Modeling the Relationship Between Overtime Rate and Manufacturing KPIs: A Case Study in the OSAT Manufacturing Company using CRISP-DM

Jonathan D. Mondero
Associate Professor
College of Business Administration and Accountancy
Graduate Studies in Business
De La Salle University - Dasmariñas
Cavite, Philippines
jdmondero@dlsud.edu.ph

Jonathan Anthony C. Ac-ac , Shenalie M. Lopena

College of Computer Studies, Graduate School
De La Salle University
Manila, Philippines
jonathan_ac-ac@dlsu.edu.ph, shenalie_lopena@dlsu.edu.ph

Christopher Josh A. Cuevas
College of Engineering, University of Santo Tomas
Manila, Philippines
christopherjoshcuevas@gmail.com

Abstract

The study aims to understand how collected data from the Assembly Wire Bond manufacturing line and enterprise systems can lead to insights regarding the relationship of overtime rate to manufacturing KPIs: output attainment, utilization, and absenteeism. Using CRISP-DM as the data mining methodology, approximately 3.8 million records for attendance, 25.7 million for lot transactions, and 546.2 million for machine utilization were extracted, transformed, and summarized into weekly buckets, resulting in three hundred fifteen (315) samples from the five (5) selected packages, representing 97% of the total volume. The samples were subjected to Multiple Linear Regression and Pearson Correlation to determine whether there was a significant relationship between Overtime Rate and the manufacturing KPIs. Results show that Overtime Rate exhibits a statistically significant relationship with all the variables ($P < 0.05$) for the overall setup. Violations of the regression assumptions are probably due to model misspecification: other variables can more effectively explain the behavior of the overtime rate, as indicated by the low R-squared values. In contrast, only selected variables are considered significant across the results by package. These differences can be attributed to the different environmental conditions of semiconductor processes and manufacturing lines.

Keywords
Big Data Analytics, Overtime Rate, Absenteeism, Utilization, Semiconductor Manufacturing

1. Introduction
Globalization and highly competitive business environments are pushing many organizations to realize numerous productivity improvement efforts to continue to survive. With stiffening global competition, companies strive for ways to gain a competitive advantage by optimizing and improving their production (Huang et al. 2003). To stay competitive, manufacturing companies must have productive facilities (Fleischer et al. 2006). This would be possible if the production losses were identified and eliminated so that the manufacturers could bring their products to the market at a minimum cost (Muchiri and Pintelon 2008).

© IEOM Society International

In today's competitive world, price, quality, and time have critical roles in the success of companies achieving success in the competition (Rounaghi et al. 2021). Improving reputation and branding and increasing revenues by reducing costs are the primary strategic objectives of any entity (Dana et al. 2021) (Rounaghi 2019). The problem in manufacturing companies arises when the actual cost price of an individual product obtained by calculating a work order exceeds the planned price of that product. To solve the problem, it is necessary to analyze the direct and indirect incurred costs by comparing them with the planned cost by active cost management, which involves a constant and systematic seeking of opportunities to cost reduction and consistent implementation of cost reduction measures and activities (Bolfek 2021). Cost reduction is a continuous activity that must be a strategic priority. Strategic cost reduction must be part of a competitive strategy that integrates the strategies for managing technology and human resources and provides a long-term approach to cost reduction (Bolfek 2021). The long-term competitive cost-benefit depends on establishing a culture of continuous improvement in quality, time, and costs through innovation (Shields et al. 1992).

Nowadays, Outsourced semiconductor assembly and test (OSAT) companies face the challenges and issues of dynamic and unpredictable business environments, the strong influence of globalization, demands for rapid development of new complex products, requests for changes during the development process, short delivery times, accurate due dates, etc. On the other side, they are facing incomplete information, lack of knowledge, unpredictable resources, difficulties in supply, and various disturbances. All these complexities must be well mastered to fulfill orders, obtain customer satisfaction, reduce costs, and stay competitive in the market. Within this setting, effective management of shop floor operations and material flow and implementation of cost reduction mechanisms such as reducing overtime costs are some of the critical factors in controlling the operational complexity of such highly dynamic manufacturing environments, specifically in OSAT manufacturing.

1.1. Objectives

This research aims to develop a model for understanding the overtime rate of the OSAT Company. The model will be studied using the overtime rate data and manufacturing variables such as output attainment, utilization, and absenteeism. The objectives are as follows:

1. Determine the relationship between overtime rate, machine utilization, output attainment, and absenteeism.
2. Evaluate the influence of machine utilization, output attainment, and absenteeism on overtime rate.

2. Literature Review

Overtime work describes as work that may be performed beyond eight (8) hours a day provided that the employee is paid for the overtime work, an additional compensation equivalent to his or her regular wage plus at least twenty-five percent (25%) (Bureau of Labor Relations 2014). It is usually encountered when there is a labor shortage, unexpected demand, employee training, and extended seasonal hours (Bishop 2022). Co-workers and supervisors are required to work overtime to cover employee absences (Society of Human Resource Management 2014), and a corresponding overtime premium is given to employees (Department of Labor and Employment 2019). Ehrenberg (1970) said that attendance rate is random and that he supported it with a graphical illustration that shows overtime and absenteeism are positively related.

Absenteeism is an employee's failure to report for or remain at work, excluding vacations, holidays, jury duty, and the like (Cascio 2003). It is also defined as the absence of workers from regular work without prior permission (Tiwari 2014). Several studies in manufacturing were done to determine the influence of absenteeism (Chowdhury 2016) and relationships (Abston et al. 2019) between pay practices, benefit programs, location characteristics, and turnover. Poor community health outcomes (McHugh et al. 2019), age, and high-physical workload (Fritzsche et al. 2014) were strongly associated with longer absenteeism, increased error rate, and tardiness. The availability and productivity of manufacturing facilities directly impact the competitiveness of manufacturing companies (Fleischer et al. 2006).

To remain competitive in an increasingly global market, manufacturing firms need to optimize their productivity (Huang et al. 2003) by improving their production equipment's effectiveness to increase production throughput (Jauregui Becker et al. 2015). This can be done by maximizing equipment availability by minimizing machine downtime.

© IEOM Society International

Machine utilization or overall equipment effectiveness (OEE) is an important KPI in the semiconductor manufacturing industry and has become a significant concern for modern manufacturing technology systems. It is beneficial for performance optimization of the existing capacity for deferring significant capital investments, overtime expenditure reductions, process variability reduction, operator performance improvement, and reduction in changeover times (Esmaeel et al. 2018).

With the increasing competition in the global market, manufacturing companies need to define strategies to stay competitive by providing high-quality, on-time deliveries of affordable products to their customers. This is possible by improving the overall equipment effectiveness and having a strategic cost management plan, which is a continuous drive for improvement. Literature shows different studies on improving revenue, improving OEE/machine utilization, and reducing unplanned/unnecessary costs. Due to the limited studies on the use of big data to see the behavior of overtime in relation to machine utilization, production throughput, and absenteeism, establishing and understanding the relation of overtime to the manufacturing KPIs and absenteeism will be beneficial to companies to control unplanned costs better, maximize production throughput, and optimize machine utilization.

3. Methods

3.1. Multiple Linear Regression

Multiple linear regression was used to test the hypothesis of whether there is a significant (linear) relationship between the response and the predictor variables. It also describes how much of the model's response variable can be effectively explained. The following equation represents the relationship between OTRate and the MFG KPIs:

$$\mathrm{OTRate} = \beta_0 + (\beta_1 * \mathrm{AbsentRate}) + (\beta_2 * \mathrm{Attainment}) + (\beta_3 * \mathrm{Utilization})$$

The established regression model was tested against the assumptions for multiple linear regression. This was to verify whether the output model effectively described the relationship between the response and predictor variables. Throwing data into linear regression without thinking critically may produce confusing and misleading results, regardless of the statistical integrity (Franzco et al. 2014).

3.2. Variance Inflation Factor

Variance Inflation Factor (VIF) measures the severity of multicollinearity in regression analyses. It indicates the increase in the variance of a regression coefficient due to collinearity. VIF can have values greater than or equal to 1, and values greater than 5 are considered problematic. (James et al. 2013).

$$\mathrm{VIF_i} = \frac{1}{(1 - R_i^2)}$$

3.3. Durbin-Watson Test

The Durbin-Watson test is a statistical test used to detect the presence of autocorrelation in the residuals from the regression analysis.

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Durbin-Watson test scores range from 0.0 to 4.0, with 1.5-2.5 indicating an appropriate range; the optimal value is the median point of 2.0, which indicates no autocorrelation between the contiguous values of the variables (Al-Rawabdeh et al. 2021).

3.4. White Test

White Test is used to formally test for homoscedasticity (White, 1980), which tests the following hypotheses:

H0: The residuals are distributed with equal variance (Homoscedastic)
HA: The residuals are NOT distributed with equal variance (Heteroscedastic)

The procedure for conducting the test is as follows:

© IEOM Society International

1.) Fit the regression model
2.) Obtain the squared residuals of the model
3.) Obtain R_X^2 by fitting a new regression model (auxiliary regression) using the squared residuals as the response variable. The predictors should be all the explanatory variables, individual squares, and cross-products.
4.) Calculate the Chi-square test statistic $\chi^2 = (nR_X^2)$, WHERE n = number of observations
5.) Obtain the corresponding P-value for $\chi 2$.

If the P-value is less than the designated significance level, the null hypothesis is rejected, concluding that the residuals are heteroscedastic. Else, the residuals are homoscedastic.

3.5. Shapiro-Wilk Test

Shapiro-Wilk Test is a statistical test for normality (Shapiro et al. 1965). It tests the following hypotheses, and Shapiro-Wilk W-statistic can be computed using the below equation.

H0: The residuals are normally distributed
HA: The residuals are NOT normally distributed

$$W = \frac{\left(\sum_{i=1}^{n} A_i X_{(i)} \right)^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}$$

If the equivalent P-value is less than the designated significance level, the null hypothesis is rejected and concludes that the residuals are not normally distributed. Else, the residuals are normally distributed.

3.6. Pearson Correlation

Pearson Correlation is a measure of linear correlation between data sets (Kent State University N.D). It describes the direction and strength of the relationship between two variables.

$$R_{xy} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sqrt{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}}$$

It can take values between -1 and 1. Values closer to the limits imply a stronger correlation.

$R_{xy} = 0$ implies **NO** correlation
$R_{xy} > 0$ implies **POSITIVE** correlation
$R_{xy} < 0$ implies **NEGATIVE** correlation

To determine the statistical significance of the correlation, a two-tailed T-test is used

$$\mathbf{t = R_{xy} \sqrt{\frac{(n-2)}{(1 - {R_{xy}}^2)}}}$$

3.7. Statistical Tool

TIBCO Statistica, a fast, efficient, and user-friendly data analysis software, was used to establish the regression model and analyze the generated statistical insights. (TIBCO Software Inc. N.D.)

4. Data Collection

4.1 CRISP-DM Model

The Cross-Industry Standard Process for Data Mining (CRISP-DM) framework was introduced in 1996. Its target is to provide a systematic and general approach for applying data mining concepts to analyze industrial operations and gain in-depth insights into business processes (Shearer 2000). It is a dominant framework for industrial data mining and data analytics for data-driven knowledge discovery. CRISP-DM Model is flexible, easily customized, and is an industry-proven way to guide data-mining efforts. The teams trained and explicitly told to implement CRISP-DM performed better than teams using other approaches. Teams that were specifically instructed to adopt CRISP-DM throughout training and, after receiving such instruction, outperformed teams utilizing alternative methods (Saltz et al. 2017). Briefly, CRISP-DM divides the data mining-related knowledge discovery process into six phases (Tripathi et al. 2021). The six phases are illustrated in Figure 1.

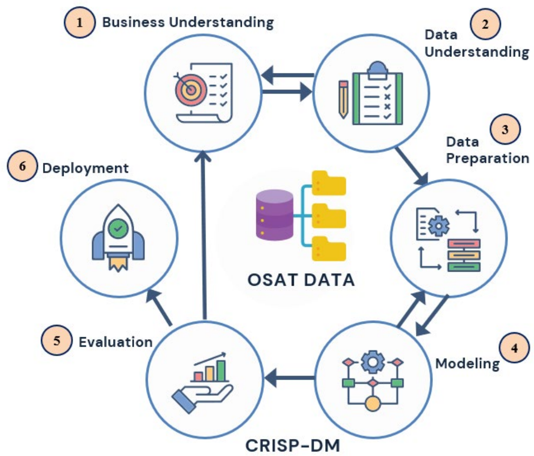

Figure 1. CRISP-DM Framework

4.2. Data Analytics Framework

To have a clear grasp of the retrieved data from diverse applications, including machine usage, production systems, personnel attendance, and consumer demand, the authors used the BDA framework (Mondero et al. 2021) aligned with the CRISP-DM framework. The authors ingested data based on the business rules of the OSAT Company utilizing SQL queries and RPG programs to prepare the "data warehouse." The extracted data also gave structured or unstructured data populated in the "data lake" of the BDA framework. The process flow is illustrated in Figure 2.

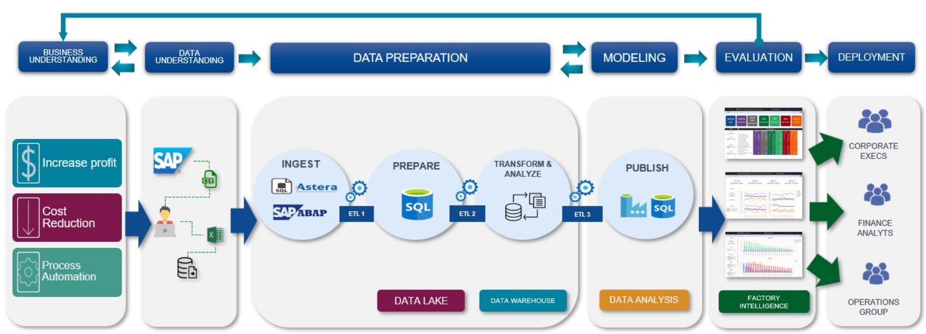

Figure 2. Big Data Analytics (BDA) and CRISP-DM Framework

5. Results and Discussion

5.1. Data Gathering & Sampling

The authors utilized CRISP-DM and judgmental sampling to acquire at least 80 % or more of the total operations goal volume to represent the data population from three (3) manufacturing plants. The top five (5) packages, which represent 97 % of the overall volume, were selected and described as follows:

- Package A is a small outline IC lead frame-based plastic encapsulated package that is well suited for applications requiring optimum performance in IC packaging.
- Package B is a plastic encapsulated with a copper lead frame substrate. This package uses perimeter lands on the bottom of the package to provide electrical contact to the Printed Wiring Board (PWB).
- Package C is an integrated circuit package that is particularly well suited for various electronic systems applications requiring broad performance characteristics.

- Package D is a package that has improved robustness by extending the molding compound encapsulation to the exposed edge areas of the top flange and corners.
- Package E is a type of laminate-based package that is compatible with surface-mounted technology, having single and multi-die layouts, stacked die, and passive component integration.

The study focuses only on the Assembly business line of the OSAT company, specifically on the Wire Bond process. Approximately 3.8 million records for attendance, 25.7 million for lot transactions, and 546.2 million for machine utilization were processed into weekly buckets, resulting in 315 samples. Descriptive statistics are shown in Table 1.

Table 1. Descriptive Statistics for Datasets

PKG	Variable	Mean	Variance	Std. Dev	Min	Q1	Median	Q3	Max
OVERALL	OTRate	25.45	104.72	10.23	4.17	16.85	26.29	34.09	46.64
	AbsentRate	3.18	4.64	2.15	0.00	1.79	2.76	3.92	15.31
	Attainment	98.83	91.25	9.55	63.92	93.85	98.49	103.67	148.87
	Utilization	75.52	29.15	5.40	57.72	71.63	75.75	79.59	86.79
A	OTRate	12.20	10.01	3.16	4.17	9.76	12.50	13.89	19.44
	AbsentRate	4.53	10.94	3.31	0.00	1.59	3.90	6.49	15.31
	Attainment	99.25	278.65	16.69	63.92	88.02	98.05	110.18	148.87
	Utilization	79.58	19.37	4.40	67.72	75.95	80.31	83.20	86.79
B	OTRate	26.91	21.78	4.67	14.58	23.78	27.50	30.45	35.58
	AbsentRate	3.15	3.69	1.92	0.00	1.94	2.71	3.94	11.48
	Attainment	96.51	60.14	7.76	78.98	91.77	96.44	102.34	112.22
	Utilization	73.30	31.13	5.58	57.72	69.95	73.20	78.74	82.01
C	OTRate	18.06	4.09	2.02	12.37	16.83	18.17	19.37	22.06
	AbsentRate	2.88	1.06	1.03	0.97	2.11	2.80	3.59	6.26
	Attainment	97.32	37.07	6.09	83.05	93.15	97.49	102.41	110.86
	Utilization	75.31	10.55	3.25	68.80	72.67	75.82	78.19	80.66
D	OTRate	34.12	32.74	5.72	20.73	28.28	34.83	38.34	45.83
	AbsentRate	2.80	1.37	1.17	0.99	1.94	2.58	3.37	6.55
	Attainment	100.19	24.36	4.94	89.49	96.85	100.11	103.46	113.31
	Utilization	71.18	14.78	3.84	61.44	69.41	71.50	73.29	79.38
E	OTRate	35.98	36.33	6.03	14.19	32.50	37.17	39.58	46.64
	AbsentRate	2.51	3.87	1.97	0.00	1.22	1.90	3.17	9.77
	Attainment	100.88	47.82	6.92	86.99	96.08	101.89	105.48	121.63
	Utilization	78.23	23.38	4.84	67.54	74.56	77.74	82.37	86.53

5.2. Regression Summary (OVERALL)

Table 2 shows the results of Multiple Linear Regression on the three hundred fifteen (315) samples gathered.

Table 2. Regression Summary (OVERALL)

N=315	Regression Summary for Dependent Variable: OTRate R= 0.4452 R^2= 0.1982 Adjusted R^2= 0.1904 F(3,311)=25.623 p<.0000 Std.Error of estimate: 9.21					
	b*	Std.Err. of b*	b	Std.Err. of b	t(311)	p-value
Intercept			45.7519	8.45	5.4148	0.0000
AbsentRate	-0.3646	0.05	-1.7328	0.24	-7.1553	0.0000
Attainment	0.1425	0.05	0.1527	0.06	2.7555	0.0062
Utilization	-0.2088	0.05	-0.3957	0.10	-4.0462	0.0001

© IEOM Society International

The P-value for the F-statistic was less than the significance value of 0.05, implying that at least 1 of the predictors is statistically significant, rendering the regression model useful. The P-value for the T-statistic of each predictor was all less than the significance level of 0.05, implying that all of the predictors and the intercept exhibit a statistically significant relationship with respect to *OTRate*. The Adjusted R^2 value of 0.1904 implies that the model is able to satisfy approximately 19.04% of the uncertainty of *OTRate*. The Std. Error implies that the average data point falls 9.21 units from the regression line. Based on the signs of the beta coefficients (b), it is implied that *Absenteeism* and *Utilization* exhibit an inverse relationship, while *Attainment* shows a direct relationship with *OTRate*.

The inverse relationship between *Absenteeism* and *OTRate* is explained by the workweek schedule scheme used, where overtime is only considered once the required 48 hours of work is exceeded for the week. Higher *Absenteeism* means a lower chance of overtime work, resulting to lower *OTRate*.

The direct relationship between *Attainment* and *OTRate* is attributed to the additional work needed to quickly resolve problems on the production floor. This is necessary to achieve the committed orders for the OSAT customers and increase profitability objectives. In terms of overtime, more work drives an increase in production attainment, considering unplanned productivity losses and target output goals.

The inverse relationship between *Utilization* and *OTRate* is due to overtime work rendered to compensate for the problems on the production floor that impact the machine's productive output, resulting in a higher *OTRate*. On the other hand, increased utilization means reduced productivity losses and unplanned/unnecessary costs, including overtime. The regression equation is now represented as below:

$$\mathbf{OTRate} = \mathbf{45.7519} + (-\mathbf{1.7328} * \mathbf{AbsentRate}) + (\mathbf{0.1527} * \mathbf{Attainment}) + (-\mathbf{0.3957} * \mathbf{Utilization})$$

Other entities can use this to estimate the overtime rate based on the given MFG KPIs. There may be a need to re-compute the coefficients based on the setup of data to produce accurate results.

5.3 Assumption Test Results - Summary (OVERALL)

Violations of the regression assumptions are probably due to model misspecification: other variables can more effectively explain the behavior of the overtime rate, as indicated by the low R-squared values. In contrast, only selected variables are considered significant across the results by package. These differences can be attributed to the different environmental conditions of semiconductor processes and manufacturing lines. Table 3 summarizes the results of the assumption tests for the regression model. Figure 3 shows the Shapiro-Wilk coefficient with its corresponding P-value and a normality histogram. The Shapiro-Wilk results align with the histogram interpretation that the data is not normally distributed.

Table 3. Assumption Test Results - Summary (OVERALL)

PKG	Predictor	VIF		D-W	White	S-W
OVERALL	Intercept	-	-	0.343 FAIL	0.000 FAIL	0.000 FAIL
	AbsentRate	1.01	PASS			
	Attainment	1.04	PASS			
	Utilization	1.03	PASS			

© IEOM Society International

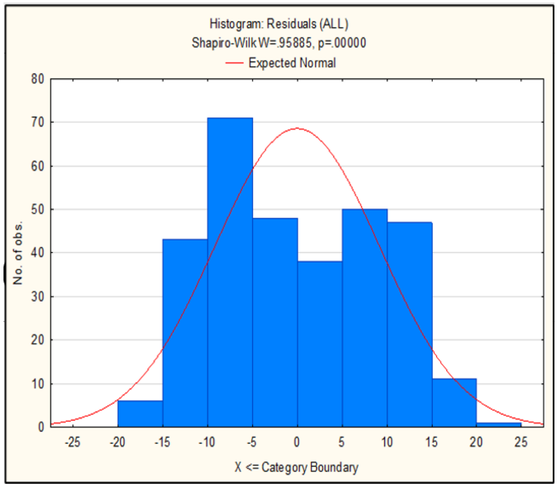

Figure 3. Shapiro-Wilk Graphical Test Result

5.4. Regression + Assumption Test Results by Package

Table 4 summarizes the regression and assumption test results by package, with 63 samples each.

Table 4. Regression + Assumption Test Results by Package

PKG	Predictor	Coeff	P (T)	P (F)	R^2	Std. Err.	VIF		D-W	White	S-W
A	Intercept	17.3238	0.0012	0.0000	0.5293	2.17	-	-	1.692 PASS	0.019 FAIL	0.097 PASS
	AbsentRate	-0.7136	0.0000				1.01	PASS			
	Attainment	-0.0001	0.9954				1.09	PASS			
	Utilization	-0.0237	0.7206				1.10	PASS			
B	Intercept	23.6106	0.0060	0.0001	0.2562	4.02	-	-	1.450 FAIL	0.007 FAIL	0.095 PASS
	AbsentRate	-1.1949	0.0001				1.09	PASS			
	Attainment	0.0757	0.3237				1.33	PASS			
	Utilization	-0.0033	0.9743				1.26	PASS			
C	Intercept	11.0181	0.1008	0.0020	0.1801	1.83	-	-	0.882 FAIL	0.004 FAIL	0.922 PASS
	AbsentRate	-0.7512	0.0036				1.20	PASS			
	Attainment	0.0660	0.1048				1.10	PASS			
	Utilization	0.0370	0.6246				1.10	PASS			
D	Intercept	-28.0041	0.1247	0.0040	0.1599	5.24	-	-	0.605 FAIL	0.008 FAIL	0.182 PASS
	AbsentRate	1.2088	0.0395				1.02	PASS			
	Attainment	0.2252	0.1028				1.01	PASS			
	Utilization	0.5083	0.0051				1.02	PASS			
E	Intercept	14.9702	0.3743	0.0097	0.1327	5.61	-	-	1.618 PASS	0.002 FAIL	0.0000 FAIL
	AbsentRate	-0.9603	0.0109				1.02	PASS			
	Attainment	0.2282	0.0320				1.02	PASS			
	Utilization	0.0052	0.9721				1.03	PASS			

The P-values for the F-statistic of each package is less than 0.05, implying that at least 1 of the predictors is statistically significant, rendering the regression model useful.

For PACKAGE A, the P-values for the T-statistics of *Intercept* and *AbsentRate* are less than 0.05, implying that they exhibit a statistically significant relationship with *OTRate*. The Adjusted R^2 value of 0.5293 implies that the model is able to satisfy approximately 52.93% of the uncertainty of *OTRate*. The Std. Error of estimate implies that the average data point falls 2.17 units from the regression line. The regression model fails the assumption of homoscedasticity as per the White test.

For PACKAGE B, the P-values for the T-statistics of *Intercept* and *AbsentRate* are less than 0.05, implying that they exhibit a statistically significant relationship with *OTRate*. The Adjusted R^2 value of 0.2562 implies that the model is able to satisfy approximately 25.62% of the uncertainty of *OTRate*. The Std. Error of estimate implies that the average data point falls 4.02 units from the regression line. The regression model fails the assumptions of independence and homoscedasticity as per the Durbin-Watson and White tests.

For PACKAGE C, the P-value for the T-statistic for *AbsentRate* is less than 0.05, implying that it exhibits a statistically significant relationship with *OTRate*. The Adjusted R^2 value of 0.1801 implies that the model is able to satisfy approximately 18.01% of the uncertainty of the *OTRate*. The Std. Error of estimate implies that the average data point falls 1.83 units from the regression line. For this package, the model fails the tests of independence and homoscedasticity, as per the Durbin-Watson and White tests.

For PACKAGE D, the P-values for the T-statistics of *AbsentRate* and *Utilization* are less than 0.05, implying that they exhibit a statistically significant relationship with *OTRate*. The Adjusted R^2 value of 0.1599 implies that the model is able to satisfy approximately 15.99% of the uncertainty of *OTRate*. The Std. Error of estimate implies that the average data point falls 5.24 units from the regression line. The regression model fails the assumptions of independence and homoscedasticity as per the Durbin-Watson and White tests.

For PACKAGE E, the P-values for the T-statistics of *AbsentRate* and *Attainment* is less than 0.05, implying that they exhibit a statistically significant relationship with *OTRate*. The Adjusted R^2 value of 0.1327 implies that the model is able to satisfy approximately 13.27% of the uncertainty of *OTRate*. The Std. Error of estimate implies that the average data point falls 5.61 units from the regression line. The regression model fails the homoscedasticity and residual normality assumptions, respectively, as per the White and Shapiro-Wilk tests.

There were cases where some assumption tests were violated for each model. Based on the results obtained, model misspecification was the most probable reason for failing some tests (Ullah 2020) (Armstrong 2016): some variables can explain *OTRate* more effectively. The failing assumptions can be attributed to the different semiconductor processes and manufacturing line environmental conditions in producing the specific packages.

5.5. Pearson Correlation

Table 5 summarizes the matrix of Pearson Correlation against *OTRate.*

Table 5. Pearson Correlation against OTRate

PKG	Predictor	R	T	P	Relationship
OVERALL	AbsentRate	-0.3806	-7.2816	0.00000	NEGATIVE
	Attainment	0.1335	2.3823	0.01780	POSITIVE
	Utilization	-0.1935	-3.4887	0.00055	NEGATIVE
A	AbsentRate	-0.7423	-8.6525	0.00000	NEGATIVE
	Attainment	-0.0145	-0.1132	0.91027	NEGATIVE
	Utilization	0.0474	0.3705	0.71232	POSITIVE
B	AbsentRate	-0.5274	-4.8476	0.00001	NEGATIVE
	Attainment	0.2663	2.1579	0.03488	POSITIVE
	Utilization	0.1415	1.1164	0.26864	POSITIVE
C	AbsentRate	-0.4229	-3.6444	0.00056	NEGATIVE
	Attainment	0.3088	2.5359	0.01379	POSITIVE
	Utilization	-0.0566	-0.4430	0.65932	NEGATIVE
D	AbsentRate	0.1926	1.5332	0.13039	POSITIVE
	Attainment	0.2039	1.6264	0.10902	POSITIVE
	Utilization	0.3316	2.7448	0.00794	POSITIVE
E	AbsentRate	-0.3264	-2.6970	0.00903	NEGATIVE
	Attainment	0.2763	2.2458	0.02835	POSITIVE
	Utilization	0.0099	0.0772	0.93874	POSITIVE

© IEOM Society International

For OVERALL, *Attainment* exhibits a POSITIVE relationship, while *AbsentRate* and *Utilization* show a NEGATIVE relationship. The P-values for all predictors are less than 0.05, implying that their correlation with *OTRate* is statistically significant.

For PACKAGE A, *Utilization* exhibits a POSITIVE relationship, while *AbsentRate* and *Attainment* show a NEGATIVE relationship. The P-value for *AbsentRate* is less than 0.05, implying that its correlation with *OTRate* is statistically significant.

For PACKAGE B, *Attainment* and *Utilization* exhibit a POSITIVE relationship, while *AbsentRate* shows a NEGATIVE relationship. The P-values for *AbsentRate* and *Attainment* is less than 0.05, implying that their correlation with *OTRate* is statistically significant.

For PACKAGE C, *Attainment* exhibits a POSITIVE relationship, while *AbsentRate* and *Utilization* show a NEGATIVE relationship. The P-values for *AbsentRate* and *Attainment* is less than 0.05, implying that their correlation with *OTRate* is statistically significant.

For PACKAGE D, All of the predictors exhibit a POSITIVE relationship. The P-value for *Utilization* is less than 0.05, implying that its correlation with *OTRate* is statistically significant.

For PACKAGE E, *Attainment* and *Utilization* exhibit a POSITIVE relationship, while *AbsentRate* shows a NEGATIVE relationship. The P-values for *AbsentRate* and *Attainment* is less than 0.05, implying that their correlation with *OTRate* is statistically significant.

The relationship between the response and predictor variables fluctuates across packages. This can be attributed to the different semiconductor processes and manufacturing line environmental conditions in producing the specific packages.

6. Conclusion and Future Research

Using three hundred fifteen (315) samples gathered from the five (5) selected packages, representing 97% of the total volume, the authors have performed Multiple Linear Regression and Pearson Correlation to determine whether there was a significant relationship between Overtime Rate and the following manufacturing KPIs: Absenteeism, Attainment, and Utilization.

With the results obtained from Multiple Linear Regression, it can be concluded that Overtime Rate exhibits a statistically significant relationship with all of the aforementioned variables for the overall setup. In contrast, only selected variables are considered significant across the regression models by package. Pearson Correlation also revealed that each setup resulted in various behaviors between the response and the predictors. These differences can be attributed to the different environmental conditions of semiconductor processes and manufacturing lines.

Violations of the regression assumptions are probably due to model misspecification, which indicates that other variables can more effectively explain the behavior of the overtime rate, as indicated by the low R-squared values. To properly model the Overtime rate in the OSAT manufacturing industry, the authors advise looking into additional production KPIs and process environmental factors, such as the man-machine ratio (MMR), machine output rate, cycle time, or machine count. In order to adequately normalize the data and provide more reliable statistical modeling, it is also advised to incorporate additional samples.

© IEOM Society International

References

Abston, K. A., Arik, M., Graves, B. C., Workforce Turnover and Absenteeism in the Manufacturing Sector, *Global Journal of Business Disciplines*, vol. 3, pp 57-79, 2019.

Al-Rawabdeh, H. O., Ghadir, H., Al-Abdallah, G., The Effects of User Generated Content and Traditional Reference Groups on Purchase Intentions of Young Consumers: A Comparative Study on Electronic Products, *International Journal of Data and Network Science*, vol. 5, no. 4, pp. 691-201, 2021.

Armstrong, D., Regression III: Non-Normality and Heteroskedasticity, *University of Wisconsin – Milwaukee – Department of Political Science*, Wisconsin, 2016.

Bishop, D., What Are Common Reasons to Work Overtime?, Available: https://www.smartcapitalmind.com/what-are-common-reasons-to-work-overtime.htm, September 19, 2022.

Bolfek, B., Contribution to the Development of the Direct and Indirect Costs Reduction Model (DICRM-8) in Manufacturing Companies, *Business Excellence*, vol. 15, no. 1, pp. 81–105, 2021.

Bureau of Labor Relations, Available: https://blr.dole.gov.ph/2014/12/11/book-iii-conditions-of-employment/, Accessed on June 1, 2022.

Cascio, W.F., *Managing Human Resources- Productivity, Quality of Work Life, Profits*, 4th Edition, McGraw-Hill, 2003.

Chowdhury, S., A Study of Absenteeism in a Manufacturing Company, *Journal of Indian Management Research & Practices*, pp. 330–351, 2016.

Franzco, P., Casson, R. J., Farmer, L. DM., Understanding and Checking the Assumptions of Linear Regression: A Primer for Medical Researchers, *Clinical & Experimental Ophthalmology*, vol. 42, no. 6, pp. 590-596, 2014.

Dana LP., Rounaghi MM., Enayati G., Increasing Productivity and Sustainability of Corporate Performance by Using Management Control Systems and Intellectual Capital Accounting Approach, *Green Finance*, vol. 3, no. 1, pp. 1–14, 2021.

Department of Labor and Employment, Available: https://www.dole.gov.ph/news/compressed-workweek-voluntary-no-cuts-in-benefits-dole/#:~:text=%E2%80%9CThe%20compressed%20workweek%20is%20not,is%20part%20of%20management%20prerogative., Accessed on June 1, 2022.

Ehrenberg, R., The Impact of the Overtime Premium on Employment and Hours in U.S. Industry, *Economic Inquiry*, vol. 9, no. 2, pp. 199 – 207, 1971.

Esmaeel, R. I., Zakuan, N., Jamal, N. M., Taherdoost, H., Understanding of Business Performance from the Perspective of Manufacturing Strategies: Fit Manufacturing and Overall Equipment Effectiveness, *Procedia Manufacturing*, vol. 22, pp. 998–1006, 2018.

Fleischer, J., Weismann, U., Niggeschmidt, S., Calculation and Optimization Model for Costs and Effects of Availability Relevant Service Elements, *Proceedings of LCE2006*, pp. 675-680, 2006

Fritzsche, L., Wegge, J., Schmauder, M., Kliegel, M., Schmidt, K.-H., Good Ergonomics and Team Diversity Reduce Absenteeism and Errors in Car Manufacturing, *Ergonomics*, vol. 57, no. 2, pp.148–161, 2014

Huang, S., Dismukes, J., Shi, J., Su, Q. K., Alabdulrazzak, M., Bodhale, R., Robinson, D., Manufacturing Productivity Improvement Using Effectiveness Metrics and Simulation Analysis, *International Journal of Production Research*, vol. 41, no. 3, pp. 513-527, 2010.

James, G., Witten, D., Hastie, T., Tibshirani, R., *An Introduction to Statistical Learning: with Applications in R*, 1st Edition, Springer, 2013.

Jauregui Becker, J. M., Borst, J., van der Veen, A., Improving the Overall Equipment Effectiveness in High-Mix-Low-Volume Manufacturing Environments, *CIRP Annals*, vol. 64, no. 1, pp. 419–422, 2015.

Kent State University, Available: https://libguides.library.kent.edu/SPSS/PearsonCorr. Accessed on June 1, 2022.

McHugh, M., French, D. D., Farley, D., Maechling, C. R., Dunlop, D. D., Holl, J. L., Community Health and Employee Work Performance in the American Manufacturing Environment, *Journal of Community Health*, vol. 44, no. 1, pp. 178–184, 2019.

Mondero, J., Caguiat, Ma. R. and Celis, N., Data Analytics for Operations Management in a Selected Outsourced Semiconductor Assembly and Test (OSAT) Manufacturing Company, *De La Salle University Research Congress Proceedings*, vol. 9, 2021.

Muchiri, P., and Pintelon, L., Performance Measurement Using Overall Equipment Effectiveness (OEE): Literature Review and Practical Application Discussion, *International Journal of Production Research,* vol. 46*,* no. 13*,* pp. 3517–3535, 2008.

© IEOM Society International

Rounaghi, MM., Economic Analysis of Using Green Accounting and Environmental Accounting to Identify Environmental Costs and Sustainability Indicators, *International Journal of Ethics and Systems*, vol. 35, no. 4, pp. 504-512, 2019.

Rounaghi, MM. Jarrar, H., Dana, L.-P., Implementation of Strategic Cost Management in Manufacturing Companies: Overcoming Costs Stickiness and Increasing Corporate Sustainability, *Future Business Journal,* vol. 7, no. 31, 2021.

Saltz, J., Shamshurin, I., Crowston, K., Comparing Data Science Project Management Methodologies via a Controlled Experiment. *Proceedings of the 50th Hawaii International Conference on System Sciences*, pp. 1013-1022, Hawaii, USA, January 4, 2017.

Shapiro, S., Wilk, M. B., An Analysis of Variance Test for Normality (Complete Samples). Biometrika, vol. 52, no. 3–4, pp. 591–611.

Shearer, C., The CRISP-DM Model: The New Blueprint for Data Mining, *Journal of Data Warehousing*, vol. 5, no. 4, pp. 13–22, 2000.

Shields, M. and Young, S.M., Effective Long-term Cost Reduction: A Strategic Perspective, *Journal of Cost Management*, vol. 6, no. 1, pp. 16-30, 1992

Society of Human Resource Management, Available: https://www.shrm.org/hr-today/news/hr-magazine/documents/kronos_us_executive_summary_final.pdf, Accessed on June 1, 2022.

TIBCO Software Inc., Available: https://docs.tibco.com/pub/stat/14.0.0/doc/html/UsersGuide/GUID-1B846317-EF4F-449D-A076-572C2F4CFCCB.html, Accessed on June 1, 2022.

Tiwari, U., Impact of Absenteeism and Labour Turnover on Organizational Performance at Iti, Nani Allahabad, India, *Abhinav National Monthly Refereed Journal of Research in commerce & Management*, vol. 34, no. 10, pp. 9-15, 2014.

Tripathi, S., Muhr, D., Brunner, M., Jodlbauer, H., Dehmer, M., Emmert-Streib, F., Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing, *Frontiers in Artificial Intelligence*, vol. 4, pp. 1-20, June 14, 2021.

Ullah, M., Reasons for Autocorrelation, Available: https://itfeature.com/time-series-analysis-and-forecasting/autocorrelation/reasons-for-autocorrelation, October 31, 2020.

White, H., A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, Econometrica, vol. 48, no. 4: pp. 817–838

Biographies

Jonathan D. Mondero, Ph.D. leads the overall daily operations of the Information Technology Division of a multi-national semiconductor manufacturing company. Outside the industry, he is an Associate Professor in the College of Business Administration and Accountancy, Graduate School of Business of De La Salle University Dasmariñas. He earned a bachelor's degree in Electronics and Communications Engineering and a Master's in Business Administration from the University of Baguio. He received a Master's degree in Information Technology from De La Salle University Manila and earned a doctorate from Angeles University Foundation.

Jonathan Anthony Ac-ac, MIT, leads the overall Enterprise Systems of a multi-national semiconductor manufacturing company for the Assembly and Test business. He received a bachelor's degree in Manufacturing Engineering and Management and a Master's in Information Technology from De La Salle University Manila, Philippines.

Shenalie M. Lopena, MIT, is a Business Analyst of a multi-national semiconductor manufacturing company. She received a bachelor's degree in Computer Science from STI College and earned her Master's degree in Information Technology from De La Salle University Manila, Philippines.

Christopher Josh A. Cuevas, ECE, is a Data Analyst of a multi-national semiconductor manufacturing company. He received a bachelor's degree in Electronics Engineering from the University of Santo Tomas Manila, Philippines.

© IEOM Society International