EDUCATUM JSMT Vol. 12 Number 2 (2025)
ISSN 2289-7070 / e-ISSN 2462-2451(86-94)
https://ejournal.upsi.edu.my/index.php/EJSMT/index

Analysis of Workers' Compensation Using Multiple Linear Regression Model: A Case Study in an Estates and Mills Company

Nor Hasliza Mat Desa[1], Mohd Saiful Adli Mohamad[1], and See Chun Kit[1]

[1]School of Quantitative Sciences, 06010 UUM Sintok, Kedah, Malaysia

**Corresponding author:* nliza@uum.edu.my

Received: 14 September 2024; **Accepted:** 13 August 2025; **Published:** 16 December 2025

To cite this article (APA): Nor Hasliza MD, Analysis of Workers' Compensation Using Multiple Linear Regression Model: A Case Study in an Estates and Mills Company, EDUCATUM Journal of Science, Mathematics and Technology, 12(2), 86–94.

To link to this article:

Abstract

Fair and open compensation is a critical component of employees' happiness and satisfaction. To provide competitive and equitable pay, an analysis of current compensation is necessary. Therefore, the main goal of this paper is to investigate the relationship between possible factors and compensation among 368 workers who work in the mills and estates sector in Malaysia. This paper also aims to identify the differences in compensation between different job types and races. The result of the multiple linear regression model showed that only five factors (monthly working days, race, gender, worker type, and job type) were significantly associated with worker's compensation. This paper also revealed that the monthly compensation for the four job types held by workers, which consist of general worker, harvester, mill worker, and rubber tapper in estates and mills, differed through the Kruskal-Wallis test. To conclude, the findings of this paper could support the strategic goals of the business, benchmark compensation, enhance compensation decision-making, and provide a more equitable compensation structure for workers.

Keywords: compensation, Kruskal-Wallis Test, Multiple Linear Regression, Spearman Correlation Analysis, sustainability

INTRODUCTION

Employers are continuously working to improve employee happiness and operational efficiency in the cutthroat business climate of today in order to attract and retain top personnel. Fair and open compensation is a critical component of employee happiness and satisfaction. Consequently, in order to guarantee that workers are paid appropriately, companies must perform thorough wage analyses taking into consideration a variety of factors, including years of experience, job function, talents, and industry norms [1].

It is crucial for the success of any organization to provide their employees with competitive and equitable pay. However, inequalities in pay packages may result in low employee satisfaction, significant staff turnover, and trouble luring top talent. Therefore, an analysis of worker's compensation in is needed in order to assess the organization's compensation competitiveness in the industry.

An analysis of compensation is important since an organization can make sure it is operating and treating its people responsibly financially by doing a compensation analysis. There are a few reasons of conducting compensation analyses such as competitive labor markets, responsible operation of an organization, avoiding bias, and performance motivation. For the competition in the job market, organizations need to hire competent workers who foster growth to prosper. By providing a competitive compensation or rate, company shows that it appreciates its employees just as much as other companies [2]. Apart from that, the organizations must avoid bias. Formal pay structures help business management treat employees with integrity and respect by preventing biases based on gender, race, or personal relationships from influencing pay [3].

Multiple linear regression (MLR) method is often preferred in compensation analysis because it allows for the simultaneous evaluation of the effects of multiple independent variables such as education, experience, job level, and performance on employee compensation. MLR method also effectively controls for confounding variables, making it ideal for identifying potential disparities and ensuring fair compensation practices across demographic groups. Moreover, recent studies highlight its interpretability and effectiveness in compliance audits and HR decision-making compared to more complex models like machine learning which may lack transparency [4, 5].

Therefore, the main goal of this paper is to investigate the relationship between the independent variables (age, years of experience, monthly working days, race, gender, worker type and job type) and compensation among workers who work in a mills and estates sector in Malaysia (dependent variable). Besides that, this paper also aims to identify the differences in compensation between different job types such as general worker, harvester, mill worker, and rubber tapper among races.

MATERIALS AND METHODS

A secondary data was obtained from Worker Wages Detail Report in a mills and estates company. In this paper, compensation of the workers (RM per month) was used as dependent variable, meanwhile seven factors were used as independent variables which consists of age, years of experience, monthly working days, race, gender, worker type and job type.

Determination of Sample Size

For most quantitative research, 40 sample size is a good quantity, although, in some circumstances, researchers can enrol fewer users [6]. A formula for the minimum sample size was proposed by [7] as follows:

$$r_s = n = \frac{\left(X^2NP(1-P)\right)}{\left(d^2(N-1)+X^2P(1-P)\right)} \tag{1}$$

with

n = sample size

X^2 = the table value of chi-square for one degree of freedom at the desired confidence level, $X^2 = 3.841$

N =population size (8483 workers)

P =The population proportion (assumed to be 0.5 since this would provide the maximum sample size)

d =ddegree of accuracy expressed as a proportion (0.05)

Then, the sample size is calculated as follows:

$$n = \frac{\left(3.841 \times 8483 \times 0.5(1-0.5)\right)}{\left(0.05^2(8483-1)+3.841 \times 0.5(1-0.5)\right)}$$

$$n = \frac{8145.8008}{22.16525} \approx 368\, workers$$

Therefore, a total of 368 sample size were selected in this compensation analysis.

Spearman's Rank Correlation Analysis

The Spearman Rank Correlation is a nonparametric measure to examine the relationship between the selected possible factors [8]. It has two assumptions as follows:

i. The data comprises a randomly selected sample of n pairs of either numeric or non-numeric observations.

ii. Each pair of observations signifies two measurements obtained from the identical object or individual, referred to as the unit of association.

The general formula for Spearman's Rank Correlation Coefficient, r_s as follows:

$$r_s = 1 - \frac{(6\sum {d_i}^2)}{(n(n^2-1))} \quad (2)$$

where,
d_i= difference between two ranked data set $i = 1,2,3,\cdots,n$
n = number of samples

The Spearman's correlation coefficient r_s varies between -1 and 1. The closer r_s is to ±1 the stronger the monotonic relationship. To test the significance of correlation coefficients, the null hypothesis and the alternative hypothesis are given as follows:

H_0: The correlation coefficient $r_s = 0$ (There is no correlation).

H_1: The correlation coefficient $r_s \neq 0$ (There is a correlation).

If the *p-value* is less than the specified significance level (usually 5%), then the null hypothesis is rejected. Otherwise, the null hypothesis is not rejected.

Kruskal-Wallis Test

The Kruskal-Wallis test is a nonparametric statistical method utilizing ranks to assess whether there exist significant differences among two or more groups of an independent variable with respect to a continuous or ordinal dependent variable. It serves as a nonparametric substitute for the one-way ANOVA and extends the capability of the Mann-Whitney U test, enabling the comparison of more than two independent groups [9]. Since the assumption in homogeneity of variance is violated in the ANOVA test, therefore the Kruskal-Wallis Test is used since it is an alternative to ANOVA test. The assumptions in the Kruskal-Wallis test as follows:

i. Ordinal Variables - The variable under consideration needs to be either ordinal or continuous, indicating some form of hierarchy.

ii. Independence - Each group should be distinct and not influenced by the others.

The general formula for statistic H in Kruskal-Wallis Test as follows:

$$H = \frac{12}{N(N+1)}\sum\frac{{R_i}^2}{n_i} - 3(N+1) \quad (3)$$

where
N = total number of observations in all groups.
n_i = number of observations in the i-th group.

R_i = total sum of ranks in the i-th group.

The null hypothesis and the alternative hypothesis are given as follows:

H_0: The independent samples all have the same central tendency and therefore come from the same population.

H_1: At least one of the independent samples does not have the same central tendency as the other samples and therefore originates from a different population

If the *p-value* is less than the specified significance level (usually 5%), then the null hypothesis is rejected. Otherwise, it is not rejected.

Multiple Linear Regression Analysis

Multiple linear regression (MLR) is a statistical method that employs multiple explanatory variables to forecast the outcome of a response variable. The primary goal of multiple linear regression modeling is to establish a linear connection between the response (dependent) variables and the explanatory (independent) variables [10].

Multiple linear regression model is used to predict the compensation of workers in estates and mills company and determine which independent variables have significant relationship with the compensation of workers. The multiple regression model relies on the following assumptions including linearity, independence of errors, normality of error, and equal variance. The general formula for the Multiple Linear Regression model can be written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \varepsilon_i \quad (4)$$

where

y = Monthly compensation of the workers (RM)
x_1 = Age (0 = <25 years, 1= 25-34 years, 2= 35-44 years, 3= 45-54 years, 4= >54 years)
x_2 = Years of Experience (0 = <10 years, 1= 10-19 years, 2= 20-29 years, 3= 30-39years, 4=>39 years)
x_3 = Monthly Working Days (0 = <10 days, 1= 10-19 days, 2= 20-29 days, 3=>29 days)
x_4 = Race (0 = Bangladeshi, 1= Indian, 2= Indonesian, 3= Malay, 4= Other Indigenous, 5= Others)
x_5 = Race (0 = Male, 1= Female)
x_6 = Worker Type (0 = Local, 1= Foreign, 2= Permanent Resident)
x_7 = Job Type (0 = General Worker, 1= Harvester, 2= Mill Worker, 3= Rubber Tapper)
β_0 = Intercept Coefficient
$\beta_1, \beta_2, \ldots, \beta_7$ = Regression Coefficient for Independent Variable of Age, Years of Experience, Monthly Working Days, Race, Gender, Worker Type and Job Type.

To test the overall fit of regression model, the null hypothesis and the alternative hypothesis are given as follows:

$H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0$ (There is no relationship between the dependent variable and the set of independent variables.)

H_1: At least one $\beta_i \neq 0$ (There is a relationship between the dependent variable and the set of independent variables.)

The hypothesized model about the individual β_k for this test is given as follows:

$H_0: \beta_k = 0$

$H_1: \beta_k \neq 0$

If the *p-value* is less than the specified significance level (usually 5%), then the null hypothesis is rejected. Otherwise, it is not rejected

RESULTS AND DISCUSSION

Demographic of Sample Data

Table 1. Summary of Descriptive Analysis

Category	Item	Frequency (n)	Percentage (%)
Age	<25	28	7.6
	25-34	100	27.2
	35-44	130	35.3
	45-54	72	19.6
	>55	38	10.3
Years of Experience	<10	247	67.1
	10-19	59	16.0
	20-29	39	10.6
	30-39	22	6.0
	>39	1	0.3
Monthly Working Days	<10	6	1.6
	10-19	36	9.8
	20-29	325	88.3
	>29	1	0.3
Race	Bangladeshi	60	16.3
	Indian	135	36.7
	Indonesian	87	23.6
	Malay	81	22.0
	Other Indigenous	4	1.1
	Others	1	0.3
Gender	Male	320	87.0
	Female	48	13.0
Worker Type	Local	170	46.2
	Foreign	197	53.5
	Permanent Resident	1	0.3
Job Type	General Worker	92	25.0
	Harvester	92	25.0
	Mill Worker	92	25.0
	Rubber Tapper	92	25.0

Table 1 shows the gender distribution of 384 workers in an estates and mills company. It indicates that 320 (87%) male workers and 48 (13%) female workers are employed in the company. The majority of workers are Indian, accounting for 36.68%, followed by Indonesians (23.64%), Malays (22.01%), and Bangladeshis

(16.30%). The remaining categories, including other indigenous and others, constituted 1.09% and 0.27%, respectively. Meanwhile, 197 (53.5%) workers are foreign workers, followed by 170 (46.2%) local workers, and 1 (0.3%) permanent resident. Four job types are identified: general worker, harvester, mill worker, and rubber tapper, with 92 workers (25%) for each category.

Table 2. Correlation Between the Demographic and Possible Factors with Workers' Compensation

Variable		Age	Years of Experience	Monthly Working Days	Race	Gender	Worker Type	Job Type
Compensation	***r***	-0.140	-0.186	0.379	0.060	-0.411	0.425	-.0312
	Sig.	0.007	0.000	0.000	0.254	0.000	0.000	0.000

Table 2 illustrates the correlation between demographic and other potential factors related to workers' compensation, indicating that only the race component is statistically insignificant ($p>0.05$). At a 5% significant level ($p<0.05$), there is a weak negative relationship between workers' compensation and factors like age, years of experience, and job type. There is also a moderate negative relationship with the gender factor. Meanwhile, worker type and monthly working days' factors show a moderate positive relationship on worker's compensation.

Table 3. The Analysis of Kruskal-Wallis Test

Total N	**368**
Test Statistics	165.000
Degree of Freedom	3
Sig.	0.000

From Table 3, the analysis of the Kruskal-Wallis Test indicated the four job types of workers (general worker, harvester, mill worker, and rubber tapper) in estates and mills do not have the same median in monthly compensation. Therefore, the null hypothesis is rejected.

Table 4. The Pairwise Comparisons of Job Type for Workers

Job Type Comparison	Test Statistic	Std. Error	Std. Test Statistic	P-value
Rubber Tapper-General Worker	60.533	15.684	3.859	0.000*
Rubber Tapper-Mill Worker	81.413	15.684	5.191	0.000*
Rubber Tapper-Harvester	196.620	15.684	12.536	0.000*
General Worker-Mill Worker	-20.880	15.684	-1.331	0.183
General Worker-Harvester	-136.087	15.684	-8.677	0.000*
Mill Worker-Harvester	115.207	15.684	7.345	0.000*

**p-value is significant at 0.05 significance level.*

Table 4 shows the pairwise comparison of job types of workers. It can be seen that all the job type of workers in estates and mills are difference in median of monthly compensation expect for the job type comparison in general worker and mill worker is 0.183, which is greater than 5% of significance level, means that this job type comparison is statistically insignificant and have the same median in monthly compensation while others job type comparison is statistically significant and do not have the same median in monthly compensation. However, in this pairwise comparison, a Type I error may occur when a statistically significant difference is found between two groups median when no true difference exists. In this case, post-hoc adjustment methods such as Bonferroni test could be considered.

There are five differences in median of monthly compensation between the job type of workers in estates and mills. Since the *p-value* in the job type comparison in general worker and mill worker is 0.183 ($p>0.05$), t this job type comparison is statistically insignificant and have the same median in monthly compensation while others job type comparison is statistically significant and do not have the same median in monthly compensation.

Table 5. The Mean Rank of Compensation for Each Job Type

	Job Type	N	Mean Rank
Compensation	General Worker	92	160.39
	Harvester	92	296.48
	Mill Worker	92	181.27
	Rubber Tapper	92	99.86
	Total	368	

From Table 5, the mean rank of the compensation for each job type is shown. The job type which has the highest and lowest compensation can be identified. The job type of Harvester has the highest mean rank followed by mill worker and general worker while the rubber tapper has the lowest mean rank. It means that although the median of compensation does not much difference for the worker except for the general worker and mill worker, but there is still a difference in the mean rank for the workers' compensation according to the job type.

.

Table 6. Overall Regression Model Summary

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	46.248	7	6.607	38.7	0.000
Residual	60.948	357	0.171		
Total	107.196	364			

Table 7. Parameter Estimation between Compensation and Selected Factors

Variable	Coefficients, B	Std. Error	t Stats	Sig.	Lower 95%	Upper 95%
Intercept	6.101	0.189	32.288	0.000*	5.729	6.473
Age	0.003	0.003	1.080	0.281	-0.003	0.009
Years of Experience	0.005	0.003	1.539	0.125	-0.001	0.012
Monthly Working Days	0.046	0.007	6.867	0.000*	0.032	0.059
Race	0.070	0.023	3.013	0.003*	0.024	0.116
Gender	-0.369	0.074	-5.009	0.000*	-0.514	-0.224
Worker Type	0.374	0.061	6.108	0.000*	0.254	0.495
Job Type	-0.139	0.020	-6.907	0.000*	-0.178	-0.099

Table 6 shows that the overall regression model is significant ($p < 0.05$). Hence, it can be concluded that there is at least one coefficient that is not equal to zero in this regression model. Other than that, the result in Table 7 reveals that five independent variables such as total working days, race, gender, worker type, and job type have a significant relationship with compensation ($p < 0.05$). The workers' gender component has a considerable impact on the worker's compensation. It is apparent from the response demographic that structural gender inequality exists. Due to the fact that this company's tasks are more related to men than to women, job segregation is the result of the worker type factor. Meanwhile, the age and years of experience have not significant with compensation in this model ($p > 0.05$). This result is consistent with Table 2, which show that there is a weak negative relationship between compensation and these two factors. Therefore, the MLR model for this compensation analysis is given as follows:

$$y = 6.101 + 0.046(\text{Monthly Working Days}) + 0.07(\text{Race}) - 0.369(\text{Gender}) + 0.374(\text{Worker Type}) - 0.139(\text{Job Type})$$

CONCLUSION

In this paper, it can be concluded that only race factor does not show a significant correlation with workers' compensation while other factors such as age, gender, worker type, job type, years of experience, and monthly working days of worker had a significant correlation with workers' compensation through the Spearman's Rank Correlation Analysis. Besides that, it is found that the job types of workers (general worker, harvester, mill worker, and rubber tapper) in estates and mills sector do not have the same median in monthly compensation through Kruskal-Wallis Test. In overall, there are only five factors that are statistically significant relationship with compensation of workers in estates and mills sector in the Multiple Linear Regression model which are race, gender, worker type, job type, and total working days. In conclusion, this paper not only addresses compensation management but also supports organizational values and strategic goals. By identifying and addressing gender wage disparities, it reinforces the company's commitment to diversity, equity, and inclusion. The findings promote a fair and transparent compensation structure, positively impacting employee morale and satisfaction. Ultimately, this paper underscores the significance of data-driven and equitable compensation practices, laying a foundation for future studies and fostering a culture of trust and fairness within the organization. In conclusion, despite providing useful information, this paper is limited by the data from a single organization.

DECLARATION OF INTEREST

There is no conflict of interest with this study.

ACKNOWLEDGEMENT

This research was supported by the Fundamental University Research Grant (GPUF2022) [2022-0130-102-01]. The authors also would like to thank the Department of Chemistry, Faculty of Science and Mathematics, UPSI for research facilities.

REFERENCES

[1] Yuan, Y., Mabberley, D.J., Steane, D.A., Olmstead, R.G. (2010). Further disintegration and redefinition of *Clerodendrum* (Lamiaceae): Implications for the understanding of the evolution of an intriguing breeding strategy. *Taxon*, 59(1), 125-133.

[2] Ebadollahi, A., Ziaee, M., Palla, F. (2020). Essential oils extracted from different species of the Lamiaceae plant family as prospective bioagents against several detrimental pests. *Molecules*, 25(7), 1556-1571.

[3] Pei, J., Chen, S.L. (1982). *Flora of China*. Science Press: Beijing, China.

[4] Li, C.Z., Su, Y.F., Jin, X.J. (2005). Advances in studies on chemical constituents from plants of *VitexL*. and their bioactivities. *Chin,ese Traditional and Herbal Drugs*, 36, 930-938.

[5] Burkill, I.H. (1966). *A Dictionary of the Economic Product of Malay Peninsula*. Kuala Lumpur, Ministry of Agriculture and Co-operatives, 1966.

[6] Ladda, P.L., Magdum, C.S. (2012). *Vitex negundo* Linn.: ethnobotany, phytochemistry and pharmacology- a review. *International Journal of Advances in Pharmacy, Biology and Chemistry*, 1(1), 111-120.

[7] Edward, F.G. (1999). *Vitex trifolia* 'Variegata' - Variegated *Vitex*. *Fact Sheet*, 1, 1-3.

[8] Villella, S. (2016). The clinical management of menstrual migraine and headache by the herbal medicine practitioner. *Australian Journal of Herbal Medicine*, 28(3), 1-5.
[9] Bramley, G.L.C. (2019). *Flora Malesiana*. Djakarta.
[10] Salleh, W.M.N.H.W., Ahmad, F., Khong, H.Y. (2014). Chemical composition of *Piper stylosum* Miq. and *Piper ribesioides* Wall. essential oils, and their antioxidant, antimicrobial and tyrosinase inhibition activities. *Boletín Latinoamericano y del Caribe de Plantas Medicinales y Aromáticas*, 13, 488-497.
[11] Adams, R.P. (2007). Identification of Essential Oil by Gas Chromatography/quadrupole Mass Spectroscopy. Allured Publishing Corporation: Carol Stream, IL, USA.
[12] Johansen, D.A. (1940). *Plant microtechnique* .New York, McGraw-Hill Book.
[13] Almatar, M., Rahmat, Z., Salleh, F.M. (2013). Preliminary morphological and anatomical study of *Orthosiphon stamineus*. *Indian Journal of Pharmaceutical and Biological Research*, 1(4), 1-6.
[14] Pires, V.A., Cardozo-Junior, E.L., Ortmann, C. F., Maraschin, J.C., Favreto, W., Donaduzzi, C.M., Reginatto, F.H., Assreuy, J. (2018). Lipid-lowering and antiatherogenic effects of *Vitex megapotamica* (Spreng.) Moldenke in a mice experimental model. *Journal of Ethnopharmacology*, 215, 14-20.
[15] Santos, N., Mariane, B., Lago, J., Sartorelli, P., Rosa, W., Soares, M.G., da Silva, A.M., Lorenzi, H., Vallim, M.A., Pascon, R.C. (2015). Assessing the chemical composition and antimicrobial activity of essential oils from Brazilian plants-*Eremanthus erythropappus* (Asteraceae), *Plectrantuns barbatus* and *P. amboinicus* (Lamiaceae). *Molecules*, 20(5), 8440-8452.
[16] Bosabalidis, A.M., Sawidis, T. (2014). Glandular and non-glandular hairs in the seasonally dimorphic *Origanum dictamnus* L. (Lamiaceae) as a means of adaptation to cold stress. *Acta Agrobotanica*, 67(1), 15-20.
[17] Simpson, M.G. (2019). Plant morphology. *Plant Systematics*, 469-535.
[18] Glas, J., Schimmel, B., Alba, J., Escobar-Bravo, R., Schuurink, R., Kant, M. (2012). Plant glandular trichomes as targets for breeding or engineering of resistance to herbivores. *International Journal of Molecular Sciences*, 13(12), 17077-17103.
[19] Kowalski, R., Kowalska, G., Jankowska, M., Nawrocka, A., Kałwa, K., Pankiewicz, U., Włodarczyk-Stasiak, M. (2019). Secretory structures and essential oil composition of selected industrial species of Lamiaceae. *Acta Scientiarum Polonorum Hortorum Cultus*, 18(2), 53-69.