has two opportunities to assess any employee or candidates during interviews for cyberloafing traits. If any organization has a standard application that can predict an employee's characteristic and based on pre-

[*] Department of Computer Science and Engineering, ASET Amity University, Noida, Uttar Pradesh, India, *E-mail: official.soham@gmail.com; Sthakur3@amity.edu*

defined weights can classify the existing employees as to whether they will use or misuse the resources assigned to them, then it will help managers to either penalize the employee or keep a track of the employee's activity for misuse.

The paper has been divided into three sections. In the first section we will follow a short overview of the technique used by us to predict an employee as a cyber-loafer or whether he is satisfied with the job. In the second section we will apply multiple linear regression over user responses followed by discussion of results. The final section will draw conclusion and future scope from our work.

II. LITERATURE REVIEW

In this section we will discuss and interpret multi linear regression and null hypothesis testing in detail along with small descriptions of the various work environment variables identified through user responses. Our work will be an extension on the work done by Banerjee et al [9].

2.1. Multiple Linear Regression

Before we move on, let us first understand what do we mean by the word regression. Regression or regression analysis is simply a statistical technique that is used for making prediction or forecast based on some independent variables over a dependent variable. To make things clear, regression basically estimates relationship among variables [17]. Regression analysis is basically used in wide spread applications to build forecasting models that are basically used to estimate or predict certain events. For example one can use age and IQ as independent variables and predict the CGPA of a student. Regression analysis contains many techniques, but our focus is particularly on multiple linear regression [14].

Consider X_1, X2.......... X_n as independent variables that are not correlated in any form of relationship with each other. Now we have a dependent variable Y which we need to estimate. Then we will fit a line along a scatter plot whose equation will be of the form of

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + ...\alpha_n X_n + \varepsilon \tag{1}$$

Where ε is the intercept that represent the residual error and α_1, α_2, ...α_n are regression coefficients of the independent variables X_1, X2.......... $X_{n.}$ One can now clearly estimate the value of Y based on the line fitted on the scatter plot generated by X and Y values as per equation (1) [1]. When we perform multiple linear regression over observational data, we test the quality of the model by computing the coefficient of determination. Whenever we perform multiple linear regression we have to understand the following terms:

- ***R^2* (R-squared):** It is the coefficient of determination that measures how well the data fits the regression line in a scatter plot. It is denoted in percentage between 0 - 100%.
- ***Adjusted R^2:*** It is simply the R^2 that has been adjusted based on the number of predictors in the regression equation. Since R^2 values are very sensitive to even one data point and cannot determine whether the estimation or predictions are biased. Hence adjusted R^2 will give an unbiased prediction based on R^2 population.
- ***Predicted R^2:*** This measure comprehends that how well a regression model is able to make predictions based on new observations or data. It basically helps to judge whether a regression model is suffering from the problem of over fitting.

2.2. Hypothesis Testing

In real world scenarios people often make assumptions and provide their interpretation of events through justification and proofs. In case of statistics, however the situation is different. A statistical claim can be validated using hypothesis testing [19]. Hypothesis testing is often applied over parameters that are population based. For e.g. if a doctor claims that if a patient has sugar levels equal to 200 then he is suffering Type 2 diabetes. It might be possible that the value can be less or more or not equal to 200 .One can validate this statement by determining the alternate hypothesis for the same and test using statistical significance measures to prove which one is appropriate [20]. Whenever we perform hypothesis testing for any claim, we address two types of hypothesis as follows:

- ***Null Hypothesis (H_{null} or H_0):*** A null hypothesis generally denotes a claim that exists and there is nothing new to be observed. Null hypothesis basically denotes that a particular parameter which is based on population is equal to the value that is being claimed. For e.g. if we say that the minimum temperature required to convert water into ice is 4 degree centigrade. Then the null hypothesis in shorthand will be denoted as H_0: $\mu = 4$.
- ***Alternative Hypothesis (H_{alt} or H_a):*** When null hypothesis is rejected, one must develope another set of hypothesis which is called alternative hypothesis. In this case the claimed value cannot be equal to the population parameter. Hence when generating an alternate hypothesis, three possible sets can be generated. For e.g. if the null hypothesis fails for H_0: $\mu = 4$, then alternate hypothesis would be H_a: $\mu \neq 4$ or H_a: $\mu > 4$ or H_a: $\mu < 4$.

Based on what we wish to conclude in our data set we can use alternate hypothesis for a new claim. Now the question arises as to select which H_a. We can select the appropriate alternate hypothesis using *p-values* and *T test statistics* that are generated during multiple linear regression. We retain null hypothesis if and only if the p-value is greater than 0.05 level of significance. Otherwise we can suggest an alternative hypothesis with proof.

2.3. Organization Work Factors

Employees in an organization handle multiple tasks of various complexities and often look for opportunities to showcase their work in front of the management. While employees look forward to salary raises and growth as key factors, learning as well as complexity of the task also effect an internal work environment within the organization since solving a task is purely based on employee skills and judgment. Based on such factors we have listed all the work related factors that will help us to determine whether the employees are satisfied or they will cyberloaf as follows:

- ***Work complexity:*** Employees often handle multiple tasks of different complexities. Harder the complexity, greater will be the challenge to resolve the work. In every organization irrespective of its standards there are certain employees who might not be able to perform the task due to high complexity [9]. This often distinguishes a good employer from an ordinary one and often managers assign complex tasks to employees who they think can do the job. This creates an opportunity for ordinary employees to slack from work or complete their allotted tasks later. The null and alternate hypothesis are:

  H_0: The work complexity faced by any cyberloafer is equal to 64.3% versus H_a: The work complexity faced by any cyberloafer is greater than 40%
- ***Growth:*** Every employee wants to rise in his career path. Authority and designation play a vital role and every employee dreams to reach the top management positions with relevant experience and credibility. Often managers and human resource executives judge the performance of employees and offer promotions to a select few [9]. Those who get the opportunity to climb will put more effort at work as compared to those who will get demotivated and may waste time in the internet rather than improving themselves [11]. Employees with good self-motivation and confidence can replenish their efforts towards work. The null and alternate hypothesis are:

H_0: The growth in career of any cyberloafer is equal to 44.5% versus H_a: The growth in career of any cyberloafer is more than 35.5%.

- ***Warnings:*** Some employees are often warned when caught by managers or IT executives when misusing the internet. Also when employees don't perform their tasks, then management uses warnings to question the employee's contribution in the overall productivity [4]. This is specially a common phenomenon in IT industry where employees are in constant pressure to deliver under strict deadlines. Warnings decrease employee moral however it is necessary to ensure that the organization and its employees follow polices and standards on strict compliance [5]. The null and alternate hypothesis are:

  H_0: A cyberloafer is warned at an average equal to 66.7% of the times he performs activities over the internet versus Ha: A cyberloafer is warned at an average less than 50% of the times he performs activities over the internet.

- ***Perks:*** Often organizations provide non cashable benefits to employees that motivate them to perform better and dedicate more time. Perks can be in the form of organizational gadgets, bill reimbursements, LTC etc. Perks are great motivators to improve employee morale, but in many scenarios it has been noticed that employees prefer raise in salary rather than perks since cash speaks more. The null and alternate hypothesis are:

  H_0: An employee will leave cyberloafing if he receives perks equal to 54.5% of the total perks covered in his package versus H_a: An employee will leave cyberloafing if he receives perks more than 54.5% of the total perks covered in his package.

- ***Learning:*** Employees join industries not only to run their stomach but to inherit on the job skills that will distinguish them from the others. A confident and motivated employee will learn more irrespective of talent. Small scale industries to large enterprises, every organization favors the fact that there should be a continuous process of learning and implementing skills [5]. Employees who cyberloaf generally make an excuse to learn using the internet and are more interested abusing it. Certain traits help managers distinguish between the lazy and the inquisitive like being proactive or accountability. The null and alternate hypothesis are:

  H_0: An employee will leave cyberloafing if he learns new skills by spending 60.2% of the total time he works in the office versus H_a: An employee will leave cyberloafing if he learns new skills by spending more than 60.2% of the total time he works in the office.

- ***Raise:*** Organizations follow performance based appraisals for employees who work hard to achieve a collective goal. Just like perks an appraisal or raise in the salary in the form of bonus can really boost the confidence and morale of employees working in the organization. Most cyberloafers pretend to work hard but the result of their work is judged based on organizational needs. In the age of globalization organizations have adopted micro level of performance as an indicator to judge an employee's proficiency in his field. In the context of most small level IT companies and SME's, low wages often force employees to leave their jobs as appraisal levels are extremely insignificant. The null and alternate hypothesis are:

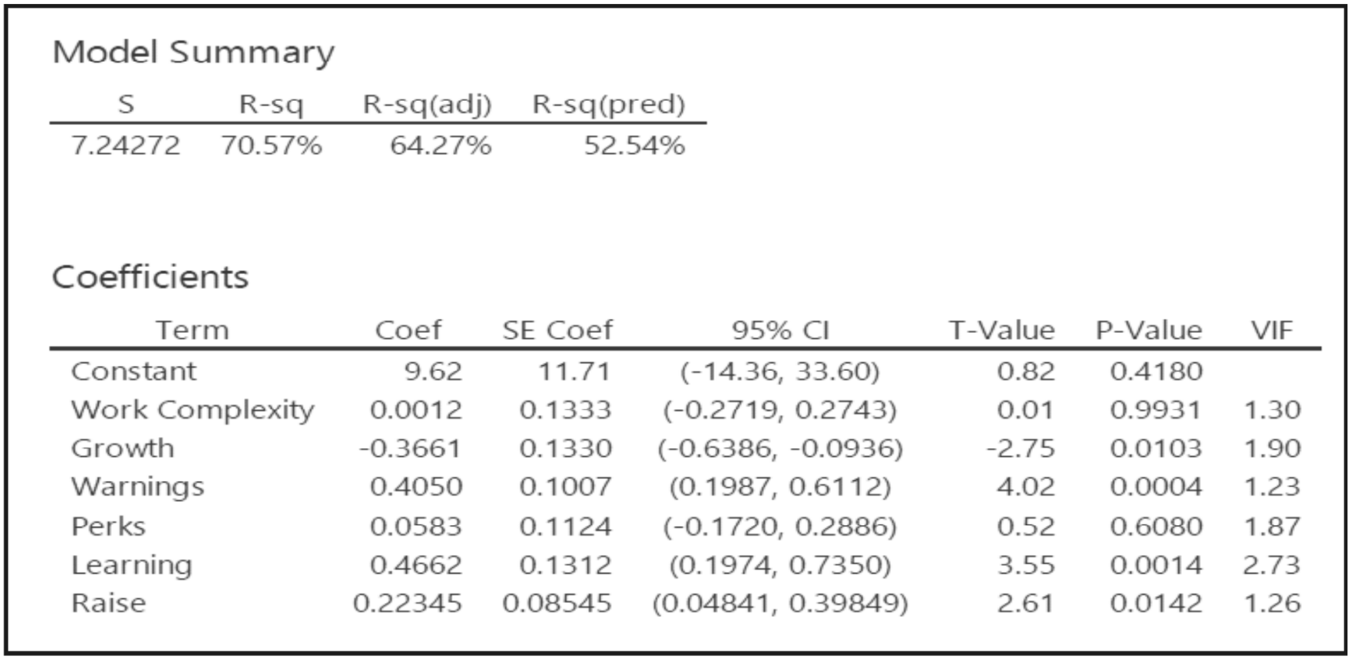
  H_0: An employee will leave cyberloafing if he gets a onetime raise of 61.8% of the total salary he earns in the office versus H_a: An employee will leave cyberloafing if he gets a onetime raise of more than 61.8% of the total salary he earns in the office.

III. EXPERIMENT AND RESULT

Based on the factors discussed in the previous section, we prepared a questionnaire for 70 participants who are working in the same organization. Out of the 70 participants, 58 responses were selected due to their completeness. 59 % males and 41% of the females had participated between the ages of 22 - 55 years. Each participant was asked to assign a score for all factors between 0 - 100. The discussed factors will be treated as independent variables for regression while the dependent variable is an overall rating w.r.t cyberloafing. Next we apply multiple linear regression over the collected responses. We have used Minitab statistical software which provides a wide range of statistical measures and provides creative visualization of results. The data can be loaded directly through any CSV or XLSX file.

3.1. Results Found

After applying the multiple linear regression model we receive the regression equation, summary for the model that contains the value of R^2, adjusted R^2 and predicted R^2. Along with these values a coefficient table is also generated with T-values and p-values as shown in Figure 1.



Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
7.24272	70.57%	64.27%	52.54%

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	9.62	11.71	(-14.36, 33.60)	0.82	0.4180	
Work Complexity	0.0012	0.1333	(-0.2719, 0.2743)	0.01	0.9931	1.30
Growth	-0.3661	0.1330	(-0.6386, -0.0936)	-2.75	0.0103	1.90
Warnings	0.4050	0.1007	(0.1987, 0.6112)	4.02	0.0004	1.23
Perks	0.0583	0.1124	(-0.1720, 0.2886)	0.52	0.6080	1.87
Learning	0.4662	0.1312	(0.1974, 0.7350)	3.55	0.0014	2.73
Raise	0.22345	0.08545	(0.04841, 0.39849)	2.61	0.0142	1.26

Figure 1: Model Summary and Coefficients along with T-values and P-values of each independent variable

Based on the coefficient table generated in Figure 1 and as per equation (1), the regression equation is as follows:

Cyberloaf = 9.62 + 0.0012 (**Work Complexity)** –0.3661(**Growth)** + 0.4050(**Warnings)** + 0.0583 (**Perks)** + 0.4662 (**Learning)** + 0.22345(**Raise)** (2)

The residual plot generated will help us to validate the amount of randomness or random error in the data as shown in Figure 2.

3.1. Observations

Based on the p-values we can remove the predictor variables that are not significant for the generated regression equation. Hence we remove the predictor variables work complexity (p-value = 0.992) and perks (p-value = 0.608) since they have p-values greater than 0.05 level of significance. This gives the new regression equation as follows:

Cyberloaf = 9.979 + 0.40522 **Warnings** + 0.4991 **Learning** + 0.23848 **Raise** –0.3660 **Growth** (3)

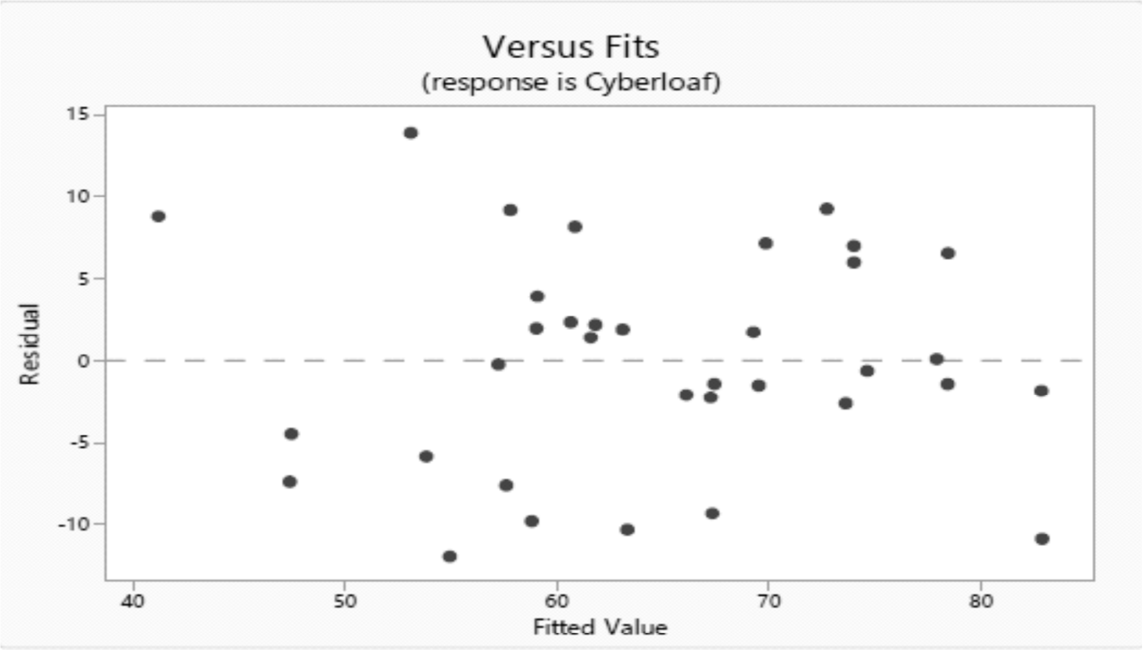
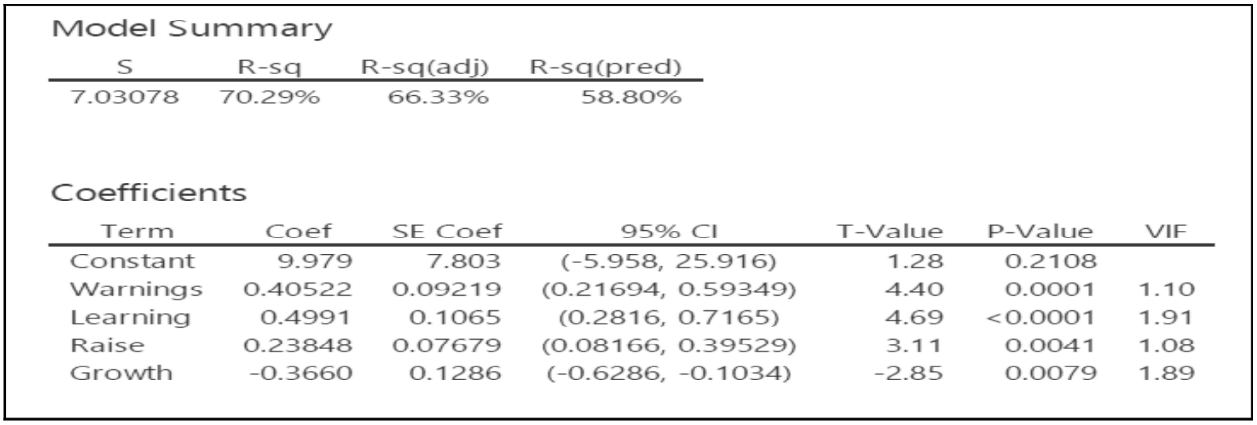


Figure 2: Residual plot that clearly shows random errors that produce residuals which are distributed normally



Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
7.03078	70.29%	66.33%	58.80%

Coefficients

Term	Coef	SE Coef	95% CI	T-Value	P-Value	VIF
Constant	9.979	7.803	(-5.958, 25.916)	1.28	0.2108	
Warnings	0.40522	0.09219	(0.21694, 0.59349)	4.40	0.0001	1.10
Learning	0.4991	0.1065	(0.2816, 0.7165)	4.69	<0.0001	1.91
Raise	0.23848	0.07679	(0.08166, 0.39529)	3.11	0.0041	1.08
Growth	-0.3660	0.1286	(-0.6286, -0.1034)	-2.85	0.0079	1.89

Figure 3: Model Summary and Coefficients along with T-values and P-values of each independent variable w.r.t to regression equation 3

The model summary and coefficients of the above equation with new p-values and T-values are shown in Figure 3. We can clearly observe that:

- All the factors in Figure 3 are significant among which *learning* is the most significant factor since it has the lowest p-value (< 0.0001) for 0.05 level of significance.
- The value of adjusted R^2 and predicted R^2 have improved as the number of predictor variables have decreased.
- Mapping human behavior based on work environment variables, we have achieved an R^2 value of greater than 50%.
- The model can forecast based on 58.8% tolerance [10].
- We can now test the null and alternate hypothesis as follows based on the p-values of predictor variables with reference to Table 1.

Table 1
Decision table for hypothesis acceptance based on p-values

Predictors	*p-value*	*HypothesisAccepted*	*Conclusion*
Work Complexity	0.9931	H_0	Work complexity is not significant for predicting cyberloafers
Growth	0.0103	H_a	Growth is significant for predicting cyberloafers
Warnings	0.0004	H_a	Warning is significant for predicting cyberloafers
Perks	0.6080	H_0	Perks is not significant for predicting cyberloafers
Learning	0.0014	H_a	Learning is significant for predicting cyberloafers
Raise	0.0142	H_a	Raise is significant for predicting cyberloafers

IV. CONCLUSION

In this paper we have discussed how multiple linear regression has been used to create a cyberloafing forecast model. We have been able to successfully retain or rejected the null hypothesis. Also we have improved the regression equation with the significant predictor variables which has improved the forecasting model as well. This work will help us to develope a possible forecasting application that will help organizations to identify cyberloafers based on the work characteristics.

REFERENCES

[1] J. Richards, "The many approaches to organisational misbehaviour." *Employee Relations*. 30, 6, 653-678 (2008),

[2] K. Askew, J. Buckner, M. Taing, A. Ilie, J. Bauer, and M. Coovert, "Explaining cyberloafing: The role of the theory of planned behavior." *Computers in Human Behavior.* 36, 510-519. (2014).

[3] J. Fichtner, and T. Strader, "Non-Work-Related Computing and Job Characteristics: Literature Review and Future Research Directions." *Journal of Psychological Issues in Organizational Culture* 4, 4, 65-79 (2014),

[4] S. Prasad, V. Lim, and D. Chen, "Self-regulation, individual characteristics and cyberloafing." *PACIS 2010 Proceedings*. 1, 1 (2010),

[5] R. Baarda, and R. Luppicini, "The Use and Abuse of Digital Democracy." *International Journal of Technoethics*. 3, 3, 50-68 (2012),

[6] O. Çinar, and F. Karcioglu, "The Relationship between Cyber Loafing and Organizational Citizenship Behavior: A Survey Study in Erzurum/Turkey." *Procedia - Social and Behavioral Sciences* 207, 444-453, (2015)

[7] L. Ivarsson, and P. Larsson, "Personal Internet Usage at Work: A Source of Recovery." *Journal of Workplace Rights*. 16, 1, 63-81.

[8] D. Rumsey, "How to Set up a Hypothesis Test: Null versus Alternative - For Dummies." http://www.dummies.com/how-to/content/how-to-set-up-a-hypothesis-test-null-versus-altern.html.

[9] S. Banerjee, S. Thakur, "A Critical Study of Factors Promoting Cyberloafing in Organizations." *International Conference on Cyber Security and Digital Forensic, ACM Proceedings* (In press) (2016)

[10] J. Frost, "Regression Analysis: How Do I Interpret R-squared and Assess the Goodness-of-Fit?" http://blog.minitab.com/blog/adventures-in-statistics/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit.

[11] H. Jia, R. Jia, and S. Karau, "Cyberloafing and Personality: The Impact of the Big Five Traits and Workplace Situational Factors." *Journal of Leadership & Organizational Studies*. 20, 3, 358-365 (2013),

[12] S. Chatterjee, and B. Price, "Regression Analysis by Example." *New York: Wiley*. (Section 3.7, p.68ff of 2nd ed. (1991).

[13] V. Lim and D. Chen, "Cyberloafing at the workplace: gain or drain on work?" *Behaviour & Information Technology*, 31, 4, pp. 343-353, (2012).

[14] T. Plotts, "A multiple regression analysis of factors concerning superintendent longevity and continuity relative to student achievement." (2011).

[15] C. Andreassen, T. Torsheim, S. Pallesen, "Predictors of Use of Social Network Sites at Work - A Specific Type of Cyberloafing." *J Comput-Mediat Comm.* 19, 906-921 (2014).

[16] L. Giangregorio, R. Cook, "Hypothesis testing in clinical and basic science research." *Transfusion.* 50, 1878-1880 (2009).

[17] L. Krantz, “An Application of Multiple Regression Analysis in Determining the Relative Contribution of Certain Components of Reading Ability in Grade and High School Achievement.” *The Journal of Experimental Education.* 23, 275-277 (1955).

[18] K. Marill, “Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression.” *Academic Emergency Medicine.* 11, 94-102 (2004).

[19] P. Sedgwick, “Pitfalls of statistical hypothesis testing: multiple testing.” *BMJ.* 349, g5624-g5624 (2014).

[20] P. Shaw, M. Proschan, “Null but not void: considerations for hypothesis testing.” *Statist. Med.* 32, 196-205 (2012).