Improving Shop Productivity in F&B Store Using Regression Models

Peng Ho Oh, Jhun Boon Tan, Jessie Yap

Singapore Management University

Abstract

The ability to identify performance gaps and areas for improvements is a critical issue for business leaders, who usually rely heavily on intuition to tackle business problems. Our paper seeks to use regression analysis as a tool to aid management in evidence-based decision-making. We utilised linear regression, multi-linear regression and one-way ANOVA models to test five different intuitive hypotheses on improving shop productivity of a Japanese F&B shop, Teppei Syokudo. We collected 2,996 hourly observations of shop sales and customer number, as well as the number and type of staff who are present during each hour. Our analysis involves testing whether several independent variables have significant impacts on shop productivity. One such example is whether the presence of a manager would motivate the staff to upsell more to customers and therefore increase the average sales dollars brought in per customer, ultimately increasing shop productivity.

We were able to identify five potential improvement opportunities for Teppei Syokudo to implement in order to improve shop productivity. Moving forward, we recommend that Teppei Syokudo take these three actionable steps.

1. *Set up controlled experiments to validate the results of these tests. One such experiment could be to staff the store with full-time cashiers and part-time cashiers on separate time periods to validate the results of this model. If it is proven that full-time cashiers earn more sales dollars per customer than part- time cashiers, then Teppei Syokudo can consider assigning full-time staff as cashiers.*
2. *Explore the feasibility of having managers to check into the store periodically since their continued presence do not significantly impact shop productivity.*
3. *Explore the feasibility of decreasing the number of labour hours on weekend lunch peaks and weekday idle periods.*

1.0 Introduction

Overall sales, profits, and customer volume are the basic targets for food and beverage (F&B) stores. The challenge comes when attempting to achieve these targets. Sales, profits, and customer volume can be easily affected by external factors such as time period, day of the week, location, and weather. One way to achieve store targets is through shop productivity. This paper looks at how shop productivity can be improved in an F&B setting, by identifying the key drivers of shop productivity using various regression models.

Shop productivity in F&B can be defined as the dollar amount produced per hour of work. There are three key quantitative variables that contribute to shop productivity. The shop productivity formula is as follows:

$$\frac{\dfrac{\boldsymbol{Sales}}{\boldsymbol{Number\ of\ Customers}} \times \boldsymbol{Number\ of\ Customers}}{\boldsymbol{Number\ of\ Hours\ Worked}}$$

Sales	The total sales dollars earned in a time period (e.g. one hour)
Number of Customers	Number of customers served in a time period (e.g. one hour)
Number of Hours Worked	Total number of staff working during that hour

We will be exploring how certain independent variables affect these one or more of these constituents of shop productivity. One example is the type of staff working in the shop. A full-timer may be more effective at bringing in sales per customer served as compared to a part-time staff as he may feel more committed to his or her job. Another example is the presence of managers in the shop. Intuitively in the presence of a manager, a staff may consciously perform better compared to when a manager is not present in the store.

To identify the factors that affect shop productivity, regression analysis can be used. Regression identifies the strength and direction of relationships between dependent and independent variables. When extrapolated, it forecasts values of the dependent variable based on the strength and direction of relationship with the independent variables. Regression has been used in the fields of econometrics and law (Sykes, n.d.), in improving students' performance

(Zakhem, Khair, & Moucary, 2011), and in setting health targets (Fukada, Nakamura, & Takano, 2002).

In performing regression analysis, data-points are first plotted in a scatter plot. Scatter plots allow the user to visually identify relationships between variables. The correlation coefficient is then found. The correlation coefficient is a standardized number between -1 and 1 that describes the strength and direction of relationship between variables. A correlation coefficient of 1 or close to 1 shows that there is a strong positive relationship between the variables. A correlation coefficient of -1 or close to -1 shows that there is a strong negative relationship between the variables. A correlation coefficient of 0 or close to 0 shows that there is no or weak relationship between the variables. Using the least-squares method, the regression line and the regression equation is found. The regression equation allows us to forecast or predict the dependent variable. Regression can be split into two types: linear and non-linear. Under each type, there are simple and multiple. In this paper, we will be using the multiple linear regression model.

Multiple linear regression allows for examining how multiple independent variables are related to a dependent variable (Higgins, 2005). R (also known as the multiple correlation coefficient) is used in multiple linear regression. It represents the strength and direction of relationship amongst a combination of variables. This differs from the simple linear correlation coefficient which only compares between two variables. In knowing the strength of relationship amongst all the variables, the multiple regression formula is formed. The multiple regression formula is as follows:

$$\boldsymbol{Y = a + b_1X_1 \ + \ b_2X_2 \ + \ \dots \ b_kX_k}$$

Y – The value of the dependent variable

a – The intercept

b – The change in Y for each incremental change in X

X – The value of the independent variable

Care must be exercised when selecting independent variables for regression models. The variables must be truly independent of each another for the regression model to work. When independent variables are correlated with each other, the model is said to have multi-collinearity.

Severe multi-collinearity can increase the variance of the coefficient estimates and make the estimates very sensitive to minor changes in the model (Farrar and Glauber, 1967). Another common problem that affects the independence of model variables is autocorrelation. The presence of significant autocorrelation in the model signals a statistical dependency between values of the same variables (Getoor, 2007). In the context of time series data, this refers to correlation between a variable's past and future values. Autocorrelation complicates the application of statistical tests by increasing the number of dependent observations. When severe multicollinearity or autocorrelation is observed in a regression model, steps should be taken to adjust the model to reduce or remove them.

2.0 The Company

Teppei Syokudo is a Japanese F&B chain, operating under the umbrella of the famous Teppei Japanese Restaurant. It operates as a takeaway kiosk concept, and specializes in Kaisendon, a bowl of barachirashidon (rice with raw fish). It offers other dons (rice bowls), tonkatsu cutlets (fried pork), tempura (fried dishes) and sashimi salads. The takeaway kiosk mainly targets lunchtime crowds, especially upmarket businessmen and women who are looking to grab a quick, healthy, and quality lunch. There are currently 4 outlet locations - Millenia Walk, Takashimaya, Republic Plaza and Ion Orchard. For this paper, we will only be analysing data from the Millenia Walk outlet.

Each outlet has its own shop manager, cashier, packer, and kitchen staff. During peak hours, the outlets may have up to seven staff to manage demand. Some staff work on a part-time basis. The store manager decides the staff roles for the day, hence some staff may be assigned different roles on different days. However, the job rotation usually occurs between part-timers, and rotation between cashier and packer role.

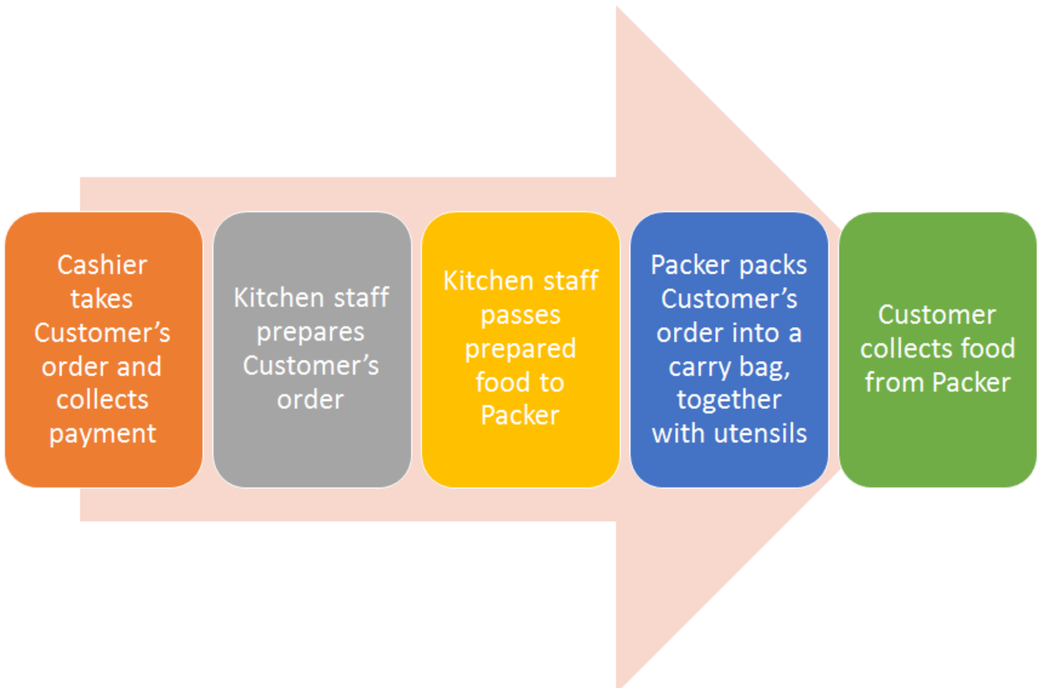


Figure 1. Sales Process

2.1 Problem Definition

One measure of shop performance is shop productivity, which can be defined as the sales dollar per hour worked, or further broken down into three constituent elements namely, sales per customer served, number of customer served and total number of labour hours in the shop. We seek to explore the effect of various variables such as the presence of managers, and quality of work by part-timer versus full-timer on the three constituent elements of shop productivity. We have come up with five hypotheses and will evaluate them using the JMP tool to identify the drivers for shop productivity.

1. We can increase shop productivity by hiring good cashiers who can upsell (increase sales dollar per customer) and serve customers faster (increase customer number).
2. We can increase shop productivity by motivating staff to work harder with the presence of managers.
3. We can increase shop productivity by increasing the number of full-time staff or managers present as they are more productive compared to part-time staff.
4. We can increase shop productivity by staffing full-timers or managers as cashiers as they are better at upselling or serving more customers than part-timers.

5. We can increase shop productivity by decreasing the number of staff on time periods where there is excess capacity.

3.0 Methodology

3.1 Data Exploration

Our dataset comprises of 48 staff and their hourly labour records from 1st June 2015 to 31st December 2015. As the shop is opened seven days a week from 9:00am in the morning to 23:00 at night, we collected 2,996 hours' worth of data over 214 days. The "Date" column indicates the date for each hourly data point, and the "Day" column indicates the corresponding day, from Monday to Sunday, as well as Public Holidays.

We allocated a number from 0 to 1 to each staff for his presence during each hour. For example, if the staff was present the full hour, he will be allocated a "1". He will be allocated "0" for the hours that he was not present and "0.5" for the hours that he was present for only half the time.

Hourly store performance figures such as store sales ("Sales" column) and customer numbers ("CustNo" column) were also included after examining data from the POS system.

Date	Day	Start Time	Sales	CustNo	Staff 1	Staff 2	Staff 3
4-Jul-15	Sat	9:00	$374.64	80	1	0	0
4-Jul-15	Sat	10:00	$95.14	33	1	0	0
4-Jul-15	Sat	11:00	$466.68	28	1	0	0
4-Jul-15	Sat	12:00	$319.08	22	1	0.5	0
4-Jul-15	Sat	13:00	$76.51	25	1	1	0
4-Jul-15	Sat	14:00	$332.75	52	1	1	0
4-Jul-15	Sat	15:00	$280.77	44	1	1	0
4-Jul-15	Sat	16:00	$254.80	45	1	1	0
4-Jul-15	Sat	17:00	$29.58	49	1	1	0
4-Jul-15	Sat	18:00	$471.35	7	1	1	0
4-Jul-15	Sat	19:00	$304.28	77	1	1	0
4-Jul-15	Sat	20:00	$387.29	46	1	1	0
4-Jul-15	Sat	21:00	$43.76	68	1	0.5	0
4-Jul-15	Sat	22:00	$127.58	93	0	0	0

Figure 2. Dataset 1

In addition, to test some of the hypotheses that we have thought of, we collected data on the staff that was manning the cash register for each hour (cashier data). The cashier data was then prepared in the same format as above. Cashier data was available from 20th October to 31st December.

Date	Day	Start Time	Sales	CustNo	Staff 1	Staff 2	Staff 3
20-Oct-15	Tue	11:00	$293.60	21	0.5	0	0.5
20-Oct-15	Tue	12:00	$16.66	8	0	0	1
20-Oct-15	Tue	13:00	$413.61	54	0	0	1
20-Oct-15	Tue	14:00	$240.26	73	0	0	1
20-Oct-15	Tue	15:00	$387.02	62	0	0	1
20-Oct-15	Tue	16:00	$280.85	94	0	0	1
20-Oct-15	Tue	17:00	$166.36	93	0	0	1
20-Oct-15	Tue	18:00	$372.70	62	0	1	0
20-Oct-15	Tue	19:00	$398.74	34	0	1	0
20-Oct-15	Tue	20:00	$111.46	74	0	1	0

Figure 3. Dataset 2

Both datasets were then imported into JMP using the “Open…” function.

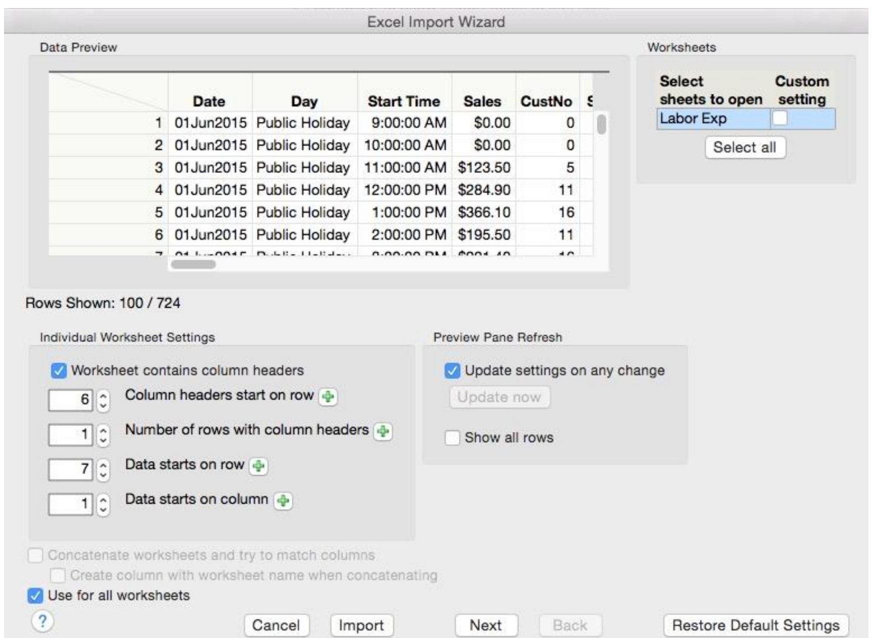


Figure 4. Importing Data

After toggling the worksheet setting such that the column headers are correctly reflected, we can proceed with importing the data using the "Import" function.


Labor Exp
Source
Columns (53/1)
Date
Day
Start Time
Sales
CustNo
Staff 1
Staff 2
Staff 3

	Date	Day	Start Time	Sales	CustNo	Staff 1	Staff 2
1	01Jun2015	Public Holiday	9:00:00 AM	$438.93	16	0	0
2	01Jun2015	Public Holiday	10:00:00 AM	$457.46	28	0	0
3	01Jun2015	Public Holiday	11:00:00 AM	$350.14	16	0	0
4	01Jun2015	Public Holiday	12:00:00 PM	$477.59	2	0	0
5	01Jun2015	Public Holiday	1:00:00 PM	$475.05	29	0	0
6	01Jun2015	Public Holiday	2:00:00 PM	$232.14	37	0	0
7	01Jun2015	Public Holiday	3:00:00 PM	$178.10	16	0	0
8	01Jun2015	Public Holiday	4:00:00 PM	$19.74	5	0	0
9	01Jun2015	Public Holiday	5:00:00 PM	$282.52	22	0	0

Figure 5. Imported Data

The following data table was then generated in JMP for further analysis.

3.1.1 "Time of the Day" Effect

We were provided with hourly and daily data for both sales and labour within a six-month period. In exploring the data given, we found that there was a bimodal pattern in the sales transactions.

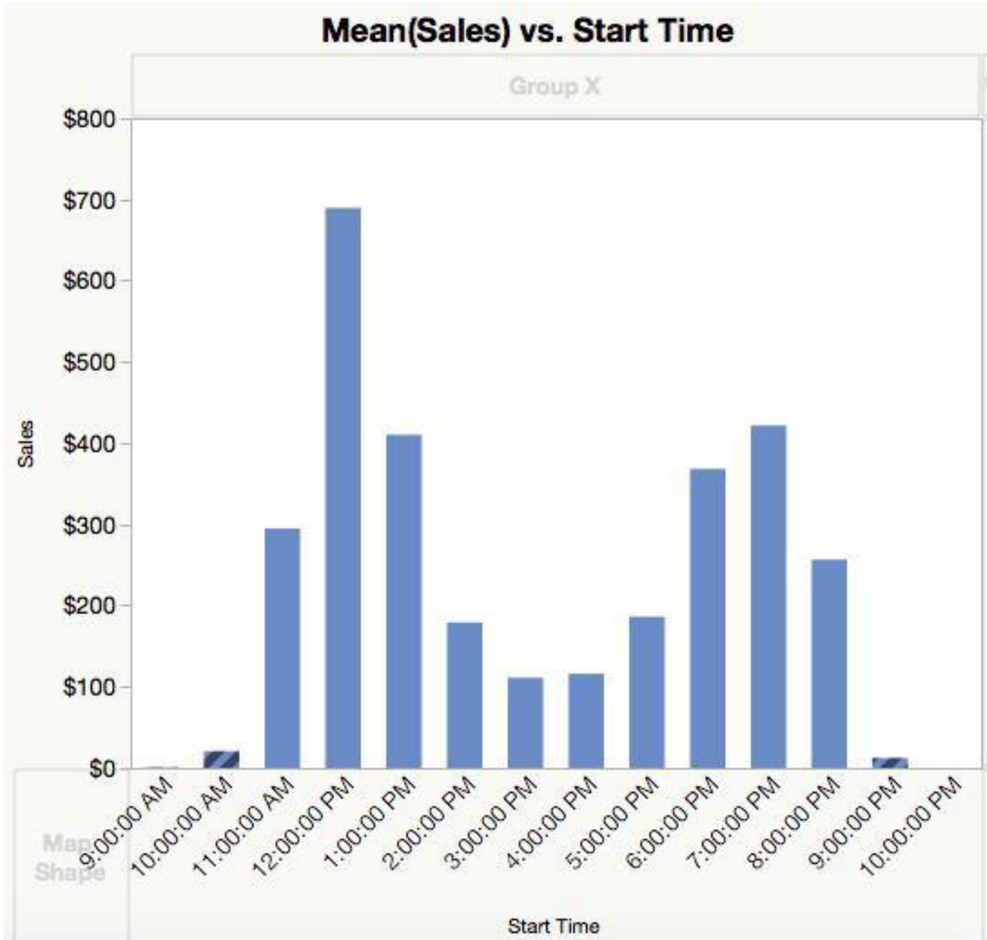


Figure 6. Mean Sales over Time in a Day

Mean sales per hour peaked at two times of the day namely, 11:00 to 13:00 and 18:00 to 20:00. We attribute the peak periods to the lunchtime and dinnertime crowds respectively. The other time periods are classified as idle time. We realise that this *"Time of the Day" effect* will have to be taken into account when evaluating shop productivity in our hypotheses so that sales toward a particular factor would not be over or under attributed. We label 11:00 to 13:00 as *Lunch Peak*, 14:00 to 17:00 as *Idle*, and 18:00 to 20:00 as *Dinner Peak*. As for time periods 9:00 to 10:00 and 21:00 to 22:00, we have decided to exclude them from our analysis due to their insignificant contribution to shop sales. Furthermore, with respect to Figure 2.1, the 75 percentile of sales for these timings are usually close to zero which implies that sales during the first two and last two hours of the day are usually zero.

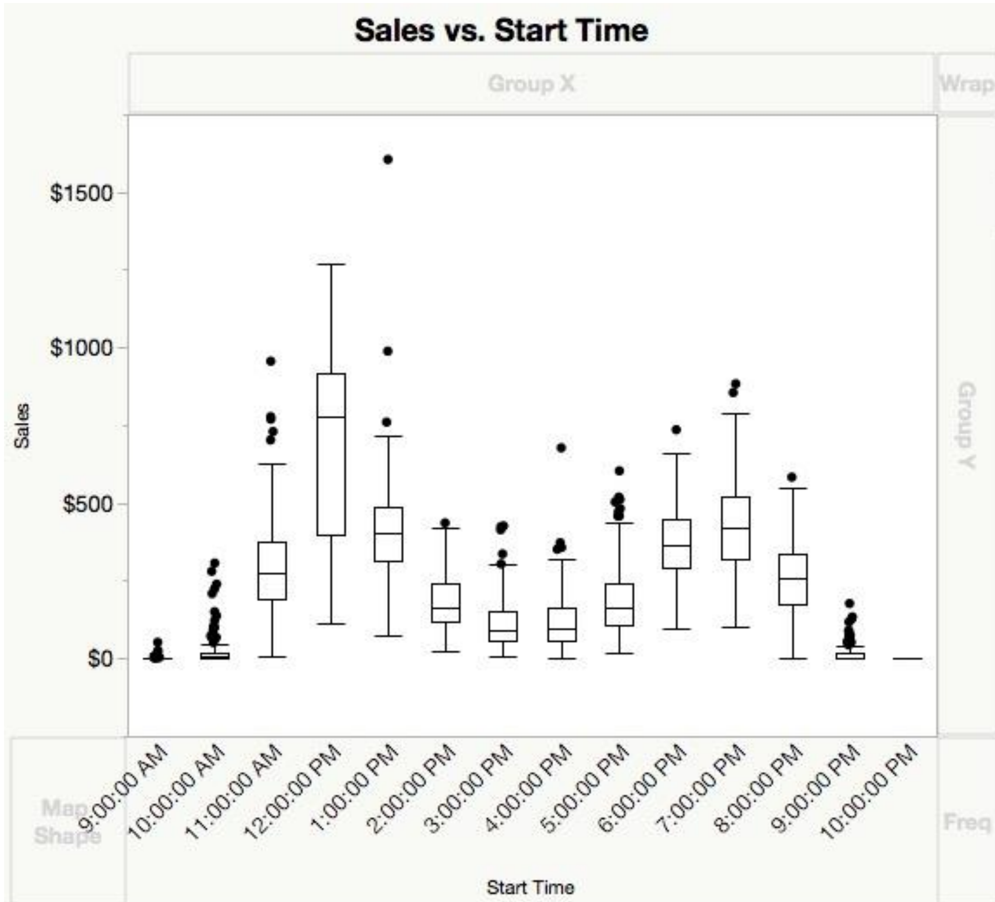


Figure 6.1. Box Plot of Sales over Time in a Day

3.1.2 "Day of the Week" Effect

We also found the mean sales for every day of the week. Accounting for the "Time of the Day" effect, we find the mean sales each day of the week. The day of the week also accounts for public holidays as crowds may be higher during public holidays. Similar to the "Time of the Day" effect, the chart below shows that there is also a *"Day of the Week" effect*.

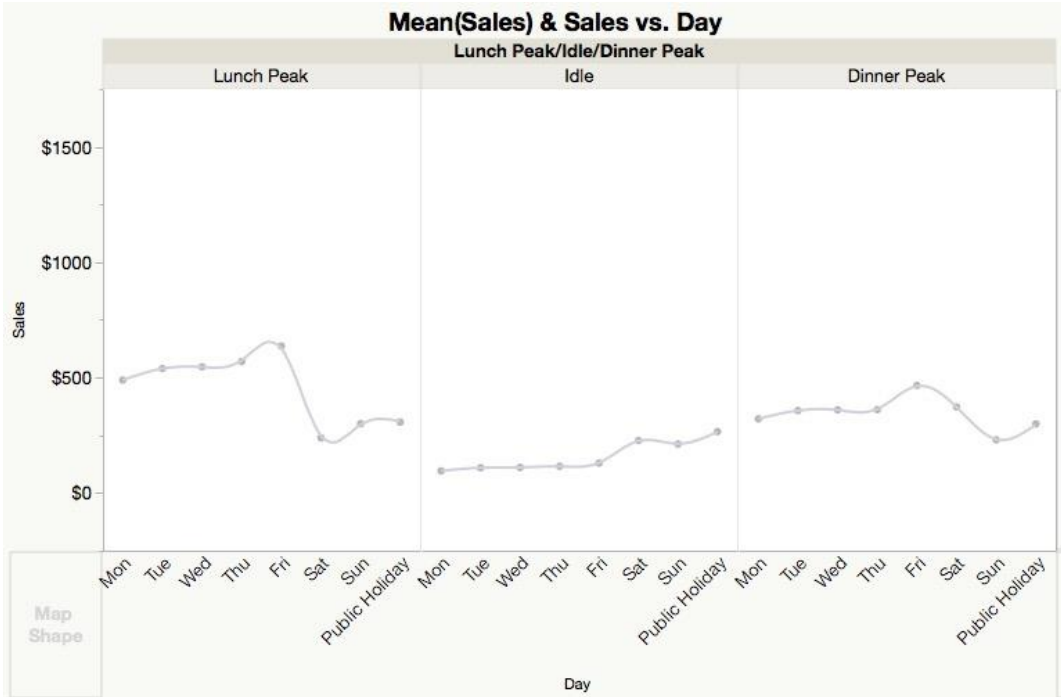


Figure 7. Mean Sales over Time over Day

During Lunch Peak and Dinner Peak, the store tends to achieve higher sales average on Fridays, with Saturdays experiencing lower mean sales. However during Idle periods, Saturdays and Public Holidays tend to experience relatively better mean sales compared to Idle periods on other days in the week.

3.2 Data Cleaning and Procedures

3.2.1 Removing Autocorrelation

The "Time of the Day" and the "Day of the Week" effects that we mentioned above indicated that the data might have significant autocorrelation due to its nature as a time-series. Significant autocorrelation would render the regression analysis inaccurate and hence we tested the dataset for autocorrelation using the Durbin-Watson test. We ran a sample regression analysis using the *Fit Model* function with the dependent variable Y as Sales/Customer Number and independent variable X as the Total number of Manager Labour Hours.

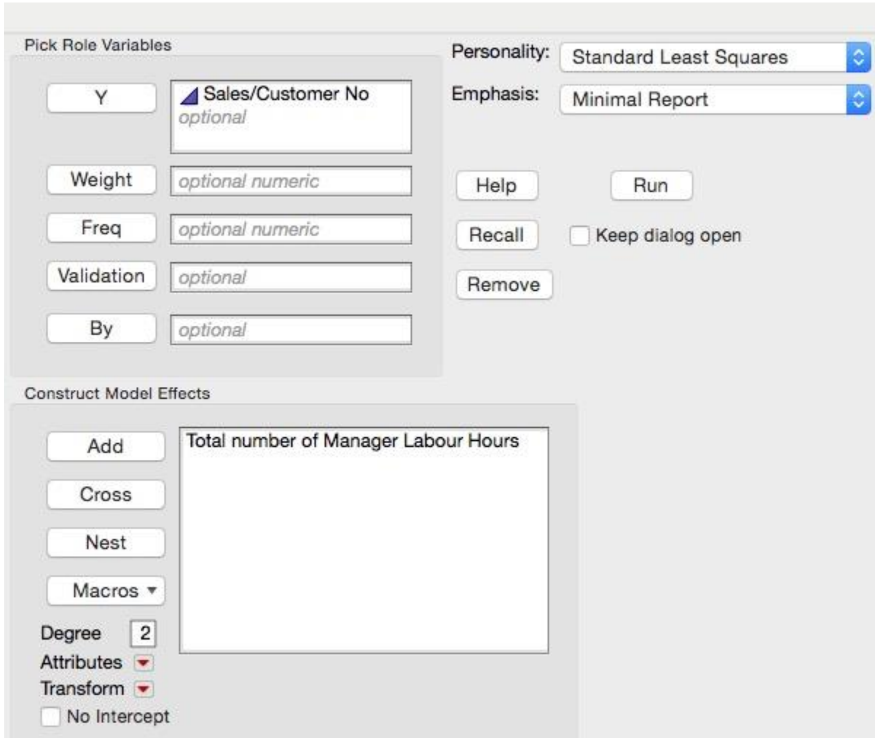


Figure 8. Fit Model (Durbin-Watson 1)

The results of the Durbin-Watson test is shown below:

Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.5102266	2124	0.2445	<.0001*

Figure 9. Durbin-Watson results 1

With p-value less than 0.05, we know that autocorrelation is present in our data. The Durbin-Watson value of 1.51 tells us that there is some positive correlation between values of the dependent variable, Sales/Customer Number, across different time periods.

In order to remove autocorrelation, we opted to partition the dataset into eight different days (including Public Holidays) and three different time periods in a day so as to account for both the "Day of the Week" and the "Time of the Day" trends. To do this in the Fit Model, we used the *By* function with the variables Day and Lunch Peak/Idle/Dinner Peak.

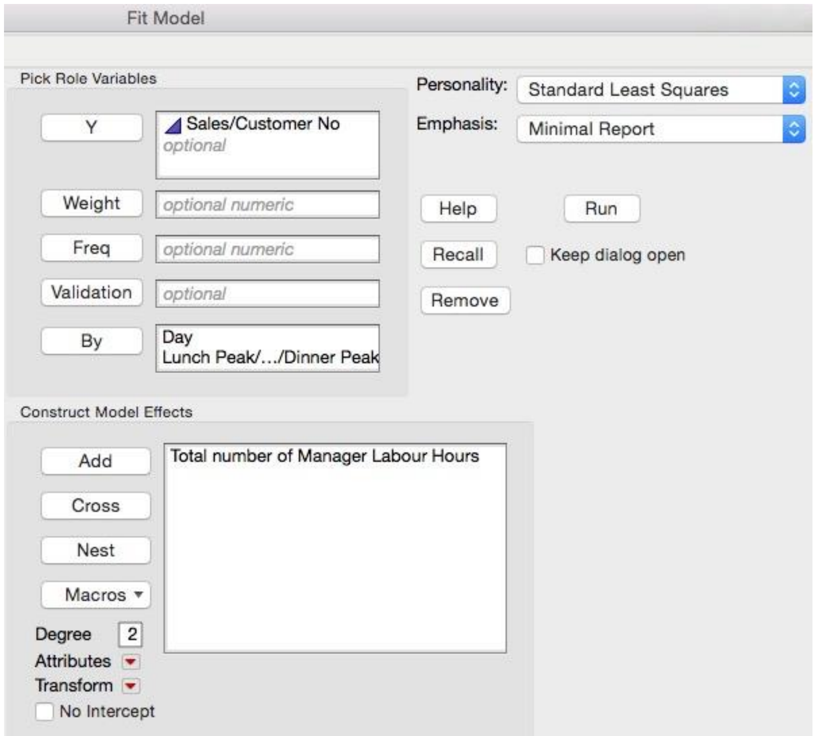


Figure 10. Fit Model (Durbin-Watson 2)

The results of the Durbin-Watson test this time is:

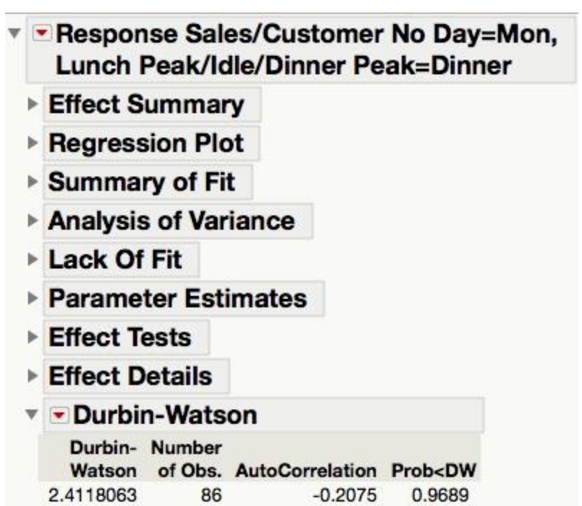
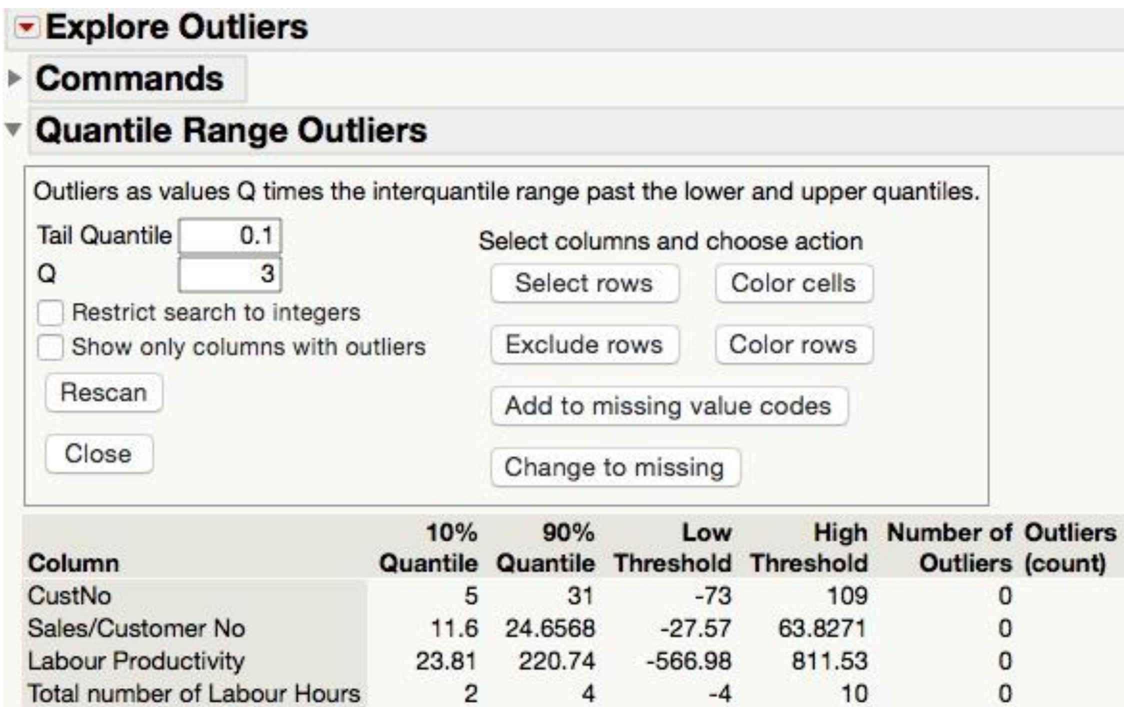


Figure 11. Durbin-Watson results 2

The p-value now is greater than 0.05, showing that the degree of autocorrelation in the partitioned data is no longer significant. Thus, for our subsequent analyses, we would run the analyses with the partitioned data instead of the full data set in order to account for the "Time of the Day" and "Day of the Week" effects.

3.2.2 Removing Outliers

Before regression analysis can be done, we must also ensure that there are no outliers that might skew the results of the analysis. We used the JMP *Remove Outlier* function to search for outliers.



Column	10% Quantile	90% Quantile	Low Threshold	High Threshold	Number of Outliers	Outliers (count)
CustNo	5	31	-73	109	0	
Sales/Customer No	11.6	24.6568	-27.57	63.8271	0	
Labour Productivity	23.81	220.74	-566.98	811.53	0	
Total number of Labour Hours	2	4	-4	10	0	

Figure 12. Exploring Outliers

The definition of an outlier in this case is any point that is 3 times the interquartile range of a certain variable. We did not identify any outliers for the three constituent elements of shop productivity.

3.2.3 Adding Calculated Columns

In order to test our hypotheses, we had to create new columns that were not included in the data set. All the column variables are indicated in the table below.

Table 1

Column Definitions

Column Variables	Existing Column/New Column	Definition
Date	Existing	Date of data point, from 1st of June 2015 to 31st of December 2015
Day	Existing	Day of data point, from Monday to Sunday, as well as Public Holidays
Start Time	Existing	Start time of data point, from 9:00 to 22:00
Sales	Existing	Sales earned during the hour
CustNo	Existing	Number of customers during the hour
Staff #	Existing	Staff # (# = 1 to 48), from 0 to 1. Indicates presence during the hour
Total number of Labour Hours	New	Total number of working hours during the hour, calculated by adding up the presence of Staff #
Sales/Customer No	New	Average sales per customer during the hour, calculated by dividing hourly sales by hourly customer number
Labour Productivity	New	Sales per labour hour, calculated by dividing hourly sales by total number of labour hours worked during the hour
Total number of Manager Labour hours	New	Sales per labour hour worked by managers, calculated by dividing hourly sales by total number of labour hours worked by managers during the hour
Total number of Full-time Labour hours	New	Sales per labour hour worked by full-time staff, calculated by dividing hourly sales by total number of labour hours worked by full-time staff during the hour
Total number of Part-time Labour hours	New	Sales per labour hour worked by part-time staff, calculated by dividing hourly sales by total number of labour hours worked by part-time staff during the hour
Lunch Peak/Idle/Dinner Peak	New	The hourly data point is classified as Lunch Peak when it lies between 11:00 to 13:00. It is classified as Idle when it lies between 14:00 to 17:00, and Dinner Peak when it lies between 18:00 to 20:00
CustNo/Total number of Labour Hour	New	The number of customers served in an hour per labour hour, calculated by dividing the number of customers served in the shop by the total number of labour hours worked by all staff during the hour

We added the new columns into the JMP table using the “Add new column” function. For example, we wanted to add a calculated column for “Total number of Labour Hours”

Figure 13. Adding New Columns

Step 1: Click Cols in the Menu Bar, then select New Column.

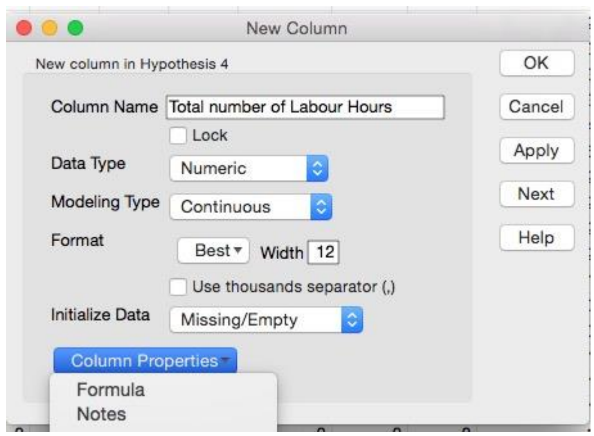


Figure 13.1. New Column Function

Step 2: Under Column Properties, select Formula.

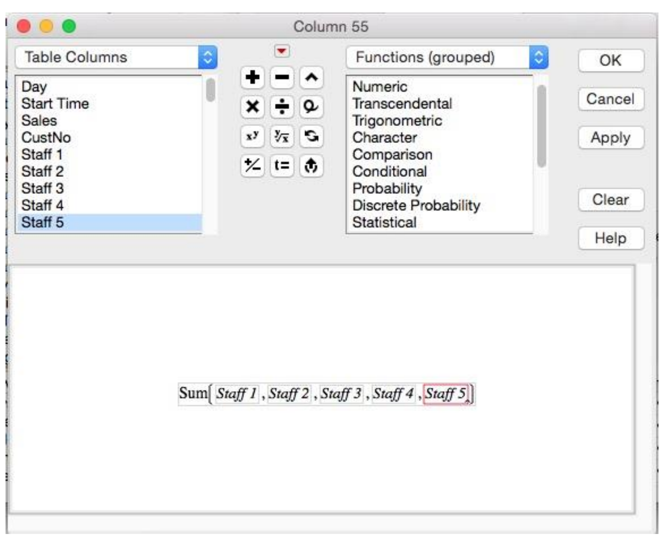


Figure 14. Formula Function

Step 3: Under Statistical, click on Sum, and then select Staff 1 to Staff 5

Step 4: Click Apply

We repeated this process for all the new columns to obtain respective calculated columns.

Hypothesis 4 2

Total number of Labour ...	Sales/ Customer No	Labour Productivity	Total number of Manager ...	Total number of Full-time ...	Total number of Part-...	Lunch Peak/ Idle/Dinner ...
0	•	•	0	0	0	Lunch Peak
1	•	0	0	0	1	Lunch Peak
2	24.7	61.75	0	0	2	Lunch Peak
3	25.9	94.966666667	0	0	3	Lunch Peak
3	22.88125	122.03333333	0	0	3	Lunch Peak
3	17.772727273	65.166666667	0	0	3	Idle
4	13.8375	55.35	0	0	4	Idle
3.5	19.05625	87.114285714	0	0	3.5	Idle
2	22.51	112.55	0	0	2	Idle
2	29.933333333	89.8	0	0	2	Dinner Peak
2	19.676923077	127.9	0	0	2	Dinner Peak
2	23.225	46.45	0	0	2	Dinner Peak
1	•	0	0	0	1	Dinner Peak
0	•	•	0	0	0	Dinner Peak

Figure 15. Calculated Columns

3.3 Analysis – Hypothesis 1

Hypothesis 1: We can increase shop productivity by hiring good cashiers who can upsell (increase sales dollar per customer) and serve customers faster (increase customer number).

From the sales process, we know that the cashiers have the most contact with customers. They are also the most likely to influence customers' purchase decisions, based on their ability to upsell and cross-sell. This hypothesis looks at identifying good cashiers who are able to consistently increase sales dollars per customer through upselling and/or cross-selling. It also looks at the speed at which the cashier serves the customers, as more customers served within an hour means higher sales, which directly relates to higher shop productivity.

For this analysis, we use simple linear regression and apply it to each cashier to see if a particular cashier is able to affect the number of customers, and sales dollar per customer. At the same time, we account for the "Time of the Day" and "Day of the Week" effects by partitioning the data as mentioned earlier.

3.3.1 Fit Model

To do so, we use JMP's Fit Model function. We first go to the Menu bar and click on *Analyze*, then on *Fit Model*.

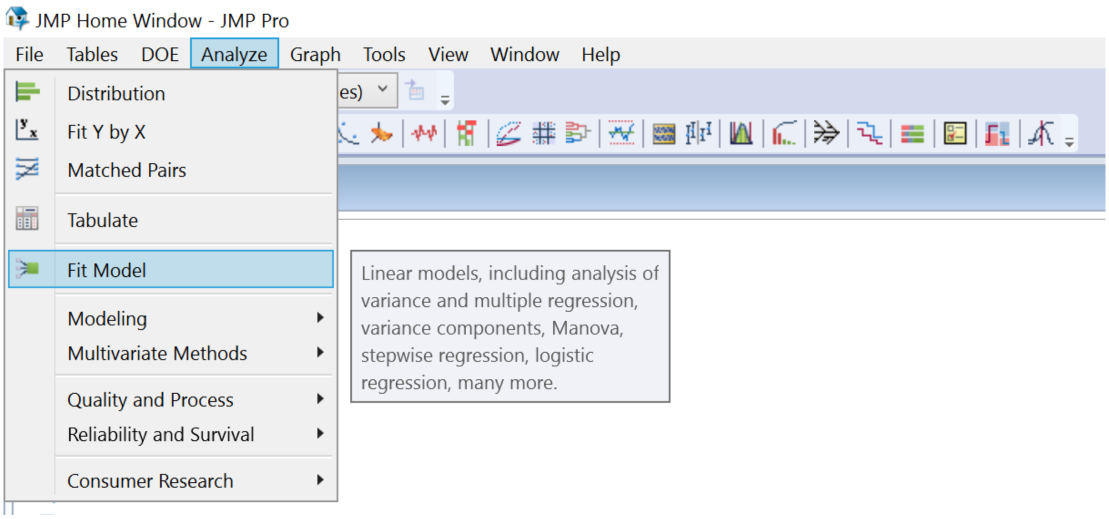


Figure 16. Fit Model Function

A pop up should appear, prompting for a JMP Data Table. Select the data table for running the regression model. The *Fit Model* Dialog will pop up.

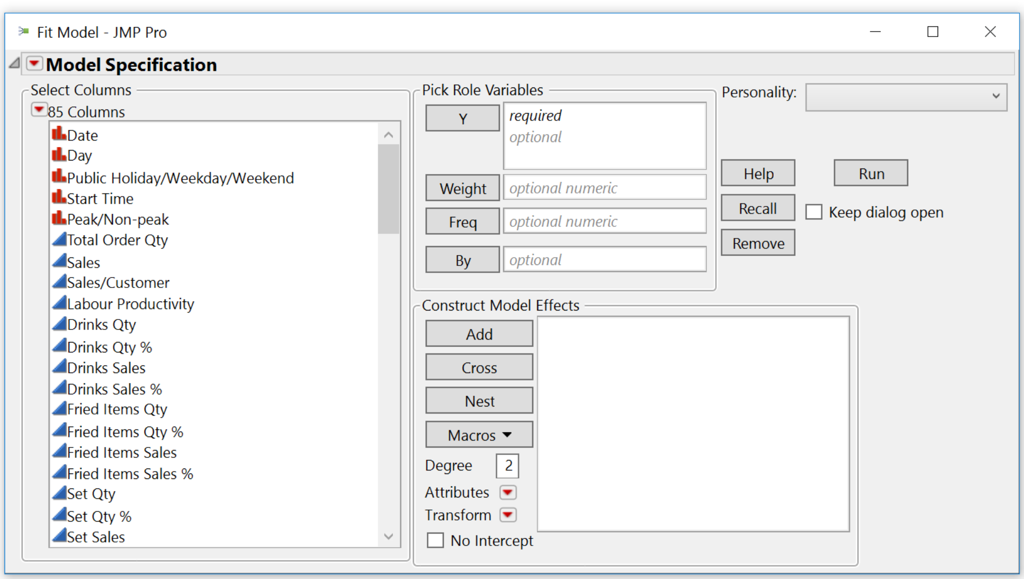


Figure 17. Fit Model Dialog

Next, select the Role Variables. *Y* refers to the dependent variable(s) that we want to analyse. For Hypothesis 1, *Y* would be *Sales/Customer* and *CustNo*. We use the *By* function to partition the data by *Day* and *Lunch Peak/Idle/Dinner Peak* to account for the "Time of the Day" and "Day of the Week" effects. Construct Model Effects is our X variables, the independent variables. We add a single cashier (eg. Part-time 1) to this box to identify the effect that Part-time 1 has when he assumes the role as a cashier.. Finally, we click on *Run* to run the analysis.

The resulting Fit Model Dialog before running the analysis should be like this:

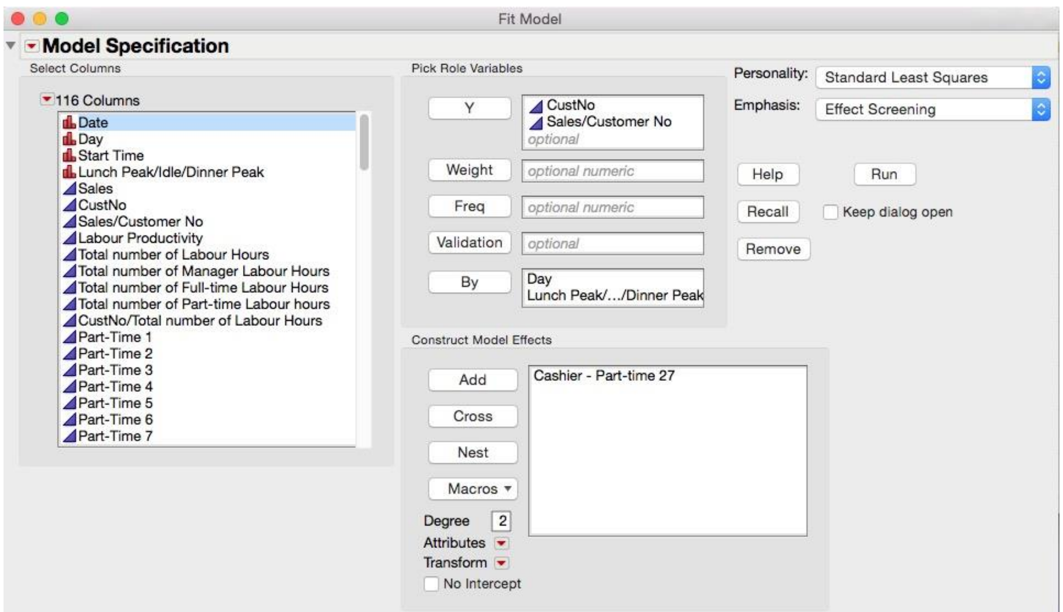


Figure 18. Fit Model Dialog (Hypothesis 1)

3.3.2 Analysis Results

Least Squares Fit Day=Sun, Lunch Peak/Idle/Dinner Peak=Dinner Peak

Response CustNo

Regression Plot

Summary of Fit

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	11.875	1.025076	11.58	<.0001*
Cashier - Part-time 27	-1.375	2.29214	-0.60	0.5534

Effect Tests

Effect Details

Response Sales/Customer No

Regression Plot

Summary of Fit

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	21.172385	1.022301	20.71	<.0001*
Cashier - Part-time 27	5.1584504	2.285934	2.26	0.0320*

Figure 19. Hypothesis 1 Results

The results show that during Dinner Peaks on Sunday, when Part-time 27 works as a cashier, he does not make a significant impact on the number of customers served in an hour as the model has a P-value that is greater than 0.05. However, Part-time 27 does make a significant impact on Sales/Customer No with P-value that is greater than 0.05. The results imply that Part-timer 27 has a significant positive impact on the Sales/Customer No when he works a cashier during dinner peaks on Sunday, bringing in $5.16 more per customer as opposed to when someone else is a cashier.

3.3.3 Implications

Teppei Syokudo should take further steps to design a controlled experiment with good and average performing cashiers to validate the results of this model. If it is proven that certain good performing cashiers consistently bring in more customers and/or increase the shop dollars per customer, Teppei Syokudo can take steps to reward and retain them.

3.4 Analysis – Hypothesis 2

Hypothesis 2: The presence of managers can positively impact shop productivity by increasing the staff's ability to upsell and serve more customers.

We have identified three shop managers present in Teppei Syokudo and we want to test whether the shop would perform better when these managers are present as the staff are more motivated to upsell and serve customers faster.

3.4.1 Fit Model

We use the Fit Model with the dependent variables *Y* as Sales/Customer and CustNo, as well as independent variables Manager 1, Manager 2 and Manager 3. Manager-variables have a value from 0 to 1, indicating their presence for every hour. Since this hypothesis involves a multi-linear regression with three independent variables, we included two-way interactions between the managers to account for the scenarios where there are two managers in the shop. As there is no scenario where there are three managers in the shop, we excluded three-way interactions.

The resulting Fit Model Dialog should look like this:

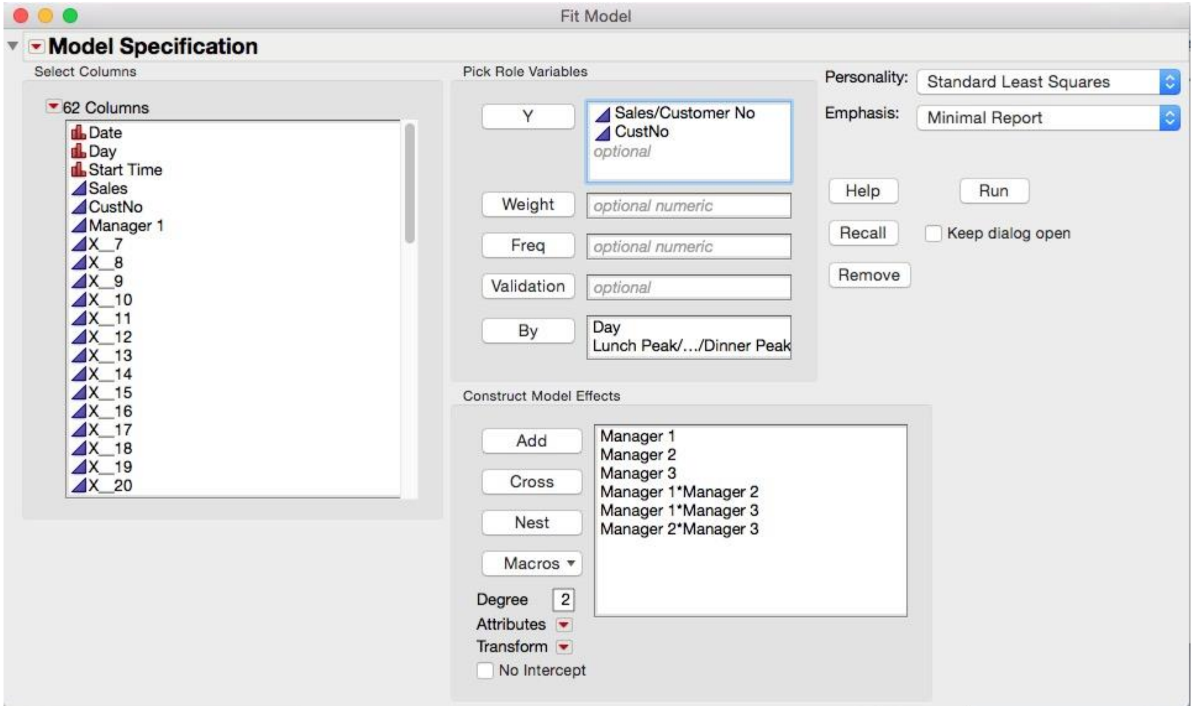
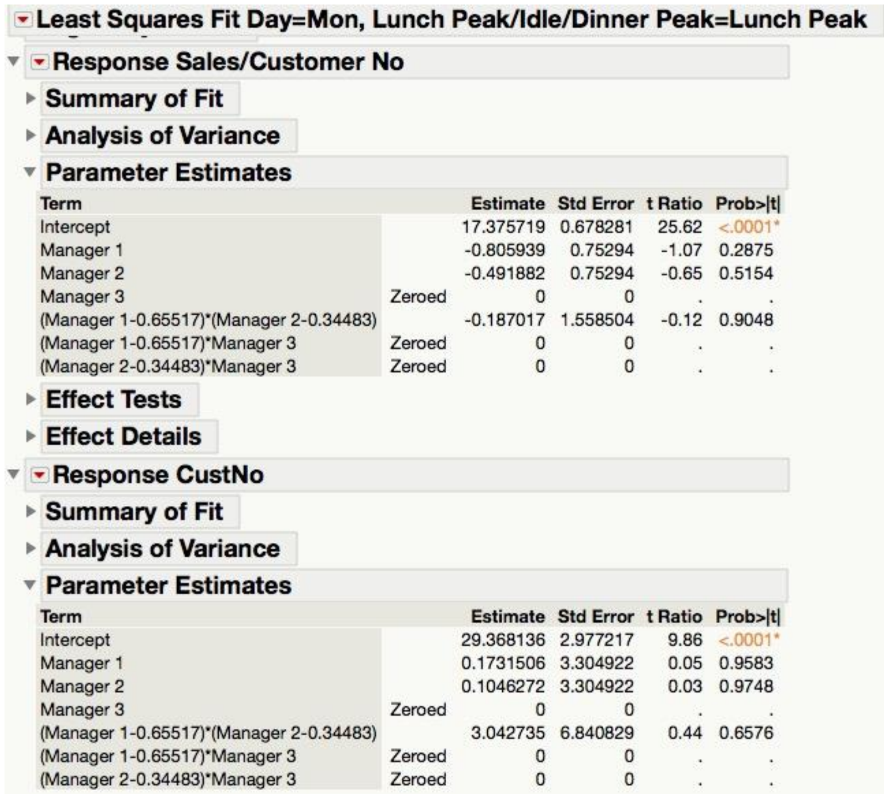


Figure 20. Fit Model Dialog (Hypothesis 2)

We ran the Fit Model with the independent variables so as to identify the effect that each manager has on the dependent variables.

3.4.2 Analysis Results



Least Squares Fit Day=Mon, Lunch Peak/Idle/Dinner Peak=Lunch Peak

Response Sales/Customer No

Summary of Fit

Analysis of Variance

Parameter Estimates

Term		Estimate	Std Error	t Ratio	Prob>|t|
Intercept		17.375719	0.678281	25.62	<.0001*
Manager 1		-0.805939	0.75294	-1.07	0.2875
Manager 2		-0.491882	0.75294	-0.65	0.5154
Manager 3	Zeroed	0	0	.	.
(Manager 1-0.65517)*(Manager 2-0.34483)		-0.187017	1.558504	-0.12	0.9048
(Manager 1-0.65517)*Manager 3	Zeroed	0	0	.	.
(Manager 2-0.34483)*Manager 3	Zeroed	0	0	.	.

Effect Tests

Effect Details

Response CustNo

Summary of Fit

Analysis of Variance

Parameter Estimates

Term		Estimate	Std Error	t Ratio	Prob>|t|
Intercept		29.368136	2.977217	9.86	<.0001*
Manager 1		0.1731506	3.304922	0.05	0.9583
Manager 2		0.1046272	3.304922	0.03	0.9748
Manager 3	Zeroed	0	0	.	.
(Manager 1-0.65517)*(Manager 2-0.34483)		3.042735	6.840829	0.44	0.6576
(Manager 1-0.65517)*Manager 3	Zeroed	0	0	.	.
(Manager 2-0.34483)*Manager 3	Zeroed	0	0	.	.

Figure 21. Hypothesis 2 Results 1

The results show that on Mondays during Lunch Peak hours, all three managers do not have significant impacts on average sales per customer (Response = Sales/Customer No) and customer number (Response = CustNo), as the Prob>|t| values are greater than 0.05.

Because there are multiple independent variables that might be correlated with each other, a multi-collinearity test should be done. The *VIF* function in JMP tests for multi-collinearity in a model. VIF values above 10 signal high multi-collinearity.

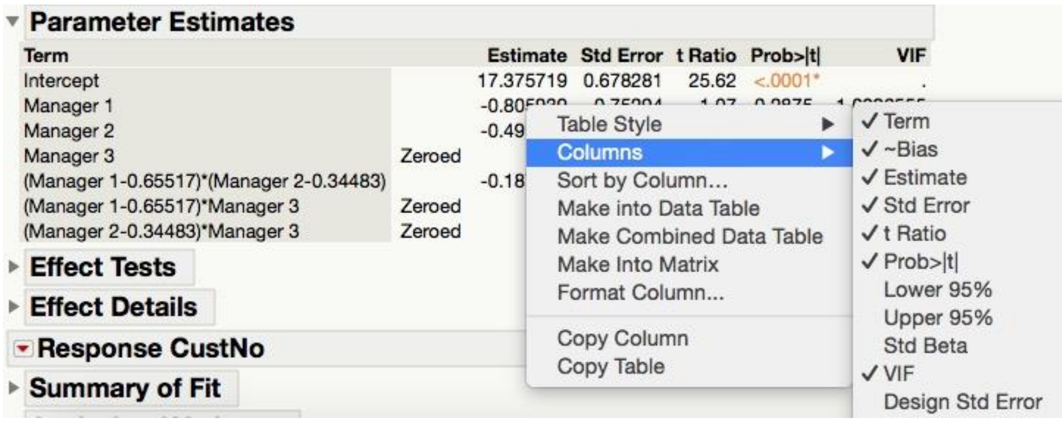


Figure 22. Show VIF

Right click on the Parameter Estimates of the Fit Model and go to Columns and then Click on VIF.

Parameter Estimates

Term		Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept		17.375719	0.678281	25.62	<.0001*	.
Manager 1		-0.805939	0.75294	-1.07	0.2875	1.0096555
Manager 2		-0.491882	0.75294	-0.65	0.5154	1.0096555
Manager 3	Zeroed	0	0	.	.	0
(Manager 1-0.65517)*(Manager 2-0.34483)		-0.187017	1.558504	-0.12	0.9048	1.0054501
(Manager 1-0.65517)*Manager 3	Zeroed	0	0	.	.	0
(Manager 2-0.34483)*Manager 3	Zeroed	0	0	.	.	0

Figure 23. Hypothesis 2 Results 2

A VIF column should appear beside the Prob>|t| column. It seems that there is low multi-collinearity for this particular model because the VIF values are small.

3.4.3 Implications

Managers have no significant impact on the staff's ability to upsell or to serve more customers on Monday lunch peaks. Teppei Syokudo can explore the feasibility of having managers to check into the store periodically since their continued presence do not motivate the staff to work harder.

3.5 Analysis – Hypothesis 3

Hypothesis 3: We can increase shop productivity by increasing the number of full-time staff present, as they are better at upselling and they serve customers faster as compared to part-time staff.

3.5.1 Fit Model

We use the Fit Model with the Sales/Customer and CustNo as dependent variables *Y*. We used the Total number of Full-time Labour Hours and Total number of Part-time Labour Hours as independent variables X. Since this hypothesis involves a multi-linear regression with two independent variables, we included two-way interactions between both independent variables to account for the scenarios where there are both full time staff and part-time staff working in the shop.

The resulting Fit Model Dialog should look like this:

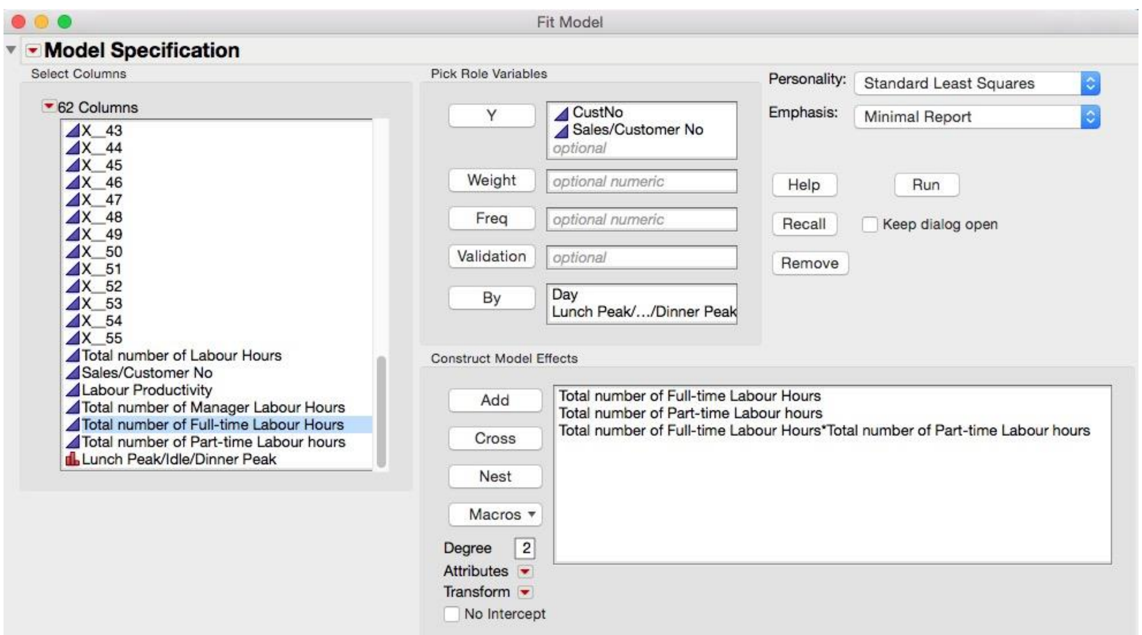


Figure 24. Fit Model Dialog (Hypothesis 3)

We ran the Fit Model with the independent variables so as to identify the effect that the number of full-time labour hours, as well as the number of part-time labour hours have on shop performance.

3.5.2 Analysis Results

Least Squares Fit Day=Mon, Lunch Peak/Idle/Dinner Peak=Lunch Peak

Effect Summary

Response CustNo

Summary of Fit

Analysis of Variance

Lack Of Fit

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	14.145881	4.971709	2.85	0.0056*	.
Total number of Full-time Labour Hours	7.0989729	2.721254	2.61	0.0108*	1.2783346
Total number of Part-time Labour hours	6.0023591	2.445647	2.45	0.0162*	1.6239117
(Total number of Full-time Labour Hours-0.77299)*(Total number of Part-time Labour hours-1.51149)	-3.565185	3.758045	-0.95	0.3455	1.3152557

Effect Tests

Effect Details

Response Sales/Customer No

Summary of Fit

Analysis of Variance

Lack Of Fit

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	15.184763	1.214939	12.50	<.0001*	.
Total number of Full-time Labour Hours	0.570042	0.664994	0.86	0.3938	1.2783346
Total number of Part-time Labour hours	0.8095288	0.597644	1.35	0.1792	1.6239117
(Total number of Full-time Labour Hours-0.77299)*(Total number of Part-time Labour hours-1.51149)	0.7914268	0.918356	0.86	0.3913	1.3152557

Figure 25. Hypothesis 3 Results

The results show that on Mondays during Lunch Peak hours, seven more customers usually accompany an hourly increase in the total number of full-time labour hours whereas six more customers usually accompany an hourly increase in the total number of part-time labour hours. The results are significant as the Prob>|t| values are smaller than 0.05.

In the same period of time, both the number of full-time and part-time labour hours do not have significant impacts on average sales per customer (Response = Sales/Customer No), as the Prob>|t| values are larger than 0.05.

It seems that there is low multi-collinearity for this particular model because the VIF values are small too.

3.5.3 Implications

Teppei Syokudo should take further steps to design a controlled experiment with the number of full-time labour and part-time labour as independent variables, to validate the results of this model. If it is proven that increasing the number of full-time labour allows the shop to serve more customers relative to increasing the number of part-time labour, then Teppei Syokudo can evaluate whether the extra revenue generated justifies the additional cost needed to hire more full-time labour.

3.6 Analysis – Hypothesis 4

Hypothesis 4: We can increase shop productivity by staffing full-timers or managers as cashiers as they are better at upselling or serving more customers than part-timers.

This hypothesis is similar to Hypothesis 1 in the sense that it focuses on identifying good cashiers that upsell more or serves more customers, ultimately increasing shop productivity. However, hypothesis 4 seeks to identify whether the type of cashier affects his ability to contribute to shop productivity. The basis of this hypothesis is that full-timers and managers should be more motivated and committed to the shop's performance and hence, when they assume the customer-facing role of a cashier, they will be more inclined to upsell products or serve customers faster relative to part-timer cashiers.

3.6.1 Fit Model

We use the Fit Model with the dependent variables *Y* as Sales/Customer and CustNo, and independent variables Manager, Full-time and Part-time. We ran the Fit Model with each individual independent variable so as to identify the effect that each type of staff has on the dependent variables when they assume the role of a cashier.

The resulting Fit Model Dialog should look like this:

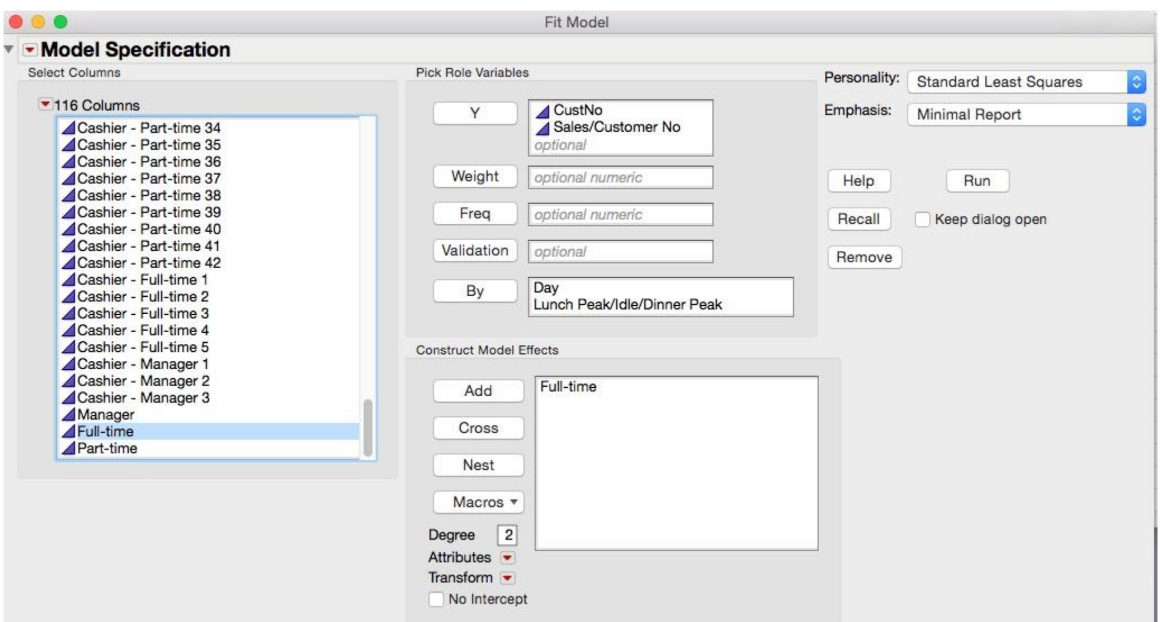
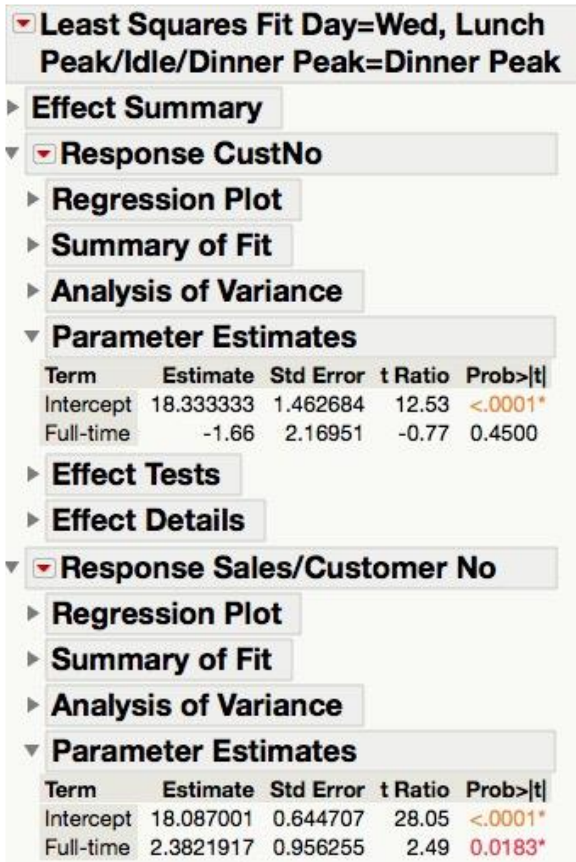


Figure 26. Fit Model Dialog (Hypothesis 4)

3.6.2 Analysis Results


Least Squares Fit Day=Wed, Lunch Peak/Idle/Dinner Peak=Dinner Peak

Effect Summary

Response CustNo

Regression Plot

Summary of Fit

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	18.333333	1.462684	12.53	<.0001*
Full-time	-1.66	2.16951	-0.77	0.4500

Effect Tests

Effect Details

Response Sales/Customer No

Regression Plot

Summary of Fit

Analysis of Variance

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	18.087001	0.644707	28.05	<.0001*
Full-time	2.3821917	0.956255	2.49	0.0183*

Figure 27. Hypothesis 4 Results

The results show that on Wednesdays during Dinner Peak hours, Full-time staff make a significant impact on Sales/Customer as P-value is greater than 0.05. However, Full-time staff do not make a significant impact on CustNo, with a P-value that is greater than 0.05. The results show that during Dinner Peak periods on Wednesdays, a full-time cashier will earn $2.38 per customer on average than if the cashier is not a full-time staff.

3.6.3 Implications

Teppei Syokudo should take further steps to design a controlled experiment with full-time cashiers and part-time cashiers to validate the results of this model. If it is proven that full-time cashiers earn more customer dollars than part- time cashiers, then Teppei Syokudo can consider assigning full-time staff as cashiers.

3.7 Analysis – Hypothesis 5

Hypothesis 5: We can increase shop productivity by decreasing the number of staff hours on time periods where there is excess capacity.

If the hypothesis is true and there were time periods where there is excess capacity, we would be able to identify time periods where the customers served, a proxy of how busy the shop is, per labour hour is significantly lower than other time periods. If the hypothesis is false, the shop should be at its optimal level of customers served per labour hour and there should be no significant difference between time periods.

As the independent X variable in this case is *Day,* which is a nominal variable, and dependent variable Y is the Number of customers served per labour hour (CustNo/Total number of Labour Hours), which is a continuous variable, we will be using the one-way ANOVA analysis to test whether there is a significant difference among the average *CustNo/Total number of Labour Hours* for each day of the week. We will repeat the analysis for the three time periods in a day to account for the time of the day effect.

3.7.1 Fit Y by X

To do so, we used JMP's Fit Y by X function. We first go to the Menu bar and click on *Analyze*, then on *Fit Y by X*.

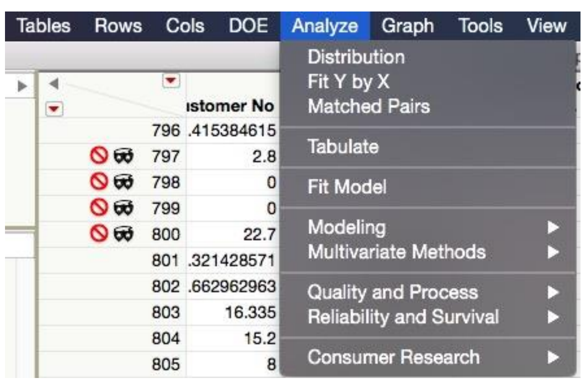


Figure 28. Fit Y by X Function

We then set *CustNo/Total number of Labour Hours* as the dependent variable *Y* and *Day* as the independent variable X.

The resulting Fit Y by X Dialog should look like this:

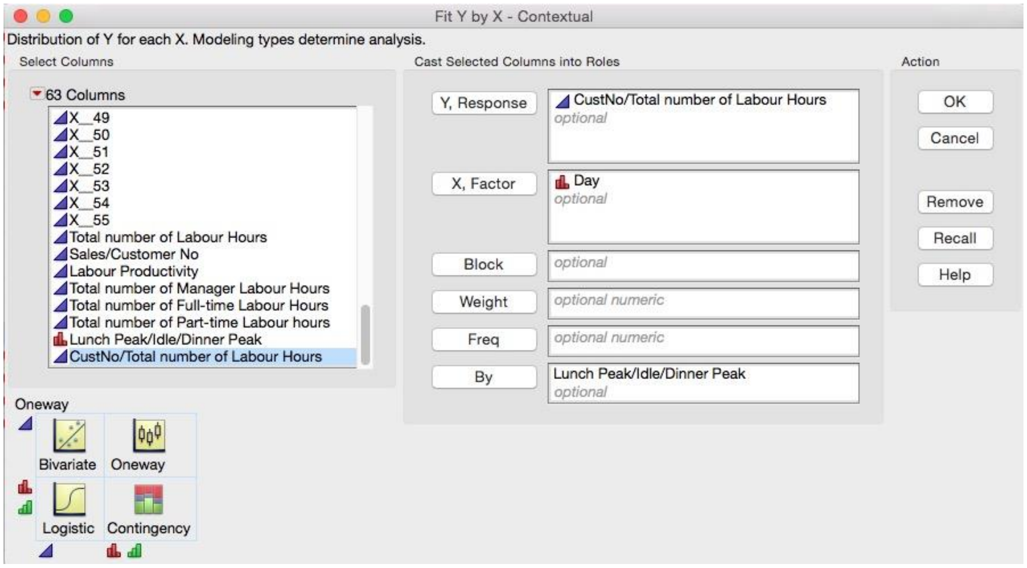


Figure 29. Fit Y by X Dialog

3.7.2 Analysis Results

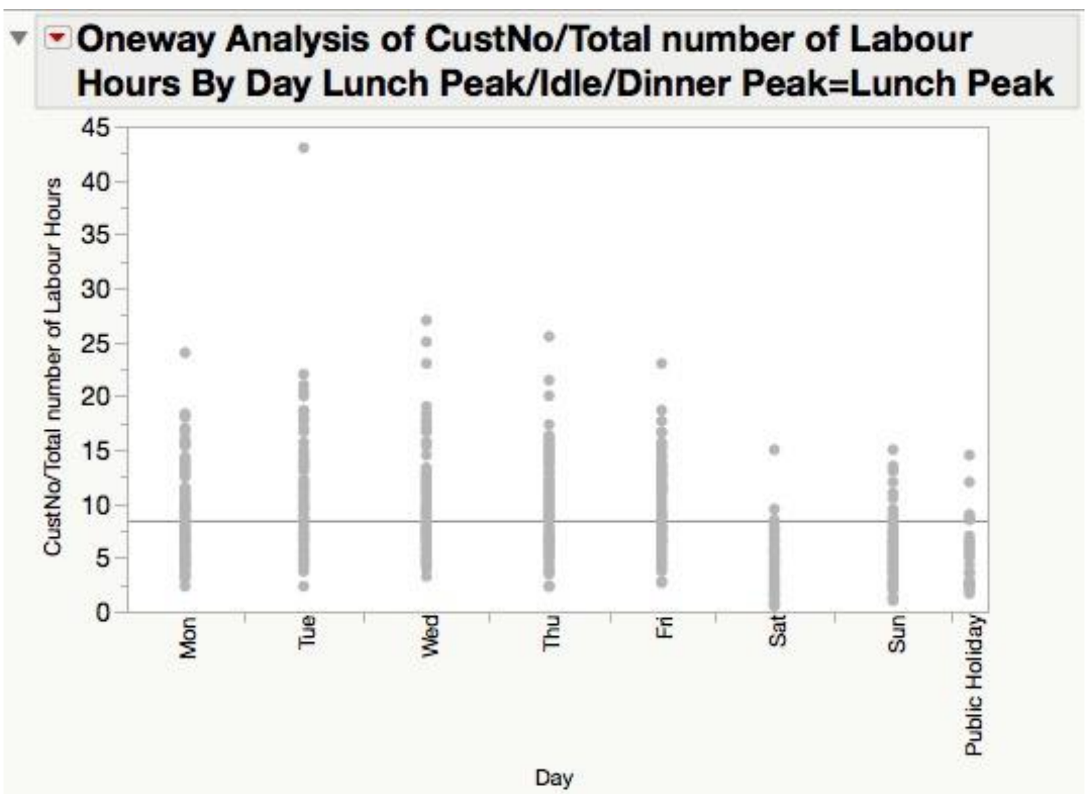


Figure 30. One-way Analysis of CustNo/Total number of Labour Hours by Day

The results show the different *CustNo/Total number of Labour Hours* for each *Day* and the line in the middle of the data points equals the mean of all the data points. To conduct the ANOVA analysis, click on the red triangle and the *Means/Anova* function.

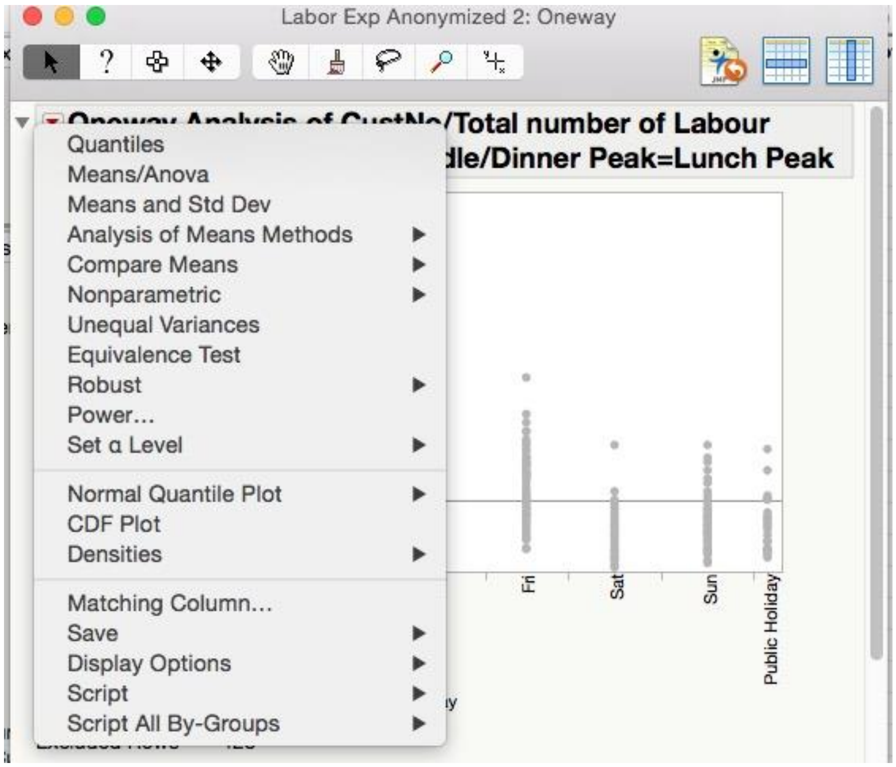
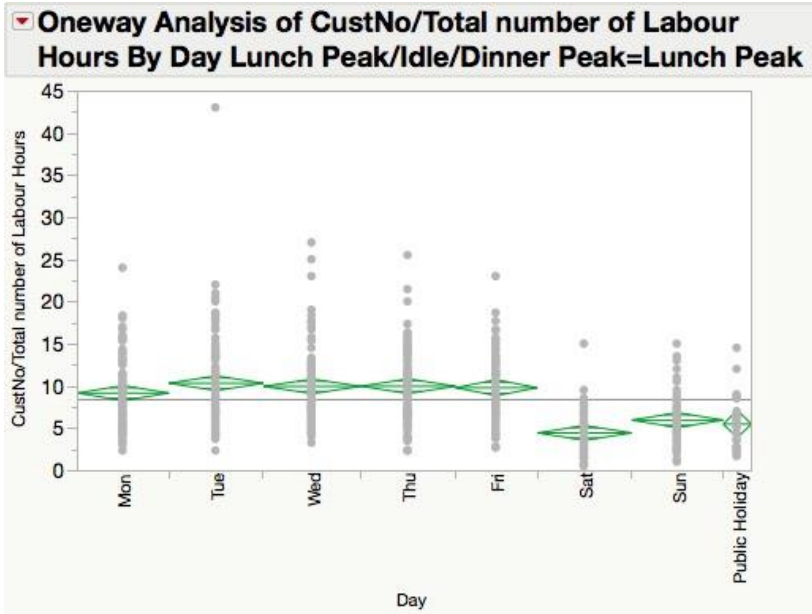


Figure 31. Conducting One-way ANOVA analysis



Excluded Rows 428

Oneway Anova

Summary of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Day	7	3148.372	449.767	25.2479	<.0001*
Error	634	11294.101	17.814		
C. Total	641	14442.473			

Figure 32. One-way ANOVA analysis results *(Lunch Peak)*

After clicking on the Means/Anova function, the one-way ANOVA analysis will be conducted. Mean diamonds representing confidence intervals will appear, with the line near the center of the diamond representing each group's mean and the vertical span of the diamond representing the 95% confidence interval for the mean of each group.

The results show that during Lunch Peak hours, the average CustNo/Total number of Labour Hours significantly differs across the days. Saturdays, Sundays and Public Holidays have significantly lower average CustNo/Total number of Labour Hours than the rest of the week as Prob > F value is less than 0.05.

Oneway Analysis of CustNo/Total number of Labour Hours By Day Lunch Peak/Idle/Dinner Peak=Lunch Peak

Oneway Anova

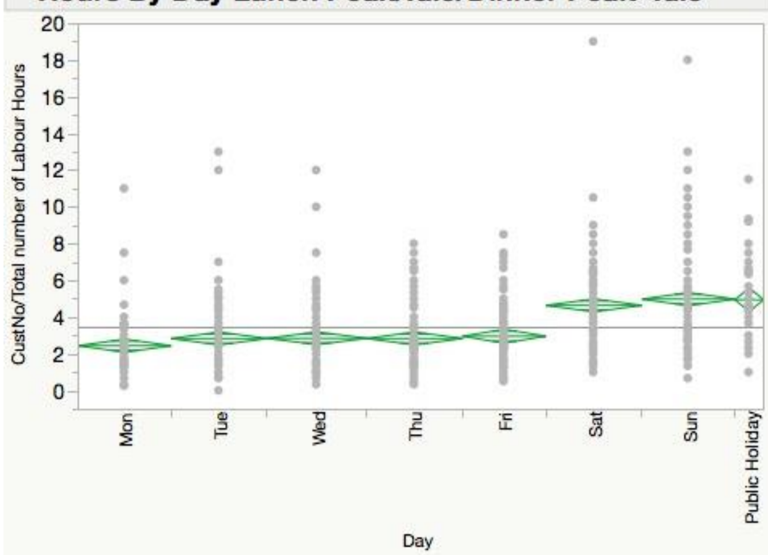
Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Mon	87	9.2281	0.45250	8.3395	10.117
Tue	90	10.3948	0.44490	9.5211	11.268
Wed	93	10.0207	0.43766	9.1612	10.880
Thu	90	10.0400	0.44490	9.1664	10.914
Fri	78	9.8633	0.47790	8.9249	10.802
Sat	90	4.5018	0.44490	3.6281	5.375
Sun	87	6.0212	0.45250	5.1327	6.910
Public Holiday	27	5.5543	0.81227	3.9593	7.149

Std Error uses a pooled estimate of error variance

Figure 33. One-way ANOVA average *CustNo/Total number of Labour Hours for each Day (Lunch Peak)*

Oneway Analysis of CustNo/Total number of Labour Hours By Day Lunch Peak/Idle/Dinner Peak=Idle



Oneway Anova

Summary of Fit

Rsquare	0.196985
Adj Rsquare	0.190357
Root Mean Square Error	1.960875
Mean of Response	3.464825
Observations (or Sum Wgts)	856

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Day	7	799.8458	114.264	29.7172	<.0001*
Error	848	3260.5869	3.845		
C. Total	855	4060.4328			

Figure 34. One-way ANOVA analysis results *(Idle)*

However, the results for the Idle time period shows that the average CustNo/Total number of Labour Hours are significantly higher during the weekends and Public Holidays than the rest of the days as Prob > F value is less than 0.05.

Oneway Analysis of CustNo/Total number of Labour Hours By Day Lunch Peak/Idle/Dinner Peak=Idle

Oneway Anova

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Mon	116	2.47654	0.18206	2.1192	2.8339
Tue	120	2.86881	0.17900	2.5175	3.2202
Wed	124	2.87889	0.17609	2.5333	3.2245
Thu	120	2.87050	0.17900	2.5192	3.2218
Fri	104	2.99396	0.19228	2.6166	3.3714
Sat	120	4.67196	0.17900	4.3206	5.0233
Sun	116	5.01408	0.18206	4.6567	5.3714
Public Holiday	36	4.97976	0.32681	4.3383	5.6212

Std Error uses a pooled estimate of error variance

Figure 35. One-way ANOVA average *CustNo/Total number of Labour Hours* for each *Day (Idle)*

3.8.3 Implications

Teppei Syokudo should explore the feasibility of decreasing the number of labour hours during lunch peak on weekends, as well as during Idle periods on weekdays as there is spare capacity during those periods.

5.0 Conclusion

Our use of regression analysis to test the effects of five different independent variables on the three constituents of store productivity yielded significant implications. We recommend that Teppei Syokudo take these three actionable steps.

1. Set up controlled experiments for hypotheses 1,3 and 4 to validate the results of the analysis.
2. Explore the feasibility of having managers to check into the store periodically since their continued presence do not significantly impact store productivity.
3. Explore the feasibility of decreasing the number of labour hours on weekend lunch peaks and weekday idle periods.

Table 2

Implications of hypotheses

No.	Hypotheses	Results	Implications
1	We can increase shop productivity by hiring good cashiers who can upsell (increase sales dollar per customer) and serve customers faster (increase customer number).	During Dinner Peaks on Sunday, when Part-time 27 works as a cashier, he does not make a significant impact on the number of customers served in an hour, but has a significant impact on Sales/Customer No The results imply that Part-timer 27 has a significant positive impact on the Sales/Customer No when he works a cashier during dinner peaks on Sunday, bringing in $5.16 more per customer as opposed to when someone else is a cashier.	Teppei Syokudo should take further steps to design a controlled experiment with good and average performing cashiers to validate the results of this model. If it is proven that certain good performing cashiers consistently bring in more customers and/or increase the shop dollars per customer, Teppei Syokudo can take steps to reward and retain them.
2	We can increase shop productivity by motivating staff to work harder with the presence of managers.	The results show that on Mondays during Lunch Peak hours, all three managers do not have significant impacts on average sales per customer	Managers have no significant impact on the staff's ability to upsell or to serve more customers on Monday lunch peaks. Teppei Syokudo can explore the feasibility of having managers to check into the store periodically since their continued presence do not motivate the staff to work harder
3	We can increase shop productivity by increasing the number of full-time staff or managers present as they are more productive compared to part-time staff.	During Lunch Peak hours on Mondays, seven more customers usually accompany an hourly increase in the total number of full-time staff hours whereas six more customers usually accompany an hourly increase in the	Teppei Syokudo should take further steps to design a controlled experiment with the number of full-time staff and part-time staff as independent variables, to validate the results of this model. If it is proven that increasing the number of full-time staff allows the shop to

		total number of part-time staff hours. In the same period of time, both the number of full-time and part-time staff hours do not have significant impacts on average sales per customer	serve more customers relative to increasing the number of part-time staff, then Teppei Syokudo can evaluate whether the extra revenue generated justifies the additional cost needed to hire more full-time labour.
4	We can increase shop productivity by staffing full-timers or managers as cashiers as they are better at upselling or serving more customers than part-timers.	During Dinner Peak hours on Wednesdays, assigning a Full-time staff as the cashier is usually accompanied by an increase in average sales dollar per customer by \$2.38 as compared to if other types of staff were assigned as the cashier.	Teppei Syokudo should take further steps to design a controlled experiment with full-time cashiers and part-time cashiers to validate the results of this model. If it is proven that full-time cashiers earn more customer dollars than part-time cashiers, then Teppei Syokudo can consider assigning full-time staff as cashiers.
5	We can increase shop productivity by decreasing the number of staff on time periods where there is excess capacity.	During Lunch Peak hours, the average CustNo/Total number of Labour Hours are significantly higher on weekends and Public Holidays than the rest of the week During Idle hours, the average CustNo/Total number of Labour Hours are significantly higher during the weekends and Public Holidays than the rest of the days as Prob > F value is less than 0.05.	Teppei Syokudo should explore the feasibility of decreasing the number of labour hours during lunch peak on weekends, as well as during Idle periods on weekdays as there is spare capacity during those periods.

6.0 Acknowledgements

We would like to thank Professor Kam Tin Seong for useful discussions and wise guidance during the formative stages of the paper.

References

Fukuda, Yoshiharu, Nakamura, Keiko, & Takano, Takehito. (2002). A combination of an extrapolation method and a benchmark method to develop quantitative health targets for Japan. *Health Policy, 61*(2), 201-212.

Farrar, D., & Glauber, R. (1967). Multicollinearity in Regression Analysis: The Problem Revisited. *The Review of Economics and Statistics*, *49*(1), 92-107.

Higgins, Jim. (2005). *The Radical Statistician.* Retrieved from http://www.biddle.com/documents/bcg_comp_chapter4.pdf

Lisa Getoor (2007). Introduction to Statistical Relational Learning

Sykes, Alan O. (n.d.). An Introduction to Regression Analysis. http://www.law.uchicago.edu/files/files/20.Sykes_.Regression.pdf

Walid Zakhem, Marie Khair, & Chady El Moucary. (2011). Improving Students Performance Using Data Clustering and Neural Networks in Foreign-Language Based Higher Education. *International Journal of ACM Jordan, 2*(1), International Journal of ACM Jordan, 01 September 2011, Vol.2(1).