# A Decision Support Model for Staffing Supply Chain Planners: A Case from the Consumer Packaged Goods Industry



by

Aura C. Castillo
B.S. Industrial Engineering, Pontificia Universidad Javeriana, 2006

and

Ethem Ucev
B.S. Industrial Engineering, Istanbul Technical University, 2005
M.S. Engineering and Technology Management, Bogazici University, 2012

Submitted to the Engineering Systems Division in Partial Fulfillment of the Requirements for the Degree of

## Master of Engineering in Logistics

at the

## Massachusetts Institute of Technology

[June 2013]



Signature of Author.........................................................
Master of Engineering in Logistics Program, Engineering Systems Division
May 10, 2013

Signature of Author.........................................................
Master of Engineering in Logistics Program, Engineering Systems Division
May 10, 2013

Certified by.........................................................
Dr. Roberto Perez-Franco
Research Associate, Center for Transportation and Logistics, Thesis Supervisor

Accepted by.........................................................
Prof. Yossi Sheffi
Elisha Gray II Professor of Engineering Systems
Director, MIT Center for Transportation and Logistics
Professor, Civil and Environmental Engineering

# A Decision Support Model for Staffing Supply Chain Planners: A Case from the Consumer Packaged Goods Industry

by

Aura Carolina Castillo

and

Ethem Ucev

Submitted to the Engineering Systems Division in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Logistics

at the

Massachusetts Institute of Technology

## ABSTRACT

Reducing or increasing labor force is not always effective when done without a thorough analysis. Organizations could face negative consequences such us unbalanced workload, inefficient procedures, lost sales, and negative work atmosphere.

An increasing number of organizations are centralizing operations in order to optimize labor costs. However, not all companies assess the new number of employees required after centralization takes place, and for those companies that actually do this analysis, there are not quantitative tools, as far as we know in the literature, that can help them estimate the workforce required.

This thesis project provides practitioners with a new mathematical model to estimate an appropriate number of production planners required for the supply chain planning department of a company in the consumer packaged goods industry. Using bivariate correlation and multiple regression analysis, we explored whether a relationship exists between the required number of production planners in the new centralized offices of the Company and 13 factors that impact employee's workload. The resulting regression model accounts for 98% of the variance of the number of planners.

Thesis Supervisor: Dr. Roberto Perez-Franco
Title: Research Associate, Center for Transportation and Logistics

## Table of Contents

**List of Figures**

**List of Tables**

# 1. INTRODUCTION

This thesis project discusses the development of a mathematical model that will provide supply chain managers at our sponsor company with an estimated number of site integrated planners needed at its planning center, where the centralized supply chain planning operations are located.

## 1.1. Topic Overview

Organizations constantly face different challenges to achieve new productivity objectives and organizational goals. Within the supply chain structures of most manufacturing companies, the centralization model is one of many options used to address these challenges. Companies are increasingly using this model to concentrate the decision-making process of supply chain planning in one specialized center.

The centralization model requires a number of changes within organizations. One of the most critical aspects of centralization is that it affects employee functions. For example, two roles can be merged into one, particularly in the case of production planners; companies can assign multiple manufacturing sites to one planner, instead of keeping one planner per plant. Having the right number of planners and size of a team is crucial for the work environment. Employees who are overloaded with work can easily become stressed, and those who are underloaded evidence inefficient use of resources. These conditions can lead to high employee turnover and negative impacts on costs, productivity, and work atmosphere.

## 1.2. Impact on the Consumer Goods Industry

To meet aggressive improvement and productivity objectives, our thesis sponsor, a global leader in consumer-packaged goods, is considering expanding the centralization model to all headquarters offices around the world. Having decided to centralize its supply chain planning activities and consolidate them into one planning office at headquarters locations, the company needs to increase the visibility of the impact on planning resources as a result of changes in business plans.

Our sponsor now faces the challenge of rightsizing the planning staffing to match any change in business plans at any given time. Therefore, the company needs to know whether the required number of planners is in place to face changes such as adding new SKUs, starting a new product line or manufacturing site, or running on tight capacity and low inventories. Additionally, our sponsor Company needs a way to identify the need for more human resource.

How do centralized supply chain managers determine the right size of their teams in order to maintain good quality and service? Traditionally, most approaches to staff sizing have been qualitative, based on experience or on average numbers from the past, or both. The problem with using these intuitive approaches is that the baseline resources already in place will inaccurately influence the new estimation.

Few quantitative techniques have been developed to determine the workforce required in a company; most of the existing literature is related to optimally allocate resources in manufacturing sites, where the number of employees required is easily associated with the equipment or machine requirements. As far as we know, no previous quantitative studies have been conducted to estimate human resources needed in the planning departments after centralizing operations, as part of the supply chain redesign or as a result of changes in business plans.

## 1.3. Research Question

Considering the problem described above, the key question in our research is how to determine the right size of the company's production planning team to achieve its business targets, after centralizing supply chain planning activities.

To address this problem, we have developed a mathematical model that provides statistical and quantitative support for staffing decisions to supply chain directors at our sponsor company. The model helps them estimate the number of production planners required, combining the human expertise and experience with a more precise quantitative tool.

## 1.4. Motivation

This model will provide the company with increased visibility in determining the right size of its production planning team and with improved adaptability to changes in business conditions. Therefore, the company will have new quantitative support that combined with current efforts will contribute to avoid any costs associated with placing either excessive or insufficient resources. Another potential benefit will be avoiding negative impacts of wrong staff sizing on the planning team, such as low employees' morale and productivity.

This research will also contribute to knowledge in the field of supply chain management on the subject of business centralization and right-sizing problems, which an increasing number of companies find relevant nowadays.

### 1.5. Thesis Outline

Chapter 2 presents a review of current and previous studies conducted in staff rightsizing and centralization. Chapter 3 presents an outline of the methodology developed to address the key question. Chapter 4 discusses the data collection and analysis required. Chapter 5 states our conclusions and findings from the analysis and presents recommendations for future research.

# 2. LITERATURE REVIEW

We reviewed literature related to centralization and current rightsizing techniques. These sources helped us understand the need for quantitative approaches in order to interrelate critical variables that affect employees' workload and to build mathematical models that can provide guidance to supply chain decision makers.

In sections 2.1 and 2.2 of this literature review, we will introduce past research conducted on the centralization and right staff-sizing topics, their relevance to our research, and the impact on the question of rightsizing teams in an organization. In section 2.3, we will examine the increasing need nowadays for quantitative tools and mathematical models as supporters for job design and staff sizing.

## 2.1. Overview of Centralization and Rightsizing

Change is an ongoing and inevitable condition in today's organizations. Companies have to remain adaptable and resilient in the face of changing conditions of their business environment in order to survive, remain competitive, and to be successful. Centralization of supply chains, which is defined as "consolidation of operations at corporate level rather than business units" (Droge & Germain, 1989) is considered a powerful tool for companies to adapt to changing business conditions and cut their costs.

Rangavittal & Sohn (2008) suggest an integrated framework for the three factors that determine the centralization or decentralization decisions of companies: customer service, supply chain management cost, and organizational control.

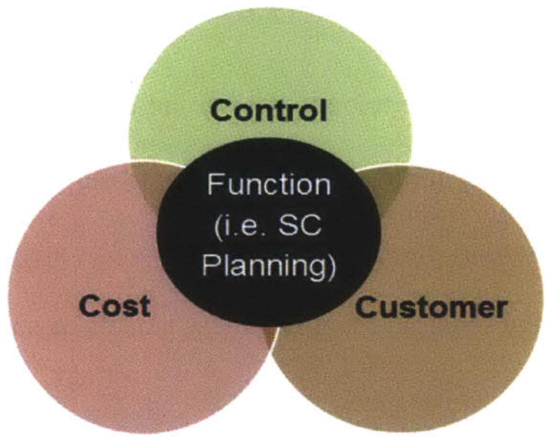



Figure 1. An integrated framework for centralization of a function (Rangavittal & Sohn, 2008)

Lewin & Minton (1986) approach the trade-offs between centralization and decentralization from an organizational effectiveness point of view. When it is time for a company to reduce costs by increasing productivity and efficiency, centralization becomes the reasonable option.

According to Rangavittal & Sohn (2008) centralization of supply chain operations and planning not only generates shared business and decision making processes within the organization, but also lets the company leverage the scale for various resources, including human resources. The opportunity for economies of scale in human resources, as a result of a centralization decision, is expected to lead to downsizing, which is the "conscious use of permanent personnel reductions in an attempt to improve efficiency and/or effectiveness" (Wilkonson, 2005).

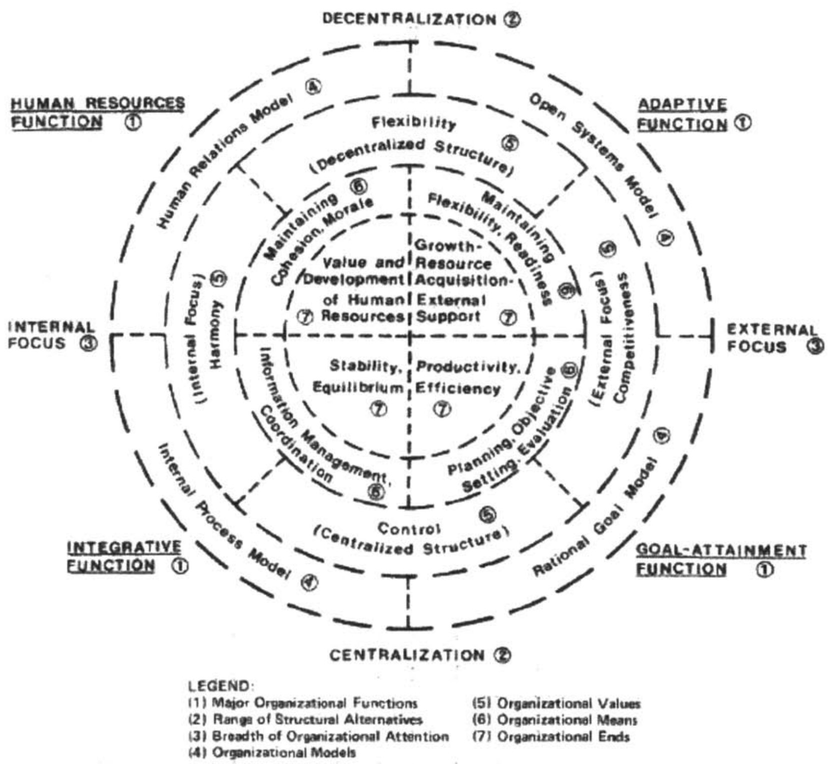



Figure 2. Spatial model of organizational effectiveness (Lewin & Minton, 1986)

## 2.2. Impact of Centralization and Rightsizing on Staffing Decisions

According to Lewin & Minton (1986), higher productivity and efficiency are two of the main drivers of centralization decisions. Inefficiencies due to nonstandard work processes for the same function of an organization in different business units, and diseconomies of scale due to administration of non-optimal size of activities and processes are likely to cause higher number of staff than required. Therefore, these inefficiencies are the target of centralization decisions.

Wilkonson (2005) claims productivity in workforce, which means savings in labor costs, is one of the expected benefits of downsizing an organization. As the essence of economies of scale implies, the average administrative time/cost decreases while the number of tasks increases. Usually same amount of people can still handle more activities and can create bigger output (Seddon, 2010).

Davison (2002), however, suggests that if management is either unaware of the right number of people their business needs, or lacks the knowledge of determining the correct number, cutting down on employees might turn out to be a wrong decision. Therefore, any centralization and rightsizing initiative requires well-established human resource planning and development strategies. "Eliminating people without first having a method for determining how many were needed to begin with is a short-term solution that is bound to create the same crisis over again and in a very short time" (Davison, 2002.)

### 2.3. Need for Staff-sizing Models

Centralization implies restructured process, and according to Morrall (1998), job design is crucial in an organization restructuring. Job design is all about providing job satisfaction to employees, which translates in increasing productivity. Managers must design jobs following strategic objectives and must also be aware of their personnel conditions (that is, whether they are understaffed or overstaffed), in order to take prompt actions.

The need to accurately size a team in an organization comes with reasons such us savings in hiring and turnover costs, reduced mismanagement and increasing workload balance. The literature is very rich about the impacts and importance of rightsizing in organizations. However, this literature is short when related to quantitative techniques that give support to qualitative ones.

As organizations become more globalized and complex, it is getting increasingly difficult to address organizational challenges without mathematical models and analytics capabilities. Our thesis project creates a mathematical model that provides a different approach to estimate human resources required within a certain organization, particularly within the supply chain-planning department.

# 3. METHODOLOGY

This section summarizes the methodology that we followed throughout our thesis to answer our research question. Our approach to the thesis question has basically two steps: identifying possible factors affecting the workloads of planners and understanding the impact of those factors on the prediction of the number of planners needed.

In order to identify possible factors we did a site visit to the sponsor company, and conducted interviews with personnel relevant to the subject. Then in order to understand impact of those factors on number of planners we collected data for the past twelve months for the key factors, chose an appropriate statistical analysis method and software, and developed our model based on the analysis that we performed on the data that we collected.

## 3.1. Site Visit, Interviews, and Identification of Inputs for Our Model

A planner's workload ultimately determines the number of planners required for the execution of a specific business function. Many factors affect that workload. These could be internal considerations such as capacities of manufacturing facilities, number of SKUs, and demand forecast accuracy, or external elements such as reliability of vendors, disruptions, and fluctuations in demand. Hence, we first wanted to define the key elements of the tasks that potentially affect the workload of production and material planners at our sponsor company. A better understanding of these factors was expected to increase prediction power of the model that we suggested to determine the right size of the production and material planning team.

### 3.1.1. Site Visit

In the very beginning of our study, we organized a site visit to the headquarters of our sponsor company. Our aim during the site visit was to meet production and material planners, their supervisors, and other planning groups with whom they work. In addition, we talked to two subject matter experts who have been with the company for almost 30 years and have a deep understanding of the dynamics of the company's supply chain and the supply chain planning activities. These experts also helped us choose our interview subjects based on the complexity of tasks within different product categories, level of interaction between different planning groups, and size of the planning group.

### 3.1.2. Interviews

We conducted interviews with a group of people who are currently doing the supply chain planning at the centralized planning office. The sample of our interview subjects included teams charged with production planning, demand planning, artwork design, material supply management, physical distribution, category supply planning, and distribution requirements planning. Through our interviews our aim was to:

- Understand the scope of production planning task as well as all supply chain planning activities in the centralized planning office
- Understand the level of standardization of processes across all planning teams in the centralized planning office
- Identify the key factors creating the complexity and affecting the workload of production and material planners

- Refine scope of our research to a level that both sponsor company and we agree on
- Build relationships for further communication and further research areas

Before our site visit and interviews we developed a set of generic interview questions aimed at identifying the drivers of the workload. These questions included:

- What are the main tasks that you spend most of your time on?
- What do you think affects your workload most?
- What kind of issues can make you spend time out of your standard tasks?
- How frequently do you work overtime?

The interviews we conducted during our site visit helped us judge the key independent variables of our potential model. Based on the learning from our interviews we could discuss in depth the key factors that subject matter experts proposed.

### 3.1.3. Identification of Key Workload Factors

The site visit and the series of interviews helped us figure out how well we understood our research problem and define scope of our project. By end of the site visit it was clear that possible key factors that affect the required number of planners were not well known. Given this fact and considering the framework presented by Perez-Franco (2012) in Figure 3, we had two options to consider:

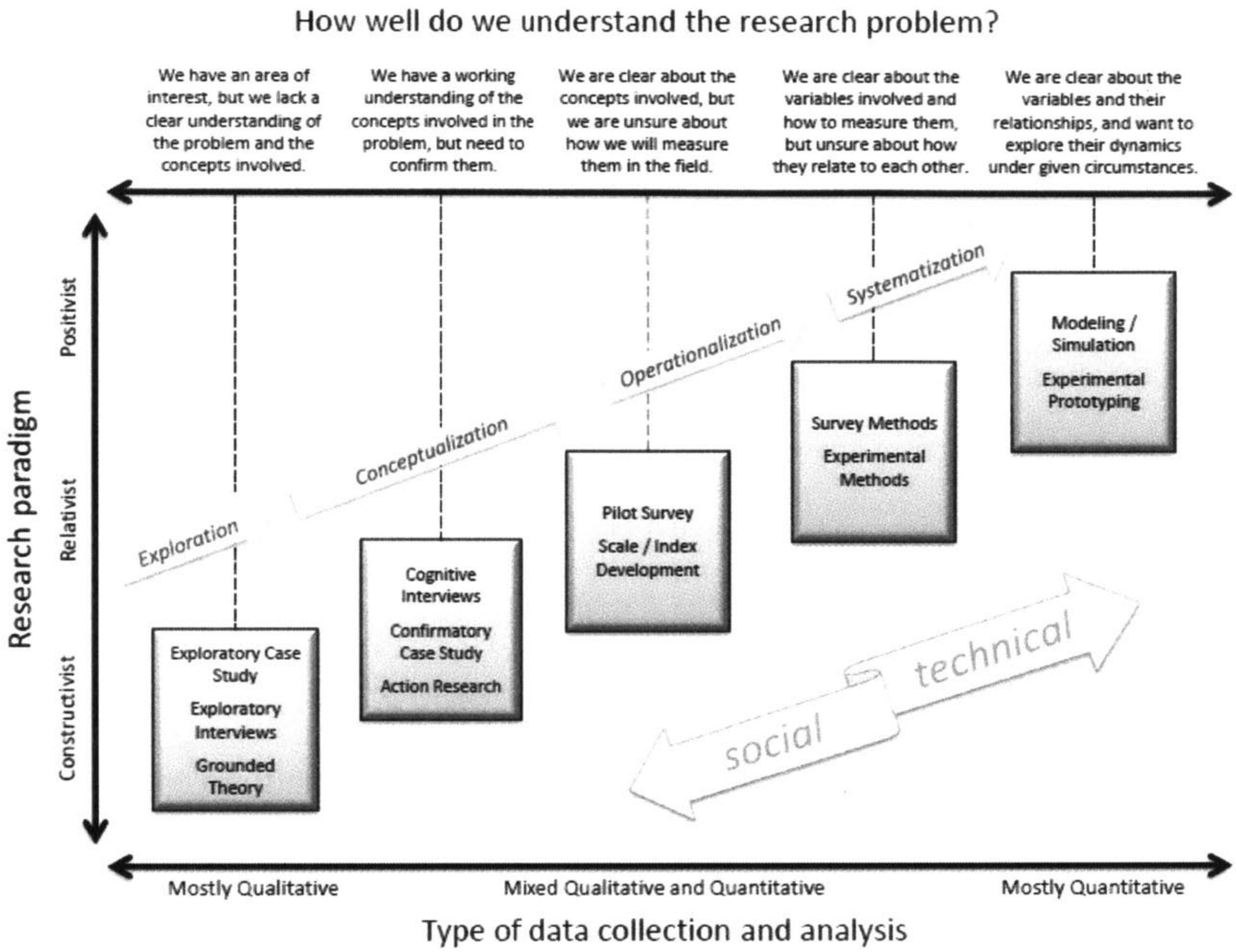



Figure 3. Research paradigm vs. type of data collection and analysis (Perez-Franco, 2012)

First option was focusing on early stages of the research, which were exploration and conceptualization, and spending our available time on identifying the key factors, so that we would be setting a solid foundation for future researchers to take over. The second option was assuming that the understanding of the experts of these variables was good enough so that we could take the variables that they would choose and proceed to the systematization phase and build our model.

These two options were presented to the subject matter experts for their decision and second option was chosen. The option that the sponsor company chose required us to assume that we started collecting data with the appropriate variables.

At the end of our visit we had a pool of 29 different possible factors affecting the required number of planners. After our discussions with our contacts in the company regarding potential factors that affect planner's workload, we ended up incorporating 13 of those 29 variables in our model based on the judgment of our contacts and our comments on each variable's potential impact and measurability.

### 3.2. Quantitative Data Collection and Analysis

After identifying key variables during our site visit, interviews with key staff, and follow-up phone meetings with experts, our next step was to collect quantitative data for these key variables.

#### 3.2.1. Data Templates

During our meetings with the subject matter experts we already identified some key characteristics of our variables, such as potential impact, unit of measurement, and type of data. We paid attention to define the variables in a way that they could be measured and past data for them could be found. Also, the way that we defined the variables was validated with the expert team. This approach helped us build templates to gather the data for the past twelve months in a format applicable to our data analysis.

We developed data templates for the 13 independent variables and one dependent variable: the required number of planners. The templates of our dependent variable were designed to collect both actual number of planners and required number of planners, just in case the actual number of planners deemed to be unsatisfactory or insufficient by subject matter experts for the same period.

Then, we communicated how to fill in these templates to our contacts in the sponsor company and got some feedback to improve the design of the templates. Our discussions during data collection showed that one of our variables—forecasted tightness of supply chain capacity—lacked almost half of the data points for the last twelve months. Population of the data for those missing points was deemed impossible. Therefore, this variable was eliminated before going to next phase of the research—data analysis.

### 3.2.2. Statistical Analysis Method and Software

After getting the data in a format that we requested, the next step was to identify the right statistical analysis method to understand how the variables relate to each other. More specifically, we were looking for a method to determine if any of our independent variables were significantly related to the dependent variable–the required number of planners.

After reviewing a series of methods such as factor analysis, MANOVA, multilevel modeling etc. we decided to use Multiple Regression for the analysis of our data. We chose Multiple Regression because it suited our purpose for this project, which was to learn more about the relationship between several independent or *predictor* variables and a dependent or *criterion* variable. As Pelham (2013) indicates, "multiple regression analyses are the most common analytic tool for determining which of several competing predictors of an outcome is the best predictor of that outcome is". We chose SAS JMP as our statistical analysis software because it has all the features required to work on our chosen statistical method, and also it was readily available at MIT's current software database.

After collecting all qualitative and quantitative data, and identifying our method for statistical analysis, we moved to the final step: data analysis, which is explained in the next chapter of the thesis.

# 4. DATA COLLECTION AND ANALYSIS

In this chapter we will assess to what extent the factors mentioned in chapter 3 are good predictors of our dependent variable: number of planners. Following the method indicated by Lehman, O'Rourke, Larry & Stepanski (2013) we checked the bivariate correlation among all variables, and performed a multiple regression analysis, which helped us identify statistically significant relationships. At the end of this chapter we will specify the regression analysis results.

## 4.1. Data Collection

### 4.1.1. Selection of variables that affect the planners' workload

Relying on the expertise of our contacts at the sponsor company, and with the findings from the interviews conducted on site, we identified 13 factors that affect the workload of the production and materials planners. These factors were then validated by our two subject expert contacts in the sponsor company, and classified as low, medium or high impact on the planners' workload.

In order to determine how good predictors these 13 factors are of the required number of production and materials planners, we requested as many data points as was possible for our sponsor company to obtain. Each set of data points was requested by product category, as each of the production and materials planners in the company is assigned to one product category. For the purpose of this thesis project, we will study sixteen product categories.

### 4.1.2. Level of measurement

According to Weinbach & Grinnell (2003), variables can be classified as listed in Table 1:

| Description of level of measurement | |
|---|---|
| **Nominal** | Uses discrete categories to classify variables. It is the least precise. Example: gender, industry type, etc. |
| **Ordinal** | Also uses discrete categories, but ones that can be ranked. Example: level of education, satisfaction, etc. |
| **Interval** | Places values on an equally spaced continuum, with a uniform unit of measurement, but without an absolute zero. Example: production or expiration date. |
| **Ratio** | Is possible when there is a fixed, absolute and non-arbitrary zero point. Numbers on a ratio scale indicate the actual amounts of the property being measured. Example: number of items manufactured. |

Table 1. Level of measurement used to classify predictor variables

Following this categorization, we listed the predictor variables in Table 2 with their correspondent level of impact on planner's workload, units of measurement, and level of measurement for each one. The company based on its expert opinions provided this information.

Our selection relies on the expertise of our two contacts at the sponsor company. The statistical analysis that we will show in the following sections is based on the assumption that all relevant variables were considered. In section 4.1.3, we provide a description for each predictor or independent variable.

| No | Variable | Potential Impact (Ranking) | Unit of measurement | Level of measurement |
|---|---|---|---|---|
| 1 | Number of re-scheduling exception messages | Medium | Number of exception messages (per category) | Ratio |
| 2 | Number of resources scheduled | Medium-High | Number of resources (per category) | Ratio |
| 3 | Master data (MD) Complexity | Medium | MD Complexity Scale 1-5 | Ordinal |
| 4 | Scheduling Complexity | High | Scheduling Complexity on Scale 1-5 (per resource) | Ordinal |
| 5 | Number of initiatives | High | Number of production versions (over the last 12 months) (per category) | Ratio |
| 6 | Number of manufacturing sites | Medium | Number of manufacturing sites (per category) | Ratio |
| 7 | Number of finished Products/Intermediate/ Raw material | Medium-High | Number of codes (per planner) | Ratio |
| 8 | Vendor Reliability | Low-Medium | Supplier schedule performance (per category) Scale 1-5 | Ratio |
| 9 | Demand Forecast Accuracy | Medium-High | MAPE (per category) | Ratio |
| 10 | Number of contractors | Medium | Number of contractors (per category) | Ratio |
| 11 | Flexibility and responsiveness of the SC | Medium | Flexibility and responsiveness (per category) Scale 1-5 | Ordinal |
| 12 | Forecasted tightness of supply chain capacity | Medium | Capacity to Demand (per category) | Ratio |
| 13 | Average Days on Hand Inventory | Medium | Average Days on Hand Inventory (per category) | Ratio |

Table 2. Characteristics of predictor variables

### 4.1.3. Description of predictor variables

*1. Number of re-scheduling exception messages*

It refers to the number of messages sent to planners to ask for changes to the current production plan. There are different types of exception messages for re-scheduling; but all can be classified under two main groups: those requiring moving production forward and those asking to delay production.

For this variable, the information available was a set of data that corresponds to a particular event on Summer 2012. One of the important assumptions made by our two subject experts is that the data points do not vary much month to month; therefore, for the purpose of our statistical analysis, the same values are repeated every month. The company will explore the possibility of measuring and recording this data in the future.

*2. Number of resources scheduled*

Typically it refers to a packing line, making line or combined resources associated by manufacturing site. Some examples are a portion of a line, a hand packing area, or physical work that needs to be done. These resources are coded by category product.

*3. Master data (MD) Complexity*

It is a characterization done by two experts at our sponsor company for material masters (MM), recipes, bill of materials (BOMs) and production versions (PV). At the code level, our sponsor company has three material types: raw materials, intermediate and finished products. The scale on this variable is from 1 to 5, where higher score reflects greater complexity.

After evaluating these factors, the sponsor company proposed the following scale (Table 3):

| Level of complexity | Description |
|---|---|
| 1 | Basic Master Data with limited BOMs |
| 2 | Basic Master Data with extensive BOMs |
| 3 | Added Complexity Factors (list) - - one of these gets a 3 |
| 4 | Added Complexity Factors (list) - - two of these gets a 4 |
| 5 | Added Complexity Factors (list) - - three or more of these gets a 5 |

Table 3. Master Data Complexity Scale

### *4. Scheduling Complexity*

It is a scale determined by our two experts in the subject, considering each product category, and key factors such us scheduling horizon, scheduling frequency, and tools to support planning tasks. The scale used for this category is described in Table 4.

| Level of complexity | Description |
|---|---|
| 1 | Low Scheduling Complexity – scheduling complexity requires well above average SIP Planner effort. |
| 2 | Below Average Scheduling Complexity – scheduling complexity below average SIP Planner effort. |
| 3 | Average Scheduling Complexity – scheduling complexity requires average SIP Planner effort. |
| 4 | Above Average Scheduling Complexity – scheduling complexity requires above average SIP Planner effort. |
| 5 | Very High Scheduling Complexity – scheduling complexity requires well above average SIP Planner effort. |

Table 4. Scheduling Complexity Scale

*5. Number of initiatives*

It refers to the count of new production versions in last twelve months; any change in a product, such us packaging, color, shape, etc., is considered an initiative. The number of initiatives can vary significantly from one category to the other.

There was not detailed data for this variable on a monthly basis, only the total for the year. Therefore, after validating with our sponsor company, we used the average of the year to populate the values per month, under the assumption that there is not significant variance month by month.

*6. Number of manufacturing sites*

It refers to the number of manufacturing sites owned by our sponsor company in the US, and where planning is engaged with a specific product category.

*7. Number of Finished Products/Intermediate/Raw material*

It refers to the total number of SKU's per category, considering all inventory types. This data can be generated anytime; however, no tracking is done on a monthly basis for active codes. This resulted in missing data that was populated with estimations from our sponsor company, assuming that no significant variations are present from one period to another within each category.

*8. Supplier Reliability*

This factor refers to how reliable the Component Material Supply Base (spanning raw materials & packing materials) is. Our sponsor company determined the scale shown in Table 5.

| Level of complexity | Description |
|---|---|
| 1 | Poor Supplier Reliability – component supply issues required well above average production and materials planner effort |
| 2 | Below Average Supplier Reliability – component supply issues required above average production and materials planner effort |
| 3 | Average Supplier Reliability – component supply issues required average production and materials planner effort |
| 4 | Above Average Supplier Reliability – component supply issues required below average production and materials planner effort |
| 5 | Exceptional Supplier Reliability – component supply issues required well below average production and materials planner effort |

Table 5. Supplier Reliability Scale

*9. Demand Forecast Accuracy*

It refers to the MAPE calculated per product category. The category Flavors & Fragrances/Chemicals is an internal customer within our sponsor company, and there is no internal demand tracking for this category, which results in missing data. According to our subject experts, some categories highly influence the value of the MAPE in this category; therefore, we calculated a weighted average based on the MAPE of those other predictor's values that influence the category Flavors & Fragrances/Chemicals.

*10. Number of contractors*

It is the number of companies that provide services to our sponsor company. Planners are in charge of monitoring that requirements from the company are fulfilled on time and in accordance with production schedule.

*11. Flexibility and responsiveness of the SC*

It is a scale from 1 to 5, where 5 represents higher responsiveness. Some key issues such as short cycle lengths in production, changeover flexibility, and staffing flexibility were considered for this ranking. The scale used is shown below in Table 6.

| Level of flexibility and responsiveness in the SC | Description |
| --- | --- |
| 1 | Very Low Supply Chain Flexibility & Responsiveness |
| 2 | Below Average Supply Chain Flexibility & Responsiveness |
| 3 | Average Supply Chain Flexibility & Responsiveness |
| 4 | Above Average Supply Chain Flexibility & Responsiveness |
| 5 | Very High Supply Chain Flexibility & Responsiveness |

Table 6. Supply Chain Flexibility and Responsiveness Scale

*12. Forecasted tightness of supply chain capacity (component, production)*

There was neither enough data available, nor any possible qualitative estimation. We found no plausible method to populate these empty cells. This variable was eliminated from the study.

*13. Average Days on Hand Inventory*

It is the amount of time that our sponsor company holds inventory before it is distributed.

### 4.2. Data Analysis

After reviewing several statistical methods, we concluded that the multiple regression approach is perfectly suited for this project as it explains to what extent a group of independent continuous variables has the power to predict one dependent continuous variable. In our study, we were able to extend the multiple regression method to include the categorical variables by using dummy variables, as explained by Field (2011).

To better understand the nature of the relationship between the dependent variables (number of planners) and the 12 predictors, we analyzed the data using several methods in JMP: univariate statistics, bivariate correlation, uniqueness index, and standardized multiple regression coefficients.

Following Lehman, Hatcher, O'Rourke and Stepanski (2013) we went through the next steps:

1. We looked at univariate statistics to check minimum and maximum values for each predictor, their distribution, and possible errors in the input data.
2. We used bivariate correlation to estimate what percentage of the variance in our dependent variable is accounted for each of the potential predictors.
3. We then reviewed the values for F-ratio and $R^2$ of the multiple regression analysis, in order to determine whether there was a significant relationship between the response variable and the multiple predictor variables examined as a group.
4. We reviewed the standard multiple regression coefficients of our analysis to check their statistical significance and the weight given to a specific predictor.

5. We performed a stepwise regression going backwards, and removed some predictors with no statistical significance for the model, in order to simplify the regression model, and avoid redundancy in the information provided by each independent variable.

### 4.2.1. Univariate Statistics

The JMP outputs of a univariate statistical analysis for each predictor variable are shown in Figures 4 through 6.

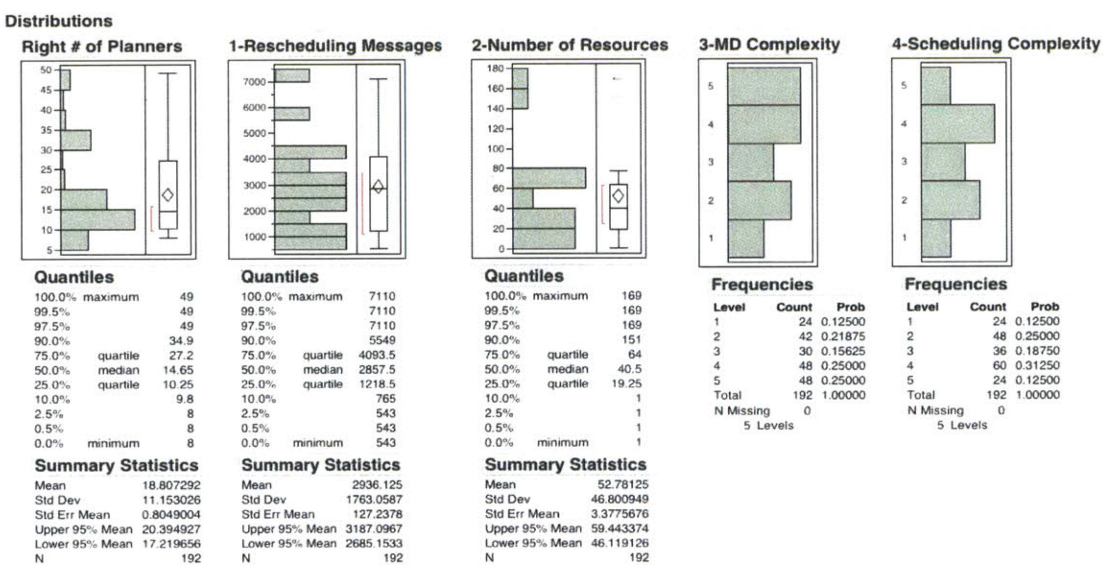
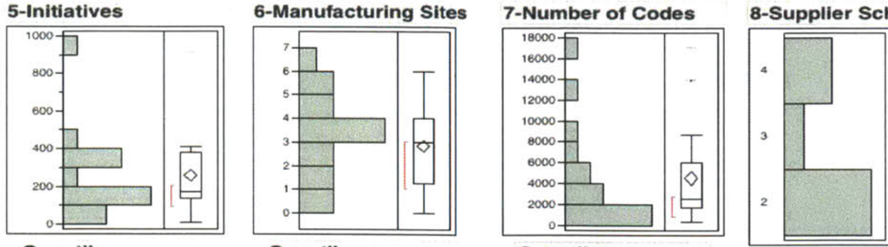



Figure 4. Univariate Statistics for Predictor Variables - Group 1

### 5-Initiatives | 6-Manufacturing Sites | 7-Number of Codes | 8-Supplier Schedule Performance



**5-Initiatives**

**Quantiles**

| | | |
|---|---|---|
| 100.0% | maximum | 909 |
| 99.5% | | 909 |
| 97.5% | | 909 |
| 90.0% | | 410 |
| 75.0% | quartile | 382.75 |
| 50.0% | median | 173.5 |
| 25.0% | quartile | 135.5 |
| 10.0% | | 90 |
| 2.5% | | 9 |
| 0.5% | | 9 |
| 0.0% | minimum | 9 |

**Summary Statistics**

| | |
|---|---|
| Mean | 260.75 |
| Std Dev | 206.52046 |
| Std Err Mean | 14.904331 |
| Upper 95% Mean | 290.14823 |
| Lower 95% Mean | 231.35177 |
| N | 192 |

**6-Manufacturing Sites**

**Quantiles**

| | | |
|---|---|---|
| 100.0% | maximum | 6 |
| 99.5% | | 6 |
| 97.5% | | 6 |
| 90.0% | | 5 |
| 75.0% | quartile | 4 |
| 50.0% | median | 3 |
| 25.0% | quartile | 1.25 |
| 10.0% | | 0 |
| 2.5% | | 0 |
| 0.5% | | 0 |
| 0.0% | minimum | 0 |

**Summary Statistics**

| | |
|---|---|
| Mean | 2.8125 |
| Std Dev | 1.7080806 |
| Std Err Mean | 0.1232701 |
| Upper 95% Mean | 3.0556456 |
| Lower 95% Mean | 2.5693544 |
| N | 192 |

**7-Number of Codes**

**Quantiles**

| | | |
|---|---|---|
| 100.0% | maximum | 17000 |
| 99.5% | | 17000 |
| 97.5% | | 17000 |
| 90.0% | | 13845 |
| 75.0% | quartile | 6025 |
| 50.0% | median | 2578.5 |
| 25.0% | quartile | 1714 |
| 10.0% | | 800 |
| 2.5% | | 394 |
| 0.5% | | 394 |
| 0.0% | minimum | 394 |

**Summary Statistics**

| | |
|---|---|
| Mean | 4550.0885 |
| Std Dev | 4666.1512 |
| Std Err Mean | 336.75046 |
| Upper 95% Mean | 5214.316 |
| Lower 95% Mean | 3885.8611 |
| N | 192 |

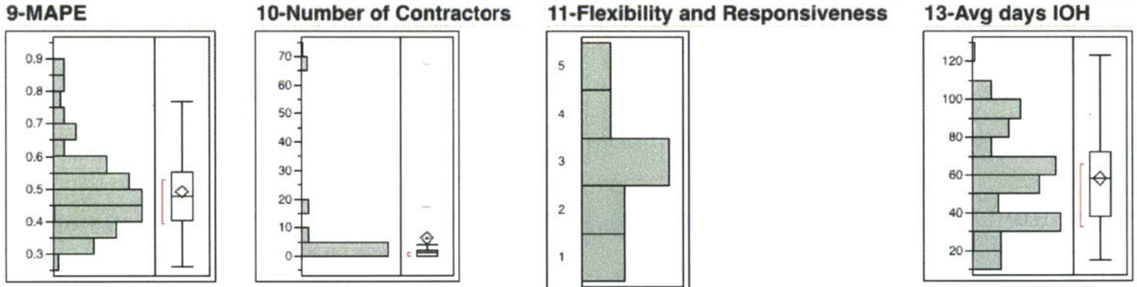
**8-Supplier Schedule Performance**

**Frequencies**

| Level | Count | Prob |
|---|---|---|
| 2 | 108 | 0.56250 |
| 3 | 24 | 0.12500 |
| 4 | 60 | 0.31250 |
| Total | 192 | 1.00000 |
| N Missing | 0 | |
| 3 Levels | | |

Figure 5. Univariate Statistics for predictor variables - Group 2

### 9-MAPE | 10-Number of Contractors | 11-Flexibility and Responsiveness | 13-Avg days IOH



**9-MAPE**

**Quantiles**

| | | |
|---|---|---|
| 100.0% | maximum | 0.897 |
| 99.5% | | 0.897 |
| 97.5% | | 0.8515 |
| 90.0% | | 0.664 |
| 75.0% | quartile | 0.55075 |
| 50.0% | median | 0.4785 |
| 25.0% | quartile | 0.40475 |
| 10.0% | | 0.3515 |
| 2.5% | | 0.30347 |
| 0.5% | | 0.261 |
| 0.0% | minimum | 0.261 |

**Summary Statistics**

| | |
|---|---|
| Mean | 0.4932479 |
| Std Dev | 0.1262651 |
| Std Err Mean | 0.0091124 |
| Upper 95% Mean | 0.5112217 |
| Lower 95% Mean | 0.475274 |
| N | 192 |

**10-Number of Contractors**

**Quantiles**

| | | |
|---|---|---|
| 100.0% | maximum | 70 |
| 99.5% | | 70 |
| 97.5% | | 68 |
| 90.0% | | 17 |
| 75.0% | quartile | 2 |
| 50.0% | median | 1 |
| 25.0% | quartile | 0 |
| 10.0% | | 0 |
| 2.5% | | 0 |
| 0.5% | | 0 |
| 0.0% | minimum | 0 |

**Summary Statistics**

| | |
|---|---|
| Mean | 6.2395833 |
| Std Dev | 16.45444 |
| Std Err Mean | 1.1874969 |
| Upper 95% Mean | 8.5818758 |
| Lower 95% Mean | 3.8972909 |
| N | 192 |

**11-Flexibility and Responsiveness**

**Frequencies**

| Level | Count | Prob |
|---|---|---|
| 1 | 36 | 0.18750 |
| 2 | 36 | 0.18750 |
| 3 | 72 | 0.37500 |
| 4 | 24 | 0.12500 |
| 5 | 24 | 0.12500 |
| Total | 192 | 1.00000 |
| N Missing | 0 | |
| 5 Levels | | |

**13-Avg days IOH**

**Quantiles**

| | | |
|---|---|---|
| 100.0% | maximum | 123 |
| 99.5% | | 123 |
| 97.5% | | 103.597 |
| 90.0% | | 96.58 |
| 75.0% | quartile | 72 |
| 50.0% | median | 58 |
| 25.0% | quartile | 37.925 |
| 10.0% | | 25.88 |
| 2.5% | | 16.6475 |
| 0.5% | | 15.3 |
| 0.0% | minimum | 15.3 |

**Summary Statistics**

| | |
|---|---|
| Mean | 58.296354 |
| Std Dev | 25.059925 |
| Std Err Mean | 1.8085443 |
| Upper 95% Mean | 61.863639 |
| Lower 95% Mean | 54.729069 |
| N | 192 |

Figure 6. Univariate Statistics for predictor variables - Group 3

After analyzing the data, we confirmed that there are no errors in the input data for the predictors subject to this analysis. For example, there is no evidence of negative values or values out of the range for ordinal variables.

### 4.2.2. Bivariate Correlation

#### 4.2.2.1. Bivariate analysis between response variable and predictor variables

We performed a *bivariate correlation* analysis in JMP between the dependent variable (number of planners) and each of the independent variables, in order to determine which variables are the best predictors for our dependent variable.

In preparation for this, we confirmed with our subject experts that the predictors are truly independent from the dependent variable, so they can perform the role of predictors in our model.

The analysis revealed seven predictor variables that were significantly related to the actual number of planners, with $r>0.45$ and low p-values. These are highlighted in Table 7.

- The most significant predictors related to the number of planners were: rescheduling messages ($r = 0.80$), number of contractors ($r = 0.75$) and scheduling complexity ($r = -0.70$). All of these correlations are significant at $p<0.001$.
- The correlation between the number of planners and the number of resources ($p = 0,67$) is not statistically significant.

| Potential Predictors | Right Number of Planners | Significance Probability (p-value) |
|---|---|---|
| 1-Rescheduling Messages | 0.80 | 0.00 |
| 10-Number of Contractors | 0.75 | 0.00 |
| 4-Scheduling Complexity | -0.70 | 0.00 |
| 7-Number of Codes | 0.66 | 0.00 |
| 11-Flexibility and Responsiveness | 0.53 | 0.00 |
| 9-MAPE | 0.48 | 0.00 |
| 3-MD Complexity | -0.47 | 0.00 |
| 5-Initiatives | 0.36 | 0.00 |
| 8-Supplier Schedule Performance | -0.26 | 0.00 |
| 13-Avg days IOH | -0.21 | 0.00 |
| 6-Manufacturing Sites | -0.16 | 0.02 |
| 2-Number of Resources | -0.03 | 0.67 |

Table 7. Bivariate correlation and significance probability between response variable and predictors

For some of the relationships (highlighted in Table 8), the expected positive or negative relationship was not validated in the correlation. For example, for variables 3-MD Complexity and 4- Scheduling complexity, we were expecting that, as the data reported higher score in these two predictors, the number of planners was also higher, but the negative sign in the correlation contradicts our intuition this time.

| Potential Predictors | Right Number of Planners | Expected Relationship |
|---|---|---|
| 1-Rescheduling Messages | 0.80 | + |
| 10-Number of Contractors | 0.75 | + |
| 4-Scheduling Complexity | -0.70 | + |
| 7-Number of Codes | 0.66 | + |
| 11-Flexibility and Responsiveness | 0.53 | - |
| 9-MAPE | 0.48 | + |
| 3-MD Complexity | -0.47 | + |
| 5-Initiatives | 0.36 | + |
| 8-Supplier Schedule Performance | -0.26 | - |
| 13-Avg days IOH | -0.21 | - |
| 6-Manufacturing Sites | -0.16 | + |
| 2-Number of Resources | -0.03 | + |

Table 8. Expected vs. actual relationship between response variable and predictors

### 4.2.2.2. Bivariate analysis among predictor variables

According to Lehman, Hatcher, O'Rourke and Stepanski (2013), the greater the correlation between the independent variables, the smaller the amount of unique variance in the dependent variable accounted for by each individual independent variable. Consequently, we wanted to determine the correlation among the independent variables. When there is high correlation between the predictors themselves, there may be redundancy between them in the prediction of the dependent variable, which results in a decrease of the total amount of variance in the dependent variable accounted for the linear combination of independent variables. We confirmed the absence of suppressor variables[1] with high correlation with other independent variables, and with zero or near-zero correlation with the dependent variable. Results are shown in Table 9:

| Variables | 1-Rescheduling Messages | 2-Number of Resources | 3-MD Complexity | 4-Scheduling Complexity | 5-Initiatives | 6-Manufacturing Sites | 7-Number of Codes | 8-Supplier Schedule Performance | 9-MAPE | 10-Number of Contractors | 11-Flexibility and Responsiveness |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Rescheduling Messages | 1.00 | | | | | | | | | | |
| 2-Number of Resources | 0.35 | 1.00 | | | | | | | | | |
| 3-MD Complexity | -0.21 | 0.29 | 1.00 | | | | | | | | |
| 4-Scheduling Complexity | -0.65 | 0.02 | 0.59 | 1.00 | | | | | | | |
| 5-Initiatives | 0.36 | -0.07 | -0.33 | -0.57 | 1.00 | | | | | | |
| 6-Manufacturing Sites | 0.03 | 0.43 | 0.53 | 0.45 | -0.27 | 1.00 | | | | | |
| 7-Number of Codes | 0.60 | 0.06 | -0.35 | -0.58 | 0.71 | -0.22 | 1.00 | | | | |
| 8-Supplier Schedule Performance | -0.40 | 0.00 | 0.18 | 0.24 | -0.36 | 0.13 | -0.25 | 1.00 | | | |
| 9-MAPE | 0.36 | 0.04 | -0.28 | -0.37 | 0.58 | -0.04 | 0.53 | -0.29 | 1.00 | | |
| 10-Number of Contractors | 0.44 | -0.29 | -0.51 | -0.53 | -0.01 | -0.47 | 0.43 | 0.02 | 0.16 | 1.00 | |
| 11-Flexibility and Responsiveness | 0.57 | 0.03 | -0.08 | -0.56 | 0.59 | 0.19 | 0.60 | -0.32 | 0.55 | 0.17 | 1.00 |
| 13-Avg days IOH | -0.20 | -0.07 | -0.03 | -0.02 | -0.12 | -0.57 | -0.13 | 0.31 | -0.22 | 0.08 | -0.36 |

Table 9. Correlation coefficients among predictor variables

As shown in Table 9, there is high correlation between some of the independent variables. For example, scheduling complexity presents a high correlation with four of the other predictors as follows: MD complexity (r=0.59); initiatives (r= -0.57); number of codes (r=-0.58), and flexibility and responsiveness (r=-0.56).

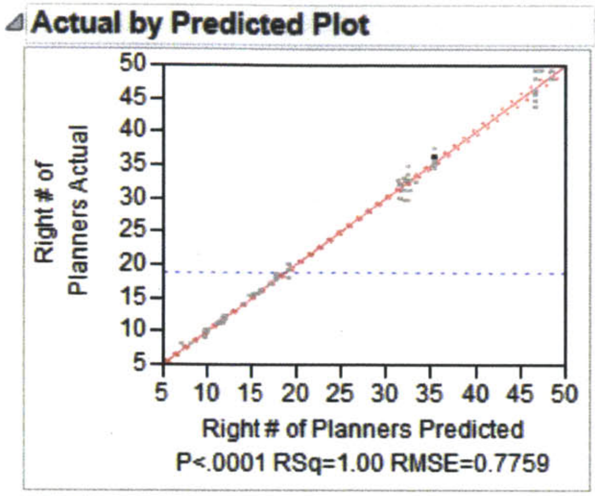
---

1 Suppressor variable is a predictor variable that improves the predictive power of a multiple regression equation by controlling for unwanted variation that it shares with other predictors.

### 4.2.3. Multiple Regression Results

We then ran the multiple regression analysis, with the following results: In Figure 7, the predicted plot graph and analysis of variance table tell us that the linear combination of all 12 predictor variables explains about 99.56% of the variance in the number of planners, the value of the $R^2$. The adjusted $R^2$, which corresponds more closely to the population value of the sample, is also high: 99.51%.

Our hypothesis is that the independent variables have an effect on our dependent variables. The null hypothesis is that all effects are zero. After running the multiple regression, the F-value (F ratio) is 1,871 and its associated probability (Prob > F) is less than 0.0001. The p-value is very small so we reject the null hypothesis and conclude that the predictors, taken as a group, do account for a significant amount of variation in the response.




**Summary of Fit**

| | |
|---|---|
| RSquare | 0.995693 |
| RSquare Adj | 0.995161 |
| Root Mean Square Error | 0.775852 |
| Mean of Response | 18.80729 |
| Observations (or Sum Wgts) | 192 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 21 | 23656.159 | 1126.48 | 1871.404 |
| Error | 170 | 102.331 | 0.60 | Prob > F |
| C. Total | 191 | 23758.490 | | <.0001* |

Figure 7. Results of multiple regression analysis

### 4.2.4. Significance Test for coefficients

We also looked at the t-ratio and p-values for each of the regression coefficients in order to accept or reject the null hypothesis of this study. Our study was based on only one sample; therefore, only one set of regression coefficients was encountered. Results from this analysis are presented in Table 10. As our study includes ordinal variables, JMP creates dummy variables to enable these ordinal predictors to be part of the regression. These dummy variables can also be seen in Table 10.

**Parameter Estimates**

| Term | | Estimate | Std Error | t Ratio | Prob>\|t\| | Std Beta |
|---|---|---|---|---|---|---|
| Intercept | | -238.5239 | 56.23884 | -4.24 | <.0001* | 0 |
| 1-Rescheduling Messages | | 0.047188 | 0.013655 | 3.46 | 0.0007* | 7.459426 |
| 2-Number of Resources | | 0.031634 | 0.017636 | 1.79 | 0.0746 | 0.132744 |
| 3-MD Complexity[2-1] | Biased | 77.359307 | 15.12124 | 5.12 | <.0001* | 2.29992 |
| 3-MD Complexity[3-2] | | 0.0220351 | 0.460053 | 0.05 | 0.9619 | 0.000941 |
| 3-MD Complexity[4-3] | | -94.36021 | 24.64832 | -3.83 | 0.0002* | -4.24131 |
| 3-MD Complexity[5-4] | | 3.2544741 | 14.58176 | 0.22 | 0.8237 | 0.126684 |
| 4-Scheduling Complexity[2-1] | Zeroed | 0 | 0 | . | . | 0 |
| 4-Scheduling Complexity[3-2] | | -82.22855 | 21.01731 | -3.91 | 0.0001* | -3.57865 |
| 4-Scheduling Complexity[4-3] | | 193.84891 | 48.98607 | 3.96 | 0.0001* | 8.644798 |
| 4-Scheduling Complexity[5-4] | | 67.209485 | 17.0937 | 3.93 | 0.0001* | 1.998162 |
| 5-Initiatives | | 0.1547719 | 0.037367 | 4.14 | <.0001* | 2.86591 |
| 6-Manufacturing Sites | | 1.4865715 | 2.207182 | 0.67 | 0.5015 | 0.227668 |
| 7-Number of Codes | | -0.004598 | 0.001748 | -2.63 | 0.0093* | -1.92373 |
| 8-Supplier Schedule Performance[3-2] | | -25.15225 | 15.05399 | -1.67 | 0.0966 | -1.12168 |
| 8-Supplier Schedule Performance[4-3] | | 18.573127 | 19.11493 | 0.97 | 0.3326 | 0.773905 |
| 9-MAPE | | -0.14777 | 0.900644 | -0.16 | 0.8699 | -0.00167 |
| 10-Number of Contractors | | 0.6307248 | 0.150848 | 4.18 | <.0001* | 0.93053 |
| 11-Flexibility and Responsiveness[2-1] | | 66.610621 | 18.45514 | 3.61 | 0.0004* | 2.337206 |
| 11-Flexibility and Responsiveness[3-2] | | -34.66367 | 7.719546 | -4.49 | <.0001* | -1.50859 |
| 11-Flexibility and Responsiveness[4-3] | | -7.868058 | 27.89303 | -0.28 | 0.7782 | -0.30627 |
| 11-Flexibility and Responsiveness[5-4] | | 10.491281 | 30.93732 | 0.34 | 0.7349 | 0.31191 |
| 13-Avg days IOH | | -0.002097 | 0.015061 | -0.14 | 0.8894 | -0.00471 |

Table 10. Parameter estimates and statistics for predictors

According to these results, the variables described in the following lines have, on one hand, very low t-ratios, and, on the other hand p-values much higher than 0.05. Details are as follows: MD complexity for values 2-3 (t-ratio= 0.05; p=0.96) and 4-5 (t-ratio= 0.22; p=0.82); manufacturing sites (t-ratio= 0.67; p=0.50); supplier schedule performance 2-3 (t-ratio= -1.67; p=0.10) and 3-4 (t-ratio= 0.97; p=0.33); MAPE (t-ratio= -0.16; p=0.87); flexibility and responsiveness for values 3-4 (t-ratio= -0.28; p=0.78) and 4-5 (t-ratio= 0.34; p=0.73); and average days of inventory on hand (t-ratio= -0.14; p=0.89). These results indicate that these predictor variables are not statistically significant for this model.

### 4.2.5. Validation of the initial model

We also checked the validity of the multiple regression model by evaluating the distribution of residual plots, multicollinearity and heteroscedasticity among variables. The residual plot in Figure 8 shows that error terms in our model are normally distributed, and the distribution is centered at zero. This result meets one of the underlined assumptions of multiple regression for the model to be valid.

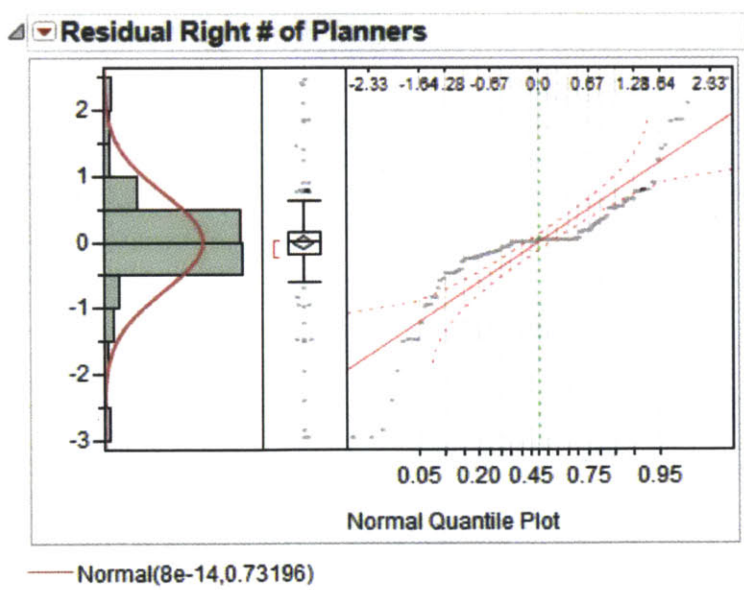



Figure 8. Normal distribution of residual plots – Initial Model

We checked multicollinearity among the predictor variables with the Variance Inflation Factor (VIF); according to Carver (2011) in general practice, VIF values over 10 may indicate a problem of collinearity. The results of our simulation are shown in Table 11.

**Parameter Estimates**

| Term | | Estimate | Std Error | t Ratio | Prob>\|t\| | VIF |
|---|---|---|---|---|---|---|
| Intercept | | -238.5239 | 56.23884 | -4.24 | <.0001* | . |
| 1-Rescheduling Messages | | 0.047188 | 0.013655 | 3.46 | 0.0007* | 183894 |
| 2-Number of Resources | | 0.031634 | 0.017636 | 1.79 | 0.0746 | 216.1745 |
| 3-MD Complexity[2-1] | Biased | 77.359307 | 15.12124 | 5.12 | <.0001* | 7976.9431 |
| 3-MD Complexity[3-2] | | 0.0220351 | 0.460053 | 0.05 | 0.9619 | 15.229005 |
| 3-MD Complexity[4-3] | | -94.36021 | 24.64832 | -3.83 | 0.0002* | 48446.067 |
| 3-MD Complexity[5-4] | | 3.2544741 | 14.58176 | 0.22 | 0.8237 | 12716.42 |
| 4-Scheduling Complexity[2-1] | Zeroed | 0 | 0 | . | . | 0 |
| 4-Scheduling Complexity[3-2] | | -82.22855 | 21.01731 | -3.91 | 0.0001* | 33022.445 |
| 4-Scheduling Complexity[4-3] | | 193.84891 | 48.98607 | 3.96 | 0.0001* | 188360.4 |
| 4-Scheduling Complexity[5-4] | | 67.209485 | 17.0937 | 3.93 | 0.0001* | 10193.756 |
| 5-Initiatives | | 0.1547719 | 0.037367 | 4.14 | <.0001* | 18895.895 |
| 6-Manufacturing Sites | | 1.4865715 | 2.207182 | 0.67 | 0.5015 | 4509.9192 |
| 7-Number of Codes | | -0.004598 | 0.001748 | -2.63 | 0.0093* | 21113.999 |
| 8-Supplier Schedule Performance[3-2] | | -25.15225 | 15.05399 | -1.67 | 0.0966 | 17788.832 |
| 8-Supplier Schedule Performance[4-3] | | 18.573127 | 19.11493 | 0.97 | 0.3326 | 25038.718 |
| 9-MAPE | | -0.14777 | 0.900644 | -0.16 | 0.8699 | 4.1034419 |
| 10-Number of Contractors | | 0.6307248 | 0.150848 | 4.18 | <.0001* | 1954.8743 |
| 11-Flexibility and Responsiveness[2-1] | | 66.610621 | 18.45514 | 3.61 | 0.0004* | 16550.195 |
| 11-Flexibility and Responsiveness[3-2] | | -34.66367 | 7.719546 | -4.49 | <.0001* | 4454.9073 |
| 11-Flexibility and Responsiveness[4-3] | | -7.868058 | 27.89303 | -0.28 | 0.7782 | 46530.36 |
| 11-Flexibility and Responsiveness[5-4] | | 10.491281 | 30.93732 | 0.34 | 0.7349 | 33390.826 |
| 13-Avg days IOH | | -0.002097 | 0.015061 | -0.14 | 0.8894 | 45.200808 |

Table 11. Multicollinearity among independent variables

Even though there are signs of collinearity among variables (most VIF values are over 10) this problem does not affect the validity of the model as a predictive tool According to Meyers, Gamst and Guarino (2010), collinearity does not matter for models that are used only with a prediction purpose, as is the case with our model.

Finally, in order to check heteroscedasticity, we looked at the residual plots and discovered that the variance of error terms around the regression line is not consistent for the different values of the dependent variable (Figure 9). This result indicated heteroscedasticity and questioned the validity of this model.

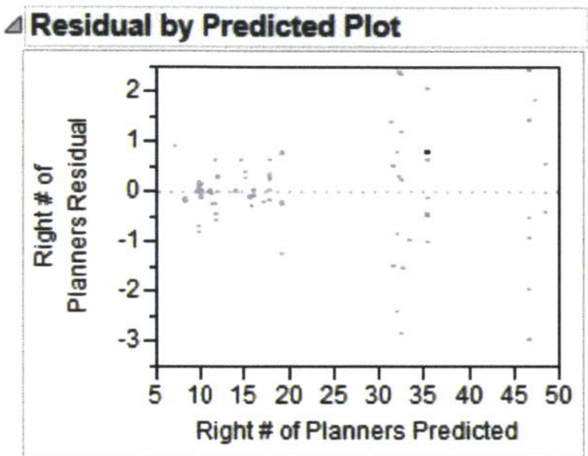



Figure 9. Residuals by Right Number of Planners Predicted

With this outcome, and looking at the results of all the previous statistical analysis we performed in JMP, we started a stepwise regression, in order to keep only those variables relevant for this study, find a more statistically valid model, and eliminate redundant variables.

### 4.2.6. Reduced Regression Model (Stepwise Approach)

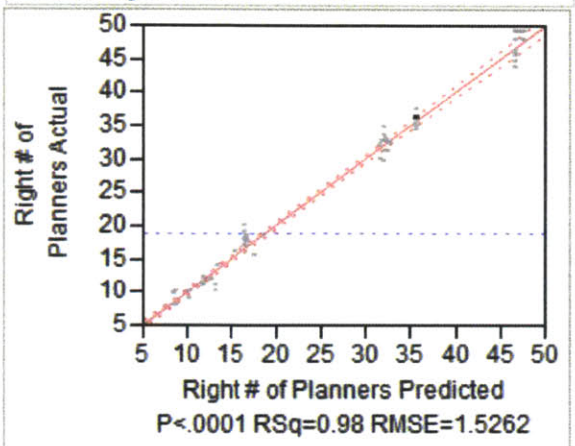
In the previous section we built a comprehensive regression model that includes all candidate independent variables. In this section we looked for a less comprehensive sub model built from the same set of candidate predictor variables by removing predictors from our initial model—in a stepwise manner—until there is no justifiable reason to remove any more. Thus, we ended up with a simpler yet powerful model. Practitioners prefer simple models because they are easier and less costly to put into practice in predicting and controlling the outcome in the future.

For our analysis we used the backward stepwise regression method in JMP as recommended by Hair, Black, Babin and Anderson (2010). We started the analysis with all of the 12 independent variables; this step yielded an $R^2$ of 99.56%. Our approach in removing a variable was to look at t- and p-statistics as well as the bivariate correlation of independent variables with the dependent variable, and inter-correlation between independent variables. For example, we started by removing the *6-Number of Manufacturing Sites* variable because the F-ratio for this variable was very low (0.45), and the p value (0.50) was very high, showing that this variable was insignificant for the model. The correlation of this variable with the number of planners was 0.16, which indicated a very low correlation. All these factors made *6-Number of Manufacturing Sites* a good candidate to remove from the model. When we deleted this variable, the $R^2$ was not affected.

As a different example, we also removed some independent variables, such as 4-Scheduling Complexity or 7-Number of Codes, which in fact had high correlations with number of planners. We could remove those variables as they were strongly correlated with some other independent variables that already accounted for the variance in the model.

Following the same approach, we removed eight of our independent variables from the initial model and ended up with a reduced model incorporating four independent variables without losing much from the prediction power. $R^2$ of the reduced model with four variables was 98.2%, compared to 99.5% for the model with 12 variables. Details of the regression model using four variables in JMP are provided in Figure 10:

**Actual by Predicted Plot**




**Summary of Fit**

| | |
|---|---|
| RSquare | 0.982254 |
| RSquare Adj | 0.981274 |
| Root Mean Square Error | 1.526214 |
| Mean of Response | 18.80729 |
| Observations (or Sum Wgts) | 192 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 10 | 23336.881 | 2333.69 | 1001.871 |
| Error | 181 | 421.609 | 2.33 | Prob > F |
| C. Total | 191 | 23758.490 | | <.0001* |

Figure 10. Results of multiple regression analysis - Reduced Model

With this reduced model, the linear combination of predictor variables explained about 98% of the variance in the number of planners. The F ratio was high and its associated probability (Prob > F) was less than 0.0001. The p-value is very small, so we could reject the null hypothesis that these independent variables have zero effect on the dependent variable, and concluded that the few remaining predictors, taken as a group, do account for a significant amount of variation in the response.

In addition, as shown below in Table 12, all variables included in the model had p-values less than 0.05, which proved that they were all significant.

**Parameter Estimates**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 21.520847 | 0.98351 | 21.88 | <.0001* |
| 1-Rescheduling Messages | 0.0008189 | 0.000153 | 5.36 | <.0001* |
| 3-MD Complexity[2-1] | -9.380846 | 0.953975 | -9.83 | <.0001* |
| 3-MD Complexity[3-2] | -4.359106 | 0.410895 | -10.61 | <.0001* |
| 3-MD Complexity[4-3] | 2.6870712 | 0.484673 | 5.54 | <.0001* |
| 3-MD Complexity[5-4] | -3.125918 | 0.446404 | -7.00 | <.0001* |
| 10-Number of Contractors | 0.2712364 | 0.016478 | 16.46 | <.0001* |
| 11-Flexibility and Responsiveness[2-1] | 1.2189386 | 0.438183 | 2.78 | 0.0060* |
| 11-Flexibility and Responsiveness[3-2] | 1.2745348 | 0.384629 | 3.31 | 0.0011* |
| 11-Flexibility and Responsiveness[4-3] | 15.120863 | 0.688616 | 21.96 | <.0001* |
| 11-Flexibility and Responsiveness[5-4] | -14.96659 | 0.998477 | -14.99 | <.0001* |

**Effect Tests**

| Source | Nparm | DF | Sum of Squares | F Ratio | Prob > F |
|---|---|---|---|---|---|
| 1-Rescheduling Messages | 1 | 1 | 66.9913 | 28.7599 | <.0001* |
| 3-MD Complexity | 4 | 4 | 1234.0798 | 132.4501 | <.0001* |
| 10-Number of Contractors | 1 | 1 | 631.1322 | 270.9501 | <.0001* |
| 11-Flexibility and Responsiveness | 4 | 4 | 1264.8707 | 135.7548 | <.0001* |

Table 12. Parameter estimates and statistics for predictors - Reduced Model

The intercept and regression coefficients, and the final equation used in predicting the required number of planners are shown in Figure 11.

**Prediction Expression**

$$
\begin{aligned}
&21.5208473774928 \\
&+0.00081894537338 * \text{Rescheduling Messages} \\
&+\text{Match}\left(\text{MD Complexity}\right)\begin{bmatrix} 1 & \Rightarrow 0 \\ 2 & \Rightarrow -9.3808462078673 \\ 3 & \Rightarrow -13.739951872512 \\ 4 & \Rightarrow -11.052880629135 \\ 5 & \Rightarrow -14.178798464721 \\ \text{else} & \Rightarrow . \end{bmatrix} \\
&+0.27123636003372 * \text{Number of Contractors} \\
&+\text{Match}\left(\text{Flexibility and Responsiveness}\right)\begin{bmatrix} 1 & \Rightarrow 0 \\ 2 & \Rightarrow 1.21893863023933 \\ 3 & \Rightarrow 2.49347342753468 \\ 4 & \Rightarrow 17.6143363446564 \\ 5 & \Rightarrow 2.64774578590246 \\ \text{else} & \Rightarrow . \end{bmatrix}
\end{aligned}
$$

Figure 11. Prediction expression - Reduced Model

### 4.2.7. Validation of the Reduced Model

When we looked at the outputs of the residual analysis in JMP for our model, we observed that the error terms are normally distributed with a mean of zero, as shown in Figure 12.

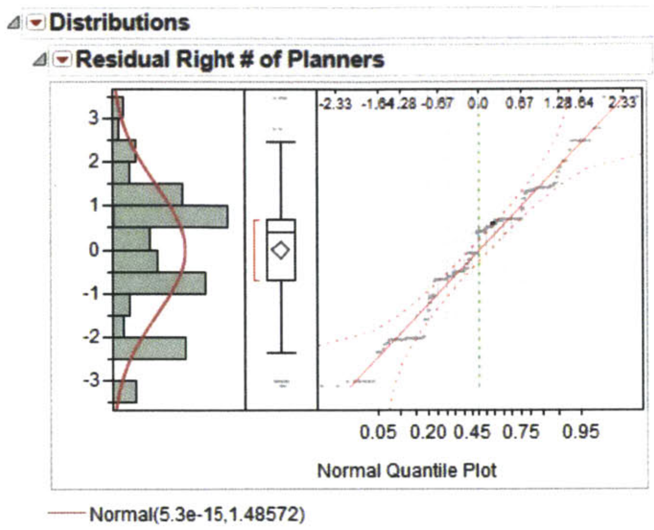



Figure 12. Statistics for residuals - Reduced Model

In addition, we observed no signs of a significant heteroscedasticity based on the residual by predicted plot. We looked at the residual plots and encountered a consistent variance of error terms around the regression line for the different values of the dependent variable (Figure 13)

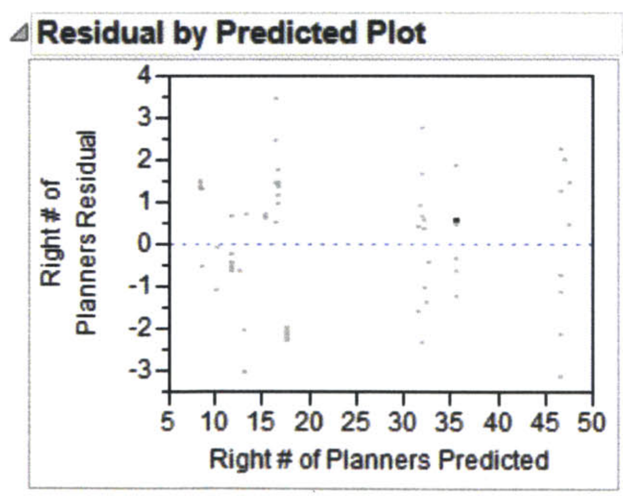



Figure 13. Residuals by Right Number of Planners Predicted - Reduced Model

Although multicollinearity does not matter for models that are used only with a prediction purpose, unlike the initial model we did not observe any multicollinearity with our new model. As shown in Figure 14, JMP output for all VIF (Variance Inflation Factors) values were less than 10, indicating that there was no multicollinearity among the variables in our model.

**Parameter Estimates**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| | VIF |
|---|---|---|---|---|---|
| Intercept | 21.520847 | 0.98351 | 21.88 | <.0001* | . |
| 1-Rescheduling Messages | 0.0008189 | 0.000153 | 5.36 | <.0001* | 5.9437191 |
| 3-MD Complexity[2-1] | -9.380846 | 0.953975 | -9.83 | <.0001* | 8.204689 |
| 3-MD Complexity[3-2] | -4.359106 | 0.410895 | -10.61 | <.0001* | 3.1393816 |
| 3-MD Complexity[4-3] | 2.6870712 | 0.484673 | 5.54 | <.0001* | 4.8406992 |
| 3-MD Complexity[5-4] | -3.125918 | 0.446404 | -7.00 | <.0001* | 3.0798323 |
| 10-Number of Contractors | 0.2712364 | 0.016478 | 16.46 | <.0001* | 6.0280249 |
| 11-Flexibility and Responsiveness[2-1] | 1.2189386 | 0.438183 | 2.78 | 0.0060* | 2.411043 |
| 11-Flexibility and Responsiveness[3-2] | 1.2745348 | 0.384629 | 3.31 | 0.0011* | 2.8580144 |
| 11-Flexibility and Responsiveness[4-3] | 15.120863 | 0.688616 | 21.96 | <.0001* | 7.3286777 |
| 11-Flexibility and Responsiveness[5-4] | -14.96659 | 0.998477 | -14.99 | <.0001* | 8.9880246 |

**Effect Tests**

| Source | Nparm | DF | Sum of Squares | F Ratio | Prob > F |
|---|---|---|---|---|---|
| 1-Rescheduling Messages | 1 | 1 | 66.9913 | 28.7599 | <.0001* |
| 3-MD Complexity | 4 | 4 | 1234.0798 | 132.4501 | <.0001* |
| 10-Number of Contractors | 1 | 1 | 631.1322 | 270.9501 | <.0001* |
| 11-Flexibility and Responsiveness | 4 | 4 | 1264.8707 | 135.7548 | <.0001* |

Figure 14. Multicollinearity among independent variables - Reduced Model

# 5. CONCLUSIONS AND RECOMMENDATIONS

This chapter provides our conclusions and recommendations based on the data analysis results.

## 5.1. Conclusions

The reduced model that we obtained in this research could serve as a tool that our sponsor company can use to estimate the number of production and material planners required in their recently centralized planning office. This model is valid within the range of data provided for the analysis. The prediction power of the model for new data beyond the ranges of past data will be lower.

The model is based on historic data, and it shows how closely the predicted number of planners matches with the sponsor company's best guess of the right number of planners subject of this study. The difference between the predicted number of planners suggested by our model and the required number of planners estimated by our sponsor company can be seen in Exhibit 1. For example, for Category 16, our model suggests that the predicted number of planners for the entire year is 0.5 people more than the required number based on expert opinion. On the other hand, for Category 15, the predicted number of planners is, on average, 2.8 people fewer than what it should be according to expert opinion.

One insight from the data analysis is that, despite what common sense would suggest about the most important factors that affect the workload of planners, i.e., number of SKUs or complexity of the supply chain, there are more critical predictors than those mentioned above that highly influence the planners' workload. Some examples are the number of rescheduling messages, and number of contractors. In fact, the number of rescheduling messages, master data complexity, number of contractors, and flexibility and responsiveness of the supply chain were the best predictors in our analysis, and these were enough to explain the variance of the required number of planners over 98%.

Although our reduced model included only four of the independent variables out of the initial 13 that were identified in the beginning of the research, some other variables are highly correlated with the number of planners, such as scheduling complexity, number of codes, and demand forecast accuracy. These are not included in the reduced model because they were highly correlated with other predictors included in the model. Nevertheless these independent variables are all useful parameters of the supply chain planning activities and need to be tracked for future research.

To summarize, after running several simulations using stepwise regression, we concluded that removing eight of the initial 12 independent variables only decreased the amount of variance explained by the final predictor variables as a group in 1.31 percentage point (From 99.5% to 98.2%). This reduction to only four variables will significantly facilitate the work on the part of the company of collecting data in order to feed the final model, while still giving a good prediction.

## 5.2. Recommendations

In order to facilitate the use of the model, we strongly suggest that the company start keeping better track of the data used in this thesis, as for some of the data points some estimations were necessary.

During our analysis, one variable was eliminated because there was not enough data or a method to estimate missing values. We recommend finding an alternative way to measure this variable and obtain the data, so that the prediction power of the variable *Forecasted tightness of supply chain capacity* can be assessed in future analyses.

Given the time frame and scope of this project, it was not possible to thoroughly examine all stages of the research problem (Figure 3). We do recommend that future research on this topic include the time to explore the problem through a very detailed exploration and conceptualization research process.

As a final recommendation, this study can be extended to other planning groups within the planning center, by following the same methodology described in this thesis project.

## EXHIBITS

## Exhibit 1 Required number of planners: predicted vs. expert opinion

| | Expert opinion | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **List of Categories** | **Jan-12** | **Feb-12** | **Mar-12** | **Apr-12** | **May-12** | **Jun-12** | **Jul-12** | **Aug-12** | **Sep-12** | **Oct-12** | **Nov-12** | **Dec-12** |
| **1** | 34.9 | 34.9 | 34.9 | 34.9 | 34.9 | 34.9 | 36.0 | 36.0 | 37.4 | 36.1 | 34.3 | 35.2 |
| **2** | 9.8 | 9.8 | 9.8 | 9.8 | 9.8 | 9.8 | 9.8 | 9.8 | 9.9 | 9.8 | 9.7 | 9.8 |
| **3** | 12.3 | 12.3 | 12.3 | 12.3 | 12.3 | 12.3 | 11.2 | 11.2 | 11.4 | 11.1 | 11.0 | 11.2 |
| **4** | 17.0 | 17.0 | 17.0 | 17.0 | 17.0 | 17.0 | 18.1 | 18.1 | 18.5 | 18.2 | 17.7 | 17.9 |
| **5** | 15.5 | 15.5 | 15.5 | 15.5 | 15.5 | 15.5 | 15.5 | 15.5 | 15.5 | 15.6 | 15.4 | 15.3 |
| **6** | 29.6 | 29.6 | 32.6 | 34.7 | 34.7 | 33.6 | 32.6 | 32.6 | 32.8 | 32.3 | 31.1 | 31.2 |
| **7** | 16.1 | 16.1 | 16.1 | 16.1 | 16.1 | 16.1 | 16.1 | 16.1 | 16.0 | 16.1 | 16.1 | 16.0 |
| **8** | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 |
| **9** | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 10.0 | 9.0 | 9.0 |
| **10** | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 |
| **11** | 49.0 | 48.0 | 49.0 | 49.0 | 49.0 | 49.0 | 48.0 | 46.0 | 45.6 | 44.6 | 43.6 | 43.6 |
| **12** | 32.0 | 32.0 | 32.0 | 32.0 | 32.0 | 32.0 | 32.0 | 32.0 | 32.0 | 30.0 | 30.0 | 30.0 |
| **13** | 11.0 | 11.0 | 11.0 | 11.0 | 11.0 | 11.0 | 11.0 | 11.0 | 11.0 | 11.0 | 11.0 | 11.0 |
| **14** | 12.0 | 12.0 | 12.0 | 12.0 | 12.0 | 12.0 | 12.0 | 12.0 | 12.0 | 12.0 | 12.0 | 12.0 |
| **15** | 19.0 | 19.0 | 19.0 | 19.0 | 19.0 | 19.0 | 19.0 | 20.0 | 20.0 | 20.0 | 20.0 | 18.0 |
| **16** | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 | 8.0 |

Table 13. Required number of planners - Expert Opinion

| | Estimation by the model | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **List of Categories** | **Jan-12** | **Feb-12** | **Mar-12** | **Apr-12** | **May-12** | **Jun-12** | **Jul-12** | **Aug-12** | **Sep-12** | **Oct-12** | **Nov-12** | **Dec-12** |
| **1** | 35.5 | 35.5 | 35.5 | 35.5 | 35.5 | 35.5 | 35.5 | 35.5 | 35.5 | 35.5 | 35.5 | 35.5 |
| **2** | 8.4 | 8.4 | 8.4 | 8.4 | 8.4 | 8.4 | 8.4 | 8.4 | 8.4 | 8.4 | 8.4 | 8.4 |
| **3** | 11.6 | 11.6 | 11.6 | 11.6 | 11.6 | 11.6 | 11.6 | 11.6 | 11.6 | 11.6 | 11.6 | 11.6 |
| **4** | 16.5 | 16.5 | 16.5 | 16.5 | 16.5 | 16.5 | 16.8 | 16.8 | 16.8 | 16.8 | 16.8 | 16.8 |
| **5** | 17.6 | 17.6 | 17.6 | 17.6 | 17.6 | 17.6 | 17.6 | 17.6 | 17.6 | 17.6 | 17.6 | 17.6 |
| **6** | 32.0 | 32.0 | 32.0 | 32.0 | 32.0 | 32.0 | 31.7 | 32.2 | 32.2 | 32.8 | 32.5 | 32.2 |
| **7** | 15.4 | 15.4 | 15.4 | 15.4 | 15.4 | 15.4 | 15.4 | 15.4 | 15.4 | 15.4 | 15.4 | 15.4 |
| **8** | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 8.7 | 8.7 | 8.7 | 8.7 | 8.7 | 8.7 |
| **9** | 10.1 | 10.1 | 10.1 | 10.1 | 10.1 | 10.1 | 10.1 | 10.1 | 10.1 | 10.1 | 10.1 | 10.1 |
| **10** | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 | 13.3 |
| **11** | 47.5 | 47.5 | 47.0 | 47.0 | 47.0 | 46.7 | 46.7 | 46.7 | 46.7 | 46.7 | 46.7 | 46.7 |
| **12** | 31.6 | 31.6 | 31.6 | 31.6 | 31.6 | 31.6 | 31.6 | 31.6 | 31.6 | 31.6 | 31.6 | 31.6 |
| **13** | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 | 13.0 |
| **14** | 12.7 | 12.7 | 12.7 | 12.7 | 12.7 | 12.7 | 12.7 | 12.7 | 12.7 | 12.7 | 12.7 | 12.7 |
| **15** | 16.5 | 16.5 | 16.5 | 16.5 | 16.5 | 16.5 | 16.5 | 16.5 | 16.5 | 16.5 | 16.5 | 16.5 |
| **16** | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 | 8.5 |

Table 14. Required number of planners - Regression Model

| | Variance | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **List of Categories** | **Jan-12** | **Feb-12** | **Mar-12** | **Apr-12** | **May-12** | **Jun-12** | **Jul-12** | **Aug-12** | **Sep-12** | **Oct-12** | **Nov-12** | **Dec-12** |
| **1** | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | 0.6 | -0.5 | -0.5 | -1.9 | -0.6 | 1.2 | 0.3 |
| **2** | -1.4 | -1.4 | -1.4 | -1.4 | -1.4 | -1.4 | -1.4 | -1.4 | -1.5 | -1.4 | -1.3 | -1.4 |
| **3** | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | 0.4 | 0.4 | 0.2 | 0.5 | 0.6 | 0.4 |
| **4** | -0.5 | -0.5 | -0.5 | -0.5 | -0.5 | -0.5 | -1.3 | -1.3 | -1.7 | -1.4 | -0.9 | -1.1 |
| **5** | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.1 | 2.0 | 2.2 | 2.3 |
| **6** | 2.4 | 2.4 | -0.6 | -2.7 | -2.7 | -1.6 | -0.9 | -0.4 | -0.6 | 0.5 | 1.4 | 1.0 |
| **7** | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.6 | -0.7 | -0.7 | -0.6 |
| **8** | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 | -1.3 | -1.3 | -1.3 | -1.3 | -1.3 | -1.3 |
| **9** | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 1.1 | 1.1 |
| **10** | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 | -0.7 |
| **11** | -1.5 | -0.5 | -2.0 | -2.0 | -2.0 | -2.3 | -1.3 | 0.7 | 1.1 | 2.1 | 3.1 | 3.1 |
| **12** | -0.4 | -0.4 | -0.4 | -0.4 | -0.4 | -0.4 | -0.4 | -0.4 | -0.4 | 1.6 | 1.6 | 1.6 |
| **13** | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 | 2.0 |
| **14** | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 |
| **15** | -2.5 | -2.5 | -2.5 | -2.5 | -2.5 | -2.5 | -2.5 | -3.5 | -3.5 | -3.5 | -3.5 | -1.5 |
| **16** | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |

Table 15. Difference between regression model and expert opinion

## BIBLIOGRAPHY


Christopher, M. (2011). *Logistics & Supply chain management* (4th ed.). Harlow, England; New York: Financial Times Prentice Hall.

Dröge, C., & Germain, R. (1989). The impact of the centralized structuring of logistics activities on span of control, formalization and performance. *Journal of the Academy of Marketing Science, 17*(1), 83.

Hunt, B., & Ivergard, T. (2010). Rightsizing and human resource development: Learning from a simulation model of the Thai public sector. *International Journal of Employment Studies, 18*(2), 90-116.

Morrall Jr., A. (1998). A human resource rightsizing model for the twenty-first century. Human Resource Development Quarterly, 9(1), 81-88.

Wilkinson, A. (2005). Downsizing, rightsizing or dumbsizing? quality, human resources and the management of sustainability. Total Quality Management & Business Excellence, 16(8), 1079-1088. doi: 10.1080/14783360500163326

Chien, C., Chen, W., & Hsu, S. (2010). Requirement estimation for indirect workforce allocation in semiconductor manufacturing. International Journal of Production Research, 48(23), 6959-6976. doi: 10.1080/00207540903352678

Seddon, J. (2010). Why do we believe in economy of scale? Retrieved from http://s3.amazonaws.com/connected_republic/attachments/33/Why_do_we_believe_in_economy_of_scale.pdf

Walch, D., & Merante, J. (2008). What is the appropriate business continuity management staff size? Journal of Business Continuity & Emergency Planning, 2(3), 240-250.

Sohn, Tae-Hee; Rangavittal, Guruprakash (Guruprakash Coimbatore). (2008) The impact and dynamics of centralization in supply chain decision-making. Massachusetts Institute of Technology. Engineering Systems Division.

Lewin, A. Y., & Minton, J. W. (1986). Determining organizational effectiveness: Another look, and an agenda for research. *Management Science, 32*(5), 514-538.

Davison, B. (2002). The Difference between Rightsizing and Wrongsizing, *The Journal of Business Strategy*, July-August, 23, 4, 31-35.

Pelham, B. W. (2013). *Intermediate Statistics: A Conceptual Course.* Multiple Regression (pp. 231). Thousand Oaks, CA: SAGE Publications.

Lehman, A., O'Rourke, N., Larry, H., & Stepanski, E. J. (2013). *JMP for Basic Univariate and Multivariate Statistics.* Cary, NC: SAS Institute.

Field, A. (2009). *Discovering Statistics using SPSS* (3rd ed.). London: SAGE Publications Ltd.

Meyers, L.; Gamst, G.; Guarino, A.J. (2013). *Applied Multivariate Research* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Carver, R. (2010). *Practical Data Analysis with JMP®*. Cary, NC: SAS Institute Inc.

Hair, J.; Black, W.; Babin, B.; Anderson, R. (2010). *Multivariate Data Analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.