Determining Factors that Significantly Impact Injury Levels in a Production Facility

T. S. Welton, *Semiconductor Manufacturing 6.780*

***Abstract*—Over an eight-month period in 2002, an analysis was performed at Michael D Computer Corporation, to determine how different factors within the production environment at the largest production facility, TMC - such as headcount levels, work day, work times, overtime amounts, and percent of temporary employees - impacted productivity. Using a linear regression model, a prediction equation was created to predict the output in the factory given the above listed factors. While the work was informative, and productivity could be modeled, these factors also impact other non-productivity outcomes, such as product quality defect levels, as well as the number of injuries seen on the factory floor. While the initial publication included models and results that could be used to predict the number of quality defects using the factors listed above, determining the number of injuries was not determined at that time. This is because modeling of injuries in the factory is much more complex because it relies on the use of more complex non-linear regression techniques. This paper uses the production information from the TMC factory, and determines which factors are important in predicting the number of injuries that will exist in the factory by using an advance non-linear regression methodology, known as the Poisson Loss Function. Through the use of this technique, it could be determined, that the major contributors to Injuries in the factory, were the Total Headcount in the factory, the Percentage of Temporary Headcount utilized in the factory, the Shift being worked, and the Day of the Week being worked.**

I. Introduction

Between the months of June and December of 2002, a study was completed at Michael D Computer Corporation on the impact of factors such as headcount levels, work day, work times, overtime amounts, and percentage of temporary employees on the floor productivity in the company's largest production facility, TMC. While the study was primarily intended for use by the staffing department to help determine appropriate headcounts needed in each of the production areas of the factory, all of the factors listed above were pertinent to determining output in the factory. For this reason, a multivariate regression model was created, and a prediction equation that could be used to accurately predict the output of the factory was found.

While this study focused mainly on the impact of adding headcount to the factory productivity, there were other outcomes that could be influenced by changing these factors as well. These non-production factors include quality defects of the products being built and the injury rates of the employees working in the factory. It was believed that while productivity was the main consideration when determining headcount levels, these other non-production outcomes (as well as some others not mentioned) should be addressed as well. While the original publication included analysis on quality defect rates, predicting injury rates in the factory was left unsolved. This is because predicting injuries in the factory is fairly complex.

The reason for the complexity is due to the nature of the injury data itself. The output of the factory was between zero and 18,000 units each shift. That spread was wide enough that it could be considered a continuous Normal distribution. Likewise, the number of defects found at the factory each day could range between zero and 700 units daily. This spread, too, is considered wide enough that the output could be considered a continuous Normal distribution as well. Because of this, traditional linear regression analysis could be done to determine the factory output and the number of quality defects found in the factory each shift.

However, determining the number of injuries that will be seen in a factory is not as straight forward. Over the timeframe of the study, the number of Injuries and Near Misses (referred to as injuries in this document) was not a normal distribution but a Poisson distribution, or a distribution that is described by integer values greater than zero, with possible values of zero, one, two, or three. Because of this, a special non-linear regression model is needed to model this prediction equation, called a Poisson Loss Function. This article walks through the Poisson Loss Function methodology used to determine the significant factors that impact the number of injuries in the factory.

II. Analysis of The Individual Data

In order to determine which factors are important in determining the number of injuries in the factory, data from one quarter (three months) for the TMC factory was gathered and analyzed. First, a qualitative study was performed to look for trends in the injury totals expected under different conditions. Different factors were individually looked at to determine if there were any patterns that could be seen from the raw data. Some of the qualitative analysis is included in the following sections.

A. Factory Output Analysis

As mentioned, one difficulty with predicting the number of injuries in the factory with many of the factors listed, such as headcount or factory output, is that the number of injuries seen in the factory over the time period of the study is Poisson distributed with possible values of 0, 1, 2, or 3. As continuous factors are compared with the actual injury rate, which is Poisson distributed, no clear pattern can be pulled out. A scatter plot that compares the factory output with the number of Injuries and Near Misses seen in the factory is shown below.

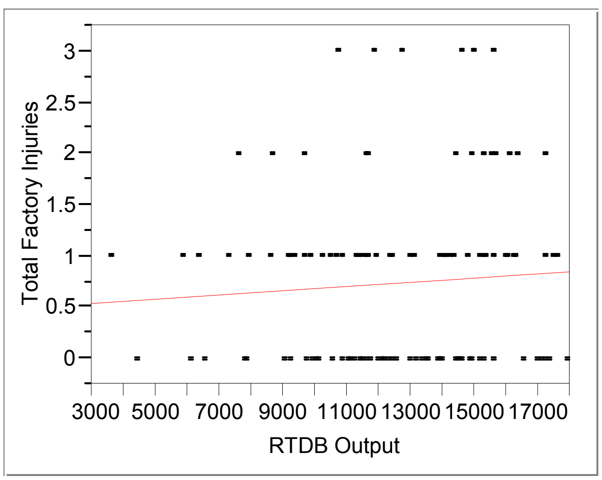


Figure 1 - Scatter plot comparing Factory Output with Factory Injuries. Red Line: Indicates general correlation.

As shown above, the comparison tells little about whether factory output is important in determining the number of injuries that will be seen in the factory. The general correlation between the output per Shift and the number of Injuries or Near Misses shows that the injury rate increases as the output volume increases, however, it is unclear whether the output is significant in determining the number of injuries that will be seen on the production floor.

B. Shift Analysis

Initially, the average number of injuries for each shift was looked at to decide if there was a noticeable difference between them. The following graph shows the output and distribution of injuries given the shift.

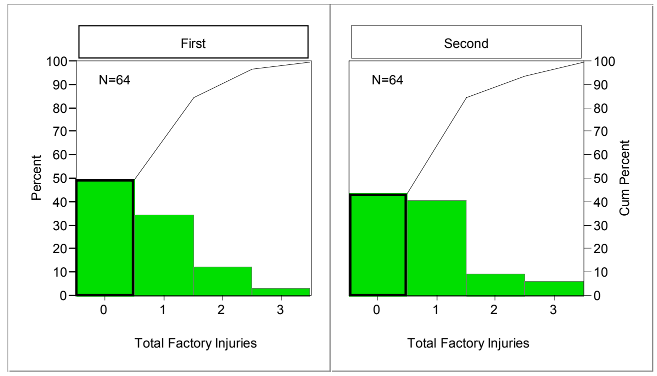
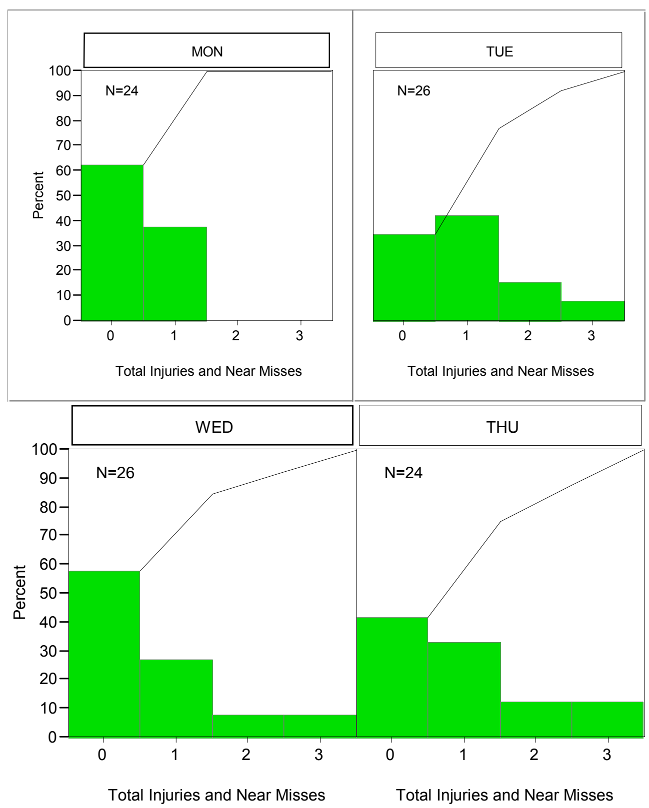


Figure 2 - Injury Distribution by Shift

From the above distribution we see that First shift has a higher percentage of days with zero injuries in the factory. Calculating the average number of injuries for each shift, we find that the First shift averages 0.64 injuries per day, and the second shift averages 0.93. That appears to be a large change, however, it is unclear if that difference is a significant difference.

C. Day of the Week Analysis

The number of injuries was looked at by the day of the week, to see if there were noticeable trends, as was done for each Shift. The following results were found.



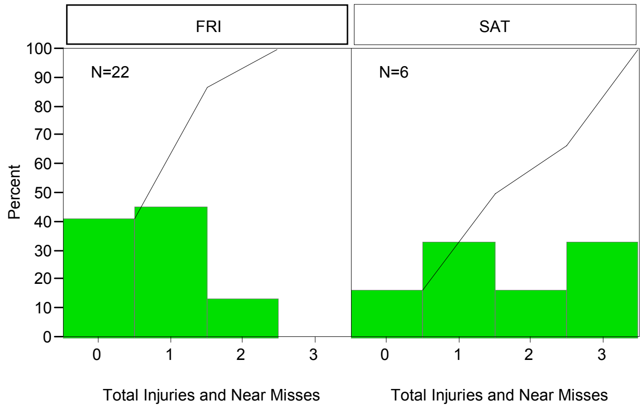


Figure 3 - Injury Distribution by Day of the Week

Looking at the above daily injury and near miss totals we see that Saturday has the largest percentage of days with Injuries or Near Misses. Additionally, Monday has no days with more than one Injury or Near Miss. The average number of Injuries and Near Misses are 0.375, 0.962, 0.654, 0.958, 0.727, and 1.667 per day for Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday respectively. One important observation is that Saturday appears to have the highest injury rate out of all the days worked. Saturdays, however, are only worked on weeks in which an extra day of production was needed to keep up with the pace of demand. If demand was not too high, Saturdays were not worked.

D. Percentage of Temporary Workers Analysis

Much like the correlation between output and the number of Injuries and Near Misses, the impact that the percentage of temporary employees on the floor has on the number of injuries is unclear. As done for the output to injury trend in the above section, a trend was done, using a scatter plot between the number of temporary workers (as a percentage of the total number of workers) compared to the total injuries in the plant. That scatter plot is shown in the Figure below.

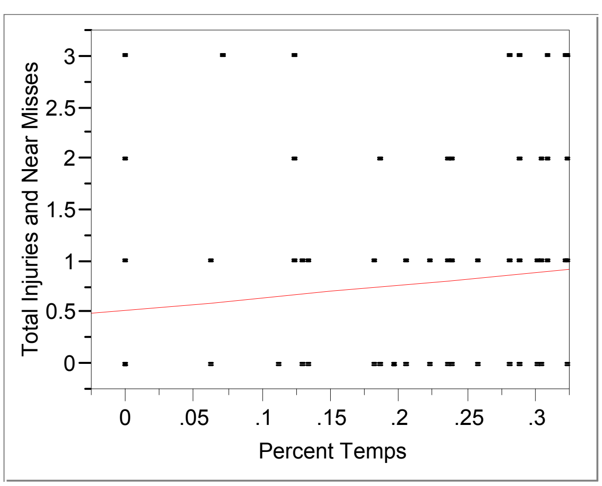


Figure 4 - Scatter Plot Comparing the Temporary Worker Percentage and the Total Number of Injuries and Near Misses. Red Line: shows general trend of the correlation.

What is interesting to note in the above plot is that by looking at this data alone, it would suggest that as the number of temporary workers increases, the total number of Injuries and Near Misses does too. The problem with this plot, is that more temporary workers are generally added when higher production dictates the need for more workers in the factory. For this reason, it is unclear whether the increase in Injuries and Near Misses is attributable to the higher volumes of production or the higher number of temporary workers on the floor. It may be true that one or the other of these factors is significant, but not both.

III. Poisson Regression

While each factor in the above section was looked at in isolation, identifying trend using that data was not reliable. A better model was needed to help identify which factors are significant in determining the Injury and Near Miss rate in the factory, and would account for other factors as the estimates are determined. To do this, a Poisson Regression is used.

Poisson Regression Models are generally used when the output data of a prediction equation is skewed, non-negative, and where the variance increases as the mean increases. Each of these might traditionally give regular linear regression models trouble. However, Poisson regression uses a log transformation called a Poisson Loss Function, which adjusts for the skewness and prevents the model from producing negative predicted values. Additionally, Poisson regression models the variance as a function of the mean, so the variance should not vary greatly as the mean increases.

As mentioned a Poisson distribution is used to model frequency counts, as it is in this case. A Poisson distribution can be described with the following formula.

(1) $$P(Y=k) = Exp(-\mu)*\mu^{n} / n!$$

In the above equation, μ , can be a single value, or a linear model with many factors. To determine the Poisson Loss Function for our model, which is used to produce non-negative numbers), the log of the equation 1 is used. This equation is shown below:

(2) $$Ln(Y) = -(N*model - \mathrm{Exp}(model)) - Log(\Gamma(N+1))$$

In the above equation, Γ is the gamma function (Γ(n+1)=n!) and e^{model} represents μ. As mentioned above, μ , could be a single value or a linear model. For the Poisson Regression, this value, also called the model, is substituted with the linear model composed of the different factors that will be used to determine the number of Injuries and Near Misses in the factory. Given the Poisson Regression model and loss function as described above, the model for the Injury prediction can be determined.

The following sequence was followed to run this model. First, the factors used to predict the Injuries and Near Misses were chosen. Secondly, a model was built that included the factors that were selected, as well as the Poisson Loss function. After that, the model was run and its output was analyzed. Within the data the significant factors were determined, and the prediction equation for the model was determined. The model as a whole could be analyzed as well for fit. Finally, the predicted number of Injuries and Near Misses could be estimated given the prediction equation. Each of these process steps is discussed in detail below.

IV. Choosing the Input Factors

To create a Poisson Regression model using the data, a list of possible factors that would be used to predict the number of Injuries and Near Misses in the factory. Over time, the data fields were narrowed down considerably. A table, which includes each type of data as well as a description of that data type, is included in Table 1.

Variable	Description
Injuries and Near Misses	This variable is the target of the study. It is a count of the Injuries that occurred in the factory, as well as incidents recorded that could have resulted in an injury on the factory floor, but was avoided. This information is collected by shift.
Shift	This is the shift of the factory for which the injury data was collected.
Total Headcount	The total headcount of the TMC factory floor, collected by the shift.
Hours in the Shifts	This is a measure of the total number of hours worked in each shift.
Total Output	This measures the total output, as the number of units shipped from the factory, for each shift.
Percent Temporary Employees	This measures the number of temporary employees used in the factory, as a percent of the total workforce. This would be used to measure the impact of using temporary employees on the number of injuries seen in the factory.
Day of the Week	This is the day of the week for the model.

Table IV-1 - Factors used in Poisson Regression

As mentioned, the above table includes the information that was included in the model to predict injuries in the factory. There were other combinations of data tested as well that were dropped. One factor was Percent Overtime, which was calculated by determining the number of hours of overtime as a percentage of total shift hours. A factor called Labor Hours was tested as well. This factor was calculated by multiplying the number of workers with the total number of hours worked during the shift. These, as well as statistics, such as Number of Units produced per person per shift, or number of units built per person per hour were used, but dropped from the final model, as they were found to not be significant.

V. Determining the Model and Loss Function

In order to create the Poisson Regression Model, two terms need to be created. First the model itself needed to be defined, and also the loss function must be determined. The model creation was a bit tricky. In this case, the data contained both categorical data, as well as continuous data. Categorical data is defined as data with a subset of choices that are possible. For this model, the Shift was categorical, because it contained only two choices for shifts, First and Second. Additionally, the Day of the Week was categorical as well. The set of possible categories for Day of the Week was Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday. The Total Headcount in the factory, the number of Hours in the Shift, the Total Output of the shift, and the Percent Temporary Employees were all continuous. They could all be feasibly any non-negative value. With that information known, the model for the function could be created. The results for the model are shown in the figure below.

$$(3)\quad model = Intercept + MatchShift + b1*TotalHeadcount + b2*HoursInShift + b3*Total\ Output + b4*PercentTemporary + MatchDayOfWeek$$

where *MatchShift* is defined as the function:

Match(Shift)("Second" -> Second =1
Else -> 0)

and where *MatchDayOfWeek* is defined as the function:

Match(DayofWeek)("MON" -> Monday =1,
"TUES" ->Tuesday =1,
"WED" -> Wednesday =1,
"FRI" -> Friday = 1,
"SAT" ->Saturday = 1,
else -> 0)

The above model is the regression model for the Poisson regression. Additionally, the loss function is needed. As introduced in Equation 2, the Poisson Loss Function for this regression can be seen below:

$$(4)\quad Ln(Y) = -(InjuriesandNearMisses*model - Exp(model)) - Log(\Gamma(InjuriesandNearMisses+1))$$

This loss function guarantees that the predicted number of Injuries and Near Misses will be a positive value. Given the model equation and the Poisson Loss Function as shown above, the Poisson regression can now be run. The results for the regression model are described below.

VI. Poisson Model Results

Given the above equations, the Poisson Model was run. Initially, the coefficients were all set to zero, and the intercept was set to the value of one. With these set, the model iterates though the different input values, and determines a best estimate for each of the coefficients and the intercept, which minimizes the sum of the residuals from our real results. The results of the Poisson Regression model are shown below.

Parameter	Current Value
Monday	-0.836201584
Tuesday	-0.011655563
Friday	-0.031761855
Saturday	1.1016554812
Wednesday	-0.422011496
Second	1.0482761722
Intercept	-0.12456072
PercentTempCoeff	-7.710753746
HrsWorkedCoeff	0.0676430084
HeadcountCoeff	0.8786549113
OutputCoeff	-0.008296988

SSE	87.957992518
N	128

Table VI-1 - Poisson Regression Model Effect Estimates

Looking at the above output, it can be seen that there were 128 data samples that were used to model this data. Each Day has its own effect estimate, except for the default case, Thursday. The effect for Thursday was equal to zero. Likewise, there is an effect (or parameter) estimate for the Second shift of 1.048. The First shift would be considered the default shift, and so the First shift would have an effect estimate of zero. Finally, the PercentTempCoeff was the effect estimate for the Percent of Workers that were Temporary employees, the HrsWorkedCoeff was the effect estimate for the Number of Hours Worked in the Shift, the HeadcountCoeff was the effect estimate for the Total Headcount, and the OutputCoeff was the effect estimate for the Total Output in the factory.

In this model, counting the number of data samples and subtracting one for every coefficient and intercept estimate in the model can determine the degrees of freedom. In this case, there were 128 samples, and 11 parameter estimates. Therefore, there are 117 degrees of freedom for this model. This becomes important as the model accuracy is judged.

To judge the overall model accuracy against the actual values predicted, the SSE number is relied on. The SSE is the Sum of Squares Estimate for the Model and is estimated to be the variation of the residuals of the model, or difference between the actual and predicted Injury and Near Misses. In order to the standard deviation of the residuals, also referred to as the MSE or Mean Square Estimate, the Sum of Squares Estimate is divided by the degrees of freedom in the model. In this case, the SSE is roughly equal to 87.95. Dividing 87.95 by 117 yields a MSE value of 0.867. This means our prediction will estimate the number of injuries in the factory within 0.867 injuries 68 percent of the time. While not completely accurate in this case, there are still some key learnings that we can pull from the data.

A. Prediction Equation

Given the above parameter estimates, a prediction equation can be generated to predict the output for the model. To predict the output, the following equation can be used.

$$(5)\quad Ln(Injuries) = Intercept + ShiftCoeff \\ + HeadcountCoeff*ScaledHeadcount \\ + HrsWorkedCoeff*HoursInShift \\ + OutputCoeff*ScaledOutput \\ + PercentTemp*PercentTemporary \\ + DayOfWeekCoeff.$$

In the above equation, Intercept, *HeadcountCoeff*, *HrsWorkedCoeff*, and *OutputCoeff* are given in the parameter estimates. Additionally the *ShiftCoeff* will be set to zero for the First shift, and 1.05 for the Second shift. Finally, the *DayOfWeekCoeff* is zero if the day of the week you are modeling is Thursday, or is found by matching the day of the week you are modeling with the parameter of the same name in the Parameter list and substituting the Current estimate value given in Table VI – 1 for the *DayOfWeekCoeff*. For example, if the day you are modeling is Wednesday, the *DayOfWeekCoeff* will be approximately equal to –0.422.

As shown through the equation, a scaled headcount and output level was needed to compute the model for the function. This is because the Poisson loss function that uses the model actually is exponentially larger than the model itself. If the model is calculated to be too high, the Poisson argument grows too quickly, and a model cannot be determined from the Poisson Regression. To handle this problem, a scaled headcount (*ScaledHeadcount*) and output (*ScaledOutput*) are determined. *ScaledHeadcount* is calculated as follows:

$$(6)\quad ScaledHeadcount \quad = (Total\ Headcount - 750)/100$$

Likewise, ScaledOutput is determined with the following equation:

$$(7)\quad ScaledOutput = (Total\ Output - 12000)/1000$$

Using these values, the Poisson Loss function will not grow arbitrarily large, and a Poisson regression model can be calculated.Finally, by knowing the values for each of the factors and the coefficients the Ln(Injuries) is found.[1] The number of Injuries or Near Misses can be determined by

[1] I am not certain of this estimate. By looking at multiple sources on the web on Poisson Loss functions, I found that the equation was equal to: $\text{Log } L(y_i,B_i) = \text{Sum}\{-\exp(x_i,B_i) + y_i\,(x_i,B_i) - \log(y_i!)\}$. This implied to me that the result could be found by $\exp(\text{Log } L(y_i,B_i))$.

taking the exponential of the the Ln(Injuries) estimate. Using this as the basis for the Poisson regression, the output can be estimated, and the residuals analyzed.

B. Residual Analysis

While the model parameter estimates have been calculated, the residuals need to be tested for randomness. If the residuals were not random, that would suggest that a significant factor of the model was not accounted for and a better regression fit might be possible. The following graph shows the residuals for the Poisson model.

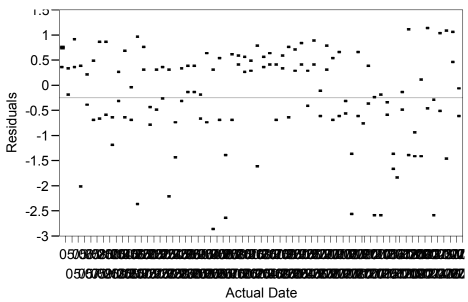


Figure 5 - Analysis of the Residuals

Given the above scatter plot, the residuals follow a mostly random pattern. There is some concern that the there are strong outliers on the low end of the model that are not explained. Additionally, the residuals tend to fall around the 0.5 and -0.5 values. This might suggest that a better model might be possible.

VII. Determining Significance of the Coefficients

Given the model and loss function in Section V, the effects of the parameters were estimated for each of the parameters in the model. With those parameter estimates, a prediction equation was generated that was used to predict the number of Injuries and Near Misses that would be expected given the changing values of our parameters. Using that prediction equation, the number of Injuries and Near Misses were predicted for each of the unique 128 data points was calculated and compared with the actual value. The residuals were analyzed as well. While that information is important, there is more information that we can obtain from this model. In addition to the parameter estimates, the error is measured for each effect as well. The following is the table of effects and the standard error for each parameter.

Parameter	Estimate	ApproxStdErr
Monday	-0.836201584	0.4054583
Tuesday	-0.011655563	0.29151622
Friday	-0.031761855	0.34095551
Saturday	1.1016554812	0.45949955
Wednesday	-0.422011496	0.32101172
Second	1.0482761722	0.38531737
Intercept	-0.12456072	1.68795816
PercentTempCoeff	-7.710753746	3.65005422
HrsWorkedCoeff	0.0676430084	0.18066858
HeadcountCoeff	0.8786549113	0.4391271
OutputCoeff	-0.008296988	0.09658109

Table VII-1 Parameter Estimates with Standard Error

Knowing the standard error for each parameter helps to identify the effects that are significant within the model. Generally, speaking the standard error term is the estimate of the standard deviation of the error for each parameter. In order to determine if a parameter is significant, an alpha, α, is chosen. Using the alpha value, and using the standard deviation of the model around the parameter estimate, confidence limits can be established in which there is a $1-\alpha$ certainty that the true estimate lies within the confidence limits. For example, if an alpha is chosen to be 0.05, the confidence limits will be determined to ensure that there is only a five percent chance that the actual estimate value will lie outside the confidence limits. Given the properties of Normal distributions, if the alpha value is set to 0.05, that is equivalent to saying the actual estimate lies between +/- 2 standard deviations of the predicted estimate.

The confidence limits, as discussed above, were calculated for the parameters in the Injury and Near Miss Poisson regression model. Their values are now shown in Table VII-2.

Parameter	Estimate	ApproxStdErr	Lower CL	Upper CL
Monday	-0.836	0.405	-1.681	-0.072
Tuesday	-0.012	0.292	-0.584	0.566
Friday	-0.032	0.341	-0.716	0.630
Saturday	1.102	0.459	0.152	1.967
Wednesday	-0.422	0.321	-1.066	0.202
Second	1.048	0.385	0.299	1.812
Intercept	-0.125	1.688	-3.360	3.277
PercentTempCoeff	-7.711	3.650	-14.885	-0.542
HrsWorkedCoeff	0.068	0.181	-0.300	0.410
HeadcountCoeff	0.879	0.439	0.025	1.750
OutputCoeff	-0.008	0.097	-0.187	0.192

To determine the Lower confidence limit for each of the parameters, the standard error was multiplied by two and subtracted from the estimate. Likewise, to determine the Upper confidence limit the standard error was multiplied by two and added to the effect estimate for each parameter. Knowing this relationship, a parameter is said to be significant if the zero value does not fall within the confidence limits established above. If the zero value actually falls within the confidence limits, there is a chance that the parameter actually has zero impact on the output determined by the model.

To start the analysis of our effect predictions, the categorical variables were looked at. Looking at the Day of the Week parameters and estimates, it can be seen that there are two days that are considered significant. First, Monday is shown to have significantly impact on the number of Injuries and Near Misses. The effect is predicted to be –0.836, and the confidence limits, which each sit 2 standard deviations from the predicted effect, are –1.681 and –0.072. Because the zero value does not fall between the Upper an Lower confidence limits, it can be said that there is greater than 95% certainty that Monday produces fewer Injuries and Near Misses than other days. Additionally, it can be seen above that Saturday has a significantly higher Injury and Near Miss rate than other days of the week. The estimated effect is 1.102 extra Injuries or Near Misses and the confidence limits are 0.152 and 1.967 for the lower and upper confidence limits respectively. It can be said that Saturday has a significantly higher defect rates than other days of the week. Of the remaining days in the model, none can be said to have a significant impact on the number of Injuries or Near Misses in the factory, because the zero value falls within the confidence interval for all of these days.

Similar to the Days of the Week, the estimate of the effect for Second shift can be determined to have a significantly higher Injury and Near Miss rate than the First shift does. In this analysis it was found that the effect estimate for the Second shift was 1.048, and its confidence limits were 0.299 and 1.812 for the lower and upper limits respectively. This means that Second shift has significantly higher defect rates than First shift does. This might cause Michael D Computer Company to look for ways to make the second shift a safer environment to work in. Whether by fatigue, or poor training, Second shift incurs higher number of injuries than the first shift does.

With the analysis of the categorical data complete, similar analysis can be completed on the continuous estimates in the model. First, it can be seen that the Number of Hours Worked and the Total Output level has negligible effect on the number of Injuries and Near Misses experienced in the factory. Both of these parameters had upper and lower confidence limits that fall on either side of the zero value. Looking at Total Output first, we see the confidence limits are –0.187 and 0.192 for the lower and upper limit respectively. Likewise, the confidence limits for the Number of Hours Worked are –0.30 and 0.41 for the upper and lower confidence limit respectively. Since the zero value falls within these limits for both variables, they are both considered to be insignificant variables.

Looking at the Percent of Workers that are Temporary, we find that as the percentage grows, there is a significant drop in the amount of Injuries and Near Misses found on the factory floor. In this case, the effect estimate is –7.71 and the confidence limits are –14.885 and –0.542 for the lower and upper limits respectively. Because zero does not fall within the confidence limits, the effect is a significant effect on the overall Injury and Near Miss average. If using the –7.71 estimate, a 10% raise in the percentage of workers that are temporary would statistically reduce the injury total by 0.77 injuries per shift. Once important note is that the temporary workers are added to the headcount as demand dictates, and so the injury rate drops in conjunction with higher demand.

Finally, the number of headcount found in the factory has a significant impact on injury rates as well. The estimate for the headcount effect is 0.879, and the confidence limits are 0.025 and 1.75 for the lower and upper confidence limits respectively. This implies that adding headcount increases the amount of Injuries or Near Misses in the factory. In this case, if the headcount of the factory were to raise by 100 workers the impact would be an Increase in Injury and Near Miss Incidents of 0.879 per shift.

Given that knowledge from the last two observations, adding headcount in the factory increases the injury rate in the factory, however, if temporary headcount is added, there will be a corresponding drop in the injury rate as well. Generally, temporary employees are added as demand volumes dictate. The drop in the injury rate due to the addition of the temporary headcount could be due to more concentration of employees on their work during high demand times or due to the temporary headcount being more attentive to their environment at work. Whatever the reason, the data shows that the times in which the headcount has a higher percentage of temporary workers there is less chance for injuries in the factory. Likewise, the addition of employees happens as the volumes increase as well. This could be a result of workers getting in each other's way or congestion. Whatever the reason, higher total headcount has a big impact on the amount of injuries seen in the factory as well.

VIII. Sensitivity Analysis

Given the confidence intervals above, some sensitivity analysis can be done to determine best and worse case scenarios for the models. Knowing the effects of each parameter as well as the confidence intervals, a manager at Michael D Computer Company can determine best and worst case scenarios for Injury and Near Miss Rates found in their plant for each of the different effect estimates. As an example, assume the day of the week effect and shift effect are effects that management believes could be incorrect. With this analysis, they can adjust their predictions to account for their lack of confidence in these two terms. Assuming the day of the week is Wednesday, and the shift being worked is Second shift. If the prediction equation for the model predicted that there would be three injuries, but the managers of the factory want to predict the best and worst cases given their lack of confidence in the Day and Shift estimates, the managers could adjust the expected Injury and Near Miss estimates to reflect their optimism or pessimism. In this case, the adjustment would be as shown

Factor	Avg Case	Best Case Adjustment	Worst Case Adjustment
Wednesday	-0.422	-0.642	0.642
Second Shift	1.048	-0.385	0.385
Total		-1.027	1.027

The best-case estimate could be as low as 1.973 injuries for the shift on Wednesday, or the worst case scenario could reflect an Injury Rate as high as 4.027 injuries.

IX. Conclusion

In the above analysis, a problem in determining the number of Injuries that could be expected in the TMC factory was discussed. Qualitatively, a number of factors that might contribute to estimating the number of injuries were looked at in isolation of the other factors.

After these factors were looked at individually, it was shown that there is a significant difference in the way regression is performed between Normal linear regressions and Poisson distributed regressions. Additionally, the methodology for determining the Poisson regression estimates was discussed as well. Knowing the methodology, the model for predicting the number of Injuries and Near Misses in the TMC factory was determined. The first output of the model was the effect estimates for each of the input factors for the model. After the effects were estimated, a prediction equation was derived that could be used to predict the number of injuries given the varying factors that determine the injury rate. After the prediction equation was determined, the predicted output was determined for all historical data in our model, and these predictions were compared to the actual outputs, through the use of residual analysis. A confidence level was determined for the model by recognizing the standard deviation of the residuals as predicted by the model was 0.867, meaning 68% of the time, our estimate will be within 0.867 injuries of the actual value.

After the residual analysis was complete, the confidence limits were determined for each individual factor in the model. By looking at the confidence limits, the significant factors, including two days of the week , Saturday and Monday, the Shift, Total Headcount and Percent Temporary workers could be found. Finally, given the confidence limits, sensitivity analysis was discussed, that managers could use to come up with conservative or aggressive estimates for their Injury and Near Miss predictions. Given these results, a manager now has the tools needed to help him understand what factors impact the number of injuries in the factory. Better data may provide better estimates.

.

REFERENCES

[1] G. Geoffrey Vining, Statistical Methods for Engineers, Pacific Grove, CA, Duxbury Press, 1997

[2] http://www.stat.uconn.edu/~nalini/intern03/notes/poissonreg2.pdf, Poisson Regression Fitting, May 13, 2003

[3] http://www.econ.kuleuven.ac.be/tew/academic/kwantmet/members/christophe/courses/busecon/poisson.pdf, The Poisson Regression Model, May 14.2003

[4] JMP4 Users Manual, Poisson Loss Function, 2002

[5] Montgomery, Peck and Vining, Introduction to Linear Regression Analysis, 3rd Edition, New York, New York, Wiley Interscience Publications, 2002.

[6] http://www.cmh.edu/stats/model/poisson.asp, May 13, 2003