iJETRM

International Journal of Engineering Technology Research & Management

ABSENTEEISM PREDICTION BASED ON THREE MODELS BY EXPLOITING HIERARCHICAL MULTIPLE REGRESSION ANALYSIS

Hnin Min Oo[*1]

San San Ei[2]

[*1]First Author (University of Computer Studies, Mandalay, Myanmar)

[2]Second Author (University of Computer Studies, Mandalay, Myanmar)

ABSTRACT

Absenteeism has become one of the organization's persistent issues. Different people usually understand absenteeism in various ways. It is generally recognized as an employee or group of employees who are either continually or repeatedly for short periods away from job. The research was studied using hierarchical multiple regression analysis to learn the causes of the company's absenteeism among the staff.In this system, the parameters such asdemographic characteristics are used to predict the effects of staff's absenteeism in organization.This system present relationship between demographic characteristics, service, work attitudes and absenteeism. Three different regression models have been constructed using hierarchical multiple regression analysis. A questionnaire consisting of twentyquestions relating totwo major variables was used. Questionnaireswere used in this study isdesigned to examine the relationships between workattitudesand absenteeism.This system applied seven hypotheses and proved the proposed hypotheses were supported or not in order to help the managers to better manage staffs' absences, resulting in increased productivity.Hierarchical Multiple Regression Analysis was used to analyze the absenteeism. The absence data used in this system is the data at Good Brother Company (Mandalay), Nay Pyi Taw (Garden and Playground Parks Department). The prediction of effects on absenteeism system has programmedusing c# programming language.

Keywords:

Absenteeism, hierarchical multiple regression, hypothesis testing, correlation analysis

INTRODUCTION

Many work in real life, absenteeism is one of the persistent challenges for which there is no clear guilty and no easy cure. In addition, it does not discriminate against individuals on the basis of gender, race and religion as a general issue. Staff attendance is based on a staff's motivation to attend as well as their ability to attend. The term "absenteeism" originates from the Latin word, "absentia". Although there is not a standard definition of absenteeism, it is extensively used to describe non-attendance of employees for scheduled work [5].

Factors such as family duties, traffic problems, accidents, etc. affect a staff's ability to attend. Once all of these factors are recognized, executives may start to realize why employees sometimes choose not to come to job when they are completely interested in attending. Work-related commitments can increase performance, reduce absenteeism and benefit both the staff and the organization. Job satisfaction has been considered one of the main determinants of absenteeism among staff. Job satisfaction and organizational commitment are closely linked to the aggregated absence period [1].

Absenteeism is commonly understood as a staff or a group of staffs remaining absent from work either continuously for a long period or repeatedly for short periods [4]. Absence can basically be separated into an involuntary part and a voluntary part. Involuntary part, e.g. accredited disease, participation at funerals, public holidays, casual leave, maternity leave or unforeseen circumstances, is beyond the unexpected control of the staffVoluntary parts, e.g. unregistered sickness or leave without pay, are controlled by employees directly and are often based on personal goals. In this system, the number of workdays which are taken off is computed by absenteeism. It is described as unpaid leave, unexpected job absence, irrespective of the reason, including long-term and short-term disability.

Hierarchical multiple regression is a type of multiple regression. Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables [7]. The variable which wants to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable) and the variables which is used to predict is called independent variable [1].

iJETRM

International Journal of Engineering Technology Research & Management

In this system, independent variables are demographic characteristics, service, work attitudes and absenteeism is dependent variable.

For this system, hierarchical multiple regression is used to evaluate the relationship between a set of independent variables and the dependent variable, controlling for or taking into account the impact of a different set of independent variables on the dependent variable. In order to test the hypothesis suggested, three different regression models are constructed using hierarchical multiple regression analysis. In this system, age, education status, marital status, number of children, service, job satisfaction and organizational commitment data will be used to predict absenteeism. The absent data is used in this project is the Staffs' absenteeism data in Good Brother Company, Mahar Aung Myay Township, Mandalay, Myanmar which is the data in 2017. Hierarchical multiple regression is also the most popular and often used to forecast the causes of absence affected by a number of factors.

Effects on Absenteeism

Businesses are needed to improve productivity while their costs reduce. Absenteeism, low level of performance, severance and distrustful behaviors are the behaviors that influence company enterprises' productivity and effectiveness. Such behaviors brings high costs to companies and therefore are the issues that must be monitored and dealt with properly.

Absences can lead to many problems for the business. If a staff doesn't go to work, his/her work will be assigned to other colleagues that have already workload. Work of the absent staff overloads another staff and this additional workload definitely affects the morale of these employees negatively. This situation not only decreases efficiencies of employees, but also increases stress. Moreover, lack of communication in the workplace and extra cost for the organization [3].

Absenteeism is extensively used to describe non-attendance of staffs for scheduled work. The statistics reflect that staff absenteeism produces dramatic costs to national economies in terms of lost productivity. Absenteeism implies an unplanned, disruptive incident but more specifically, it can be seen as non-attendance when a staff is scheduled for work [2]. Absenteeism cannot be predicted easily and it is a serious problem which occurs at workplace and it is an expensive occurrence for the organization and the staffs. The company faces high level of absenteeism. So, by finding the factors causing absenteeism and finding the ways to overcome it the problem of absenteeism can be reduced.

Unscheduled absences affect almost every type of organization. Hoque and Islam (2003, p. 81) describe absenteeism as a "subject to be studied, matter to be thought over and a problem to be solved" [2]. To the manager, absence is often seen as a problem to be solved, but to the staff it can take on a very different meaning.

Impact of Demographic Characteristics on Absenteeism

The importance of demographic characteristics in explaining absenteeism was acknowledged with the appearance of Price's (1995) model. According to Price (1995) demographic characteristics direct affect absenteeism. Price (1995) further postulates that demographic variables can assist in the construction of causal models and assist in the management of organizations [2]. Information of this nature can, for example, assist with recruitment and selection decisions in organizations.

Impact of Work Attitudes on Absenteeism

Work attitudes are considered important determinants of absenteeism behavior. Questionnaires used in this study is designed to examine the relationships between job satisfaction, organizational commitment and absenteeism. While job satisfaction is considered the main determinant of absenteeism, organizational commitment forms are considered the main predictor of absenteeism intentions [6]. Sagie (1998) found in that regard that organizational commitment and job satisfaction were strongly related to the aggregated duration of voluntary absence [1].

METHODOLOGY

Hierarchical Multiple Regression Analysis is used to understand the factors that effects absenteeism behavior and to find which factors are most influencing absenteeism intention. The proposed methods of hierarchical multiple regression analysis are provided.

In terms of data analysis, a data transform was initially performed to provide information pertaining to the demographics of the respondents. Then, hierarchical multiple regression analyses were performed to test the hypothesis suggested between the whole set of independent variables and the dependent variables under the current study. Each of the regressions in this study was conducted using the following models. In the first model,

IJETRM

International Journal of Engineering Technology Research & Management

the four demographic variables (age, educations status, marital status, number of children) were entered into equation. In the second model, demographic variables and service were entered. In the third model, demographic variables, service and work attitudes were entered. Models were tested by comparing a model with no predictors to the model that specify. Subsequently, unstandardized residuals from the regression analyses were calculated. Later, hypothesis testing was conducted to determine which hypothesis proposed based on this study are strongly supported. Lastly, implications of the results from the analysis are described with complete explanations and step by step implementations. Figure 1 shows the system flow diagram of this system.

Hierarchical Regression Model

Hierarchical regression model is used to state the hypothesized relationships between absenteeism [8]. To build a hierarchical model, the independent variables are added one at a time in a theoretical manner. The variable that is believed to be the biggest predictor or most likely to explain the variation in the dependent variable is added first, followed by the second, then third. Each time we add a variable, we examine the model output and interpret the critical-value of the variable and the R-squared value of the model.
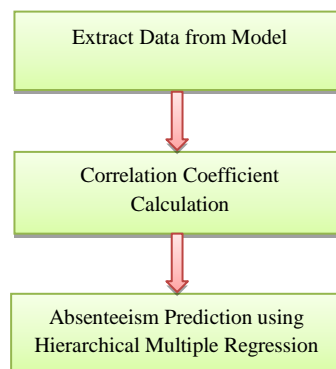
Fig. 1: Absenteeism Prediction Framework

Regression Coefficient

Model 1 (Demographic Characteristics), Model 2 (Demographic Characteristics and Service) and Model 3 (Demographic Characteristics, Service and Organizational Commitment) have been selected to demonstrate hypotheses that have a strong impact on absenteeism.

Model and Hypothesis of the Study

The purpose of this system is to examine the impact of demographic characteristics, service and work attitudes on staffs' absenteeism data. The data are collected from Good Brother Company (Mandalay) and Nay PyiTaw (Garden and Playground Parks Department). Daily absences of staffs have been also collected for 12 months. The data about the demographic characteristics and service of staffs were taken from the human resources department. Education status is defined as high school =1, bachelor = 2, honors = 3 and master = 4. Marital status is defined as a dichotomous variable (married = 1, single = 0). Work attitude was administered to staffs working in a wide variety of jobs in different work organizations with questionnaires that consists of 20 questions. Three-level like scale: strongly agree (3), neutral (2) and disagree (1).

Suggested hypotheses are tested by hierarchical multiple regression analysis. To build a hierarchical multiple regression analysis, one independent variable is added at a time in a theoretical manner. The variable that is believed to be the more than one independent variable is added first, followed by the second, then third. State the hypothesis:

- Null hypothesis- Assume there is no relationship between independent variable and dependent variable.
- Alternative hypothesis- Assume there is relationship between independent variable and dependent variable.

To analyze the relationship between demographic characteristics, service, work attitudes and absenteeism, each of the regressions is conducted using the following models. In the first model, the four demographic characteristics (age, educations status, marital status and number of children) are entered.

iJETRM
International Journal of Engineering Technology Research & Management

1) H_1: Age will have an inverse relationship with absenteeism. Older staff will have less absenteeism.
2) H_2: Staff with high education level will have less absenteeism.
3) H_3: Married staffs will have higher absenteeism.
4) H_4: Staffs with more children will have higher absenteeism.
   In the second model, demographic characteristics and service are entered.
5) H_5: Staff with high service will have higher absenteeism.
   In the third model, demographic characteristics, service and work attitudes are entered.
6) H_6: Job satisfaction will be related to absenteeism. Lower job satisfaction will lead to higher absenteeism.
7) H_7: Organizational Commitment will be related to absenteeism. Higher levels of organizational commitment will reduce absenteeism.

Hypothesis Testing

A hypothesis test is rule that specifies whether to accept or reject a claim about a population depending on the evidence provided by a sample of data. A hypothesis test examines two opposing hypotheses about a population: the null hypothesis and the alternative hypothesis. Critical value is used to make the determination. If test-statistic is larger than the critical value is based on distribution table, then we can accept the alternative hypothesis.

- F-test is used to test the overall significance of the model.
- T-test is used to test the significance of individual independent variable.

Hierarchical Multiple Regression Analysis

Multiple regression is an extension of sample linear regression. It is used when we want to predict a single dependent variable(Y) based on one or more other independent variables(X). The prediction of Y is expressed by the following equation:

$$Y' = b_0 + b_1 x_1 + b_2 x_2 + \; ... + \; b_k x_k \tag{1}$$

In above equation, Y' represents a predicted value of Y, b_0 indicates the intercept of Y', b_1, b_2, ..., b_k represents coefficients of independent variables, x_1, x_2, ..., x_k represents the independent variables and k indicates the number of independent variables.

The estimates for the coefficients are computed so as to minimize the sum of squares of differences between the observed dependent variable (values of the variable being predicted) in the data. The coefficient of independent variables are estimated using ordinary least squares estimator (OLS). Regression coefficients represent the mean change in the dependent variable for one unit of change in the independent variable while holding the other independent variables in the model constant.

Table 1: Correlation Coefficient

	Model 1	Model 2	Model 3 (Job Satisfaction)	Model 3 (Organizational Commitment)
Constant (b_0)	**4.081**	**2.925**	**4.036**	**6.737**
Age (b_1)	**-0.010**	**0.042**	**0.044**	**0.045**
Education Status (b_2)	**0.565**	**0.599**	**0.649**	**0.710**
Marital Status (b_3)	2.086	2.061	2.109	2.097
Number of Children (b_4)	-0.190	-0.107	-0.131	-0.096
Service (b_5)		-0.083	-0.085	-0.102
Job Satisfaction (b_6)			-0.060	
Organizational Commitment (b_7)				-0.188

iJETRM

International Journal of Engineering Technology Research & Management

F-Test

The F-test are used to assess whether the set of independent variables collectively predicts the dependent variable.

$$F = \frac{MS_{Regression}}{MS_{Residual}} \tag{2}$$

$$MS_{Regression} = \frac{SS_{Regression}}{k} \tag{3}$$

The sum of squared regression measures how much of the variation in the dependent variable by the model explained.

$$SS_{Regression} = \sum_{i=1}^{n}(Y' - \overline{Y})^2 \tag{4}$$

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y}{n} \tag{5}$$

where, Y' is a $predicted$ $value$ of Y, $\overline{Y}$ is the mean of Y, n is the sample size and Y is dependent variable.Critical value of F-Test is defined by as follow:

$$K_f = t_{\alpha ,k, \ n-k-1} \tag{6}$$

where, α is the significance level, k is the number of independent variables and n is the sample size. According to decision rule, $F > K_f$ is the accept model and otherwise is reject model. In hypothesis testing, the significance level (also called α) is the criterion used for rejecting the null hypothesis. In general, this is an arbitrary decision, though, as we have said, the 1% , 5% and 10%levels for α are the most used in practice. The comparison of F-test and critical value with three **α** values is displayed in Table 2.

*Table 2: Comparison of F-Test and Critical Value with **α** Value*

Model	F-Test	Critical value			Comparison			Results			Confidence Level		
		0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1	0.01
Model 1	2.873	2.435	1.985	3.453	2.873 > 2.435	2.873 > 1.985	2.873 < 3.453	Accepted	Accepted	Not Accepted	95 %	90 %	99 %
Model 2	2.386	2.277	1.888	3.149	2.386 > 2.277	2.386 > 1.888	2.386 < 3.149	Accepted	Accepted	Not Accepted	95 %	90 %	99 %
Model 3 (Job Satisfaction)	2.044	2.163	1.816	2.932	2.044 < 2.163	2.044 > 1.816	2.044 < 2.932	Not Accepted	Accepted	Not Accepted	95 %	90 %	99 %
Model 3 (Organizational Commitment)	2.911	2.163	1.816	2.932	2.911 > 2.163	2.911> 1.816	2.911< 2.932	Accepted	Accepted	Not Accepted	95 %	90 %	99 %

T-Test

The T-test is used to determine whether each of the individual independent variables is significant.

$$t = \frac{b}{S_{b_i}} \tag{7}$$

iJETRM

International Journal of Engineering Technology Research & Management

In above equation, b is the coefficient of independent variable and S_{b_i} is standard error of each coefficient. Standard error of coefficient (S_{b_i}) is as follow:

$$S_{b_i} = \sqrt{\frac{1 - R^2_{YH}}{(1 - R^2_{X_k G_k}) \times (n - k - 1)}} \times \frac{S_Y}{S_{X_k}} \tag{8}$$

where, R^2 is the coefficient of determination, Y is dependent variable, H is the set of all the X (independent) variables, G_k is the set of all the X variables, except X_k. $R^2_{YH} = R^2$ obtained by regression Y on x_1, x_2 and so on. Critical value of T-test is

$$K_t = t_{\frac{\alpha}{2}, \, n-k-1} \tag{9}$$

where, α is the significance level, k is the number of independent variables, n is the sample size. According to decision rule, $|t| \geq K_t$, is the accept model and otherwise is reject model.

EXPERIMENTAL RESULTS

In model 1, age, education status, marital status and number of children were analyzed as demographic characteristics. Regression estimates of Model 1 shows that the demographic characteristics were strongly associated with absenteeism. Table 3 shows the hypothesis results of model 1 with two α values.

Table 3: Hypothesis Results of Model 1 with α Value

Variable	T-Test	Critical Value		Comparison		Results		Confidence Level	
		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
Age	0.163	1.976	1.656	$|t| < 1.976$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %
Education Status	1.247	1.976	1.656	$|t| < 1.976$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %
Marital Status	**2.972**	**1.976**	1.656	$\mathbf{|t| > 1.976}$	$|t| > 1.656$	**Accepted**	Accepted	**95%**	90 %
Number of Children	0.397	1.976	1.656	$|t| < 1.976$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %

When the results are examined in Model 2, it is concluded that demographic characteristics, service and absenteeism have a significantly linked with absenteeism. The hypothesis results of model 1 with two α values are shown in Table 4.

The results of model 3 shows significant relationship between demographic characteristics, service, organizational commitment and absenteeism with (T= 3.021) and (>critical value = 1.977). Based on this result, we can accept model 3. Table shows the hypothesis results of model 3 with two α values.
Table 6 shows the summary of hypothesis results at Good Brother Company dataset. This dataset shows that marital status and organizational commitment (OC) was strongly related on absenteeism with significant level α = 0.05 and α = 0.1.

Table 4. Hypothesis Results of Model 2 with α Value

Variable	T-Test	Critical Value		Comparison		Results		Confidence Level	
		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
Age	0.429	1.977	1.656	$|t| < 1.977$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %
Education Status	1.310	1.977	1.656	$|t| < 1.977$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %
Marital Status	**2.928**	**1.977**	1.656	$\mathbf{|t| > 1.977}$	$|t| > 1.656$	**Accepted**	Accepted	**95%**	90 %
Number of Children	0.215	1.977	1.656	$|t| < 1.977$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %
Service	0.691	1.977	1.656	$|t| < 1.977$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %

iJETRM

International Journal of Engineering Technology Research & Management

Table 5. Hypothesis Results of Model 3 (OC) with α Value

Variable	T-Test	Critical Value		Comparison		Results		Confidence Level	
		0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
Age	0.472	1.977	1.656	$|t| < 1.977$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %
Education Status	1.567	1.977	1.656	$|t| < 1.977$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %
Marital Status	**3.021**	**1.977**	1.656	$\mathbf{|t| > 1.977}$	$|t| > 1.656$	**Accepted**	Accepted	**95%**	90 %
Number of Children	0.197	1.977	1.656	$|t| < 1.977$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %
Service	0.859	1.977	1.656	$|t| < 1.977$	$|t| < 1.656$	Not Accepted	Not Accepted	95%	90 %
Organizational Commitment	**2.278**	**1.977**	1.656	$\mathbf{|t| > 1.977}$	$|t| < 1.656$	**Accepted**	Not Accepted	**95%**	90 %

Table 6: Summary of Hypothesis Results (α = 0.05 and α = 0.1) at GBC

Hypotheses	Results
H_3 : Married staffs will have higher absenteeism. H_7: Organizational Commitment will be related to absenteeism. Higher levels of organizational commitment will reduce absenteeism.	Accepted Accepted

CONCLUSION

The purpose of this framework is to comprehend the benefit of hierarchical multiple regression assessment on the absenteeism process. According to the results of the system, there is a relationship between dependent variable of absenteeism and independent variables. The results show marital status and organizational commitment strongly significant on absenteeism. The organization has to make the married staff go with an authorized leave or else it would bring down the absenteeism by offering counseling services to staff by making them aware of these leaves. Organization will require various level of analysis to identify the factors that impact absenteeism for a specific employer.

The result of hierarchical multiple regression analysis is based on the selected independent variables that is derived from a hypothesis or theory. In order to test the hypothesis suggested, this system can be extended by other methods such as multiple regression, simultaneous multiple regression and stepwise multiple regression.

REFERENCES

[1] Cohen, A. and Golan, R., 2007. Predicting absenteeism and turnover intentions by past absenteeism and work attitudes: An empirical examination of female employees in long term nursing care facilities. Career Development International, 12(5), pp.416-432.

[2] Josias, B.A., 2005. The relationship between job satisfaction and absenteeism in a selected field services section within an electricity utility in the Western Cape (Doctoral dissertation, University of the Western Cape).

[3] Senel, B. and Senel, M., 2012. The cost of absenteeism and the effect of demographic characteristics and tenure on absenteeism. Interdisciplinary Journal of Contemporary Research in Business, 4(1), pp.1142-1151.

[4] Murthy, S.A.S., (1954): A study on absenteeism, Abhinav Journal of Social Work, 14 (1-4), P.132.

[5] Akgeyik, T., 2014. Factors affecting employee absenteeism (A study on a sample of textile workers). European Journal of Management, 14(3), pp.69-76.

[6] https://bspace.buid.ac.ae/bitstream/1234/193/1/90129.pdf

iJE**T**RM

International Journal of Engineering Technology Research & Management

[7] Petrocelli, J.V., 2003. Hierarchical multiple regression in counseling research: Common problems and possible remedies. Measurement and evaluation in counseling and development, 36(1), pp.9-22.

[8] [8] Hoyt, W.T., Leierer, S. and Millington, M.J., 2006. Analysis and interpretation of findings using multiple regression techniques. Rehabilitation Counseling Bulletin, 49(4), pp.223-233.