Improved Federal IT Project Schedule Performance Prediction Model

by Raymond C. Reese II

B.S. in Mechanical Engineering, May 2007, Purdue University
M.S. in Systems Engineering, May 2011, The George Washington University
M.B.A., May 2015, College of William and Mary

A Praxis submitted to

The Faculty of
The School of Engineering and Applied Science
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Engineering

January 10, 2019

Praxis directed by

Ebrahim Malalla
Visiting Associate Professor of Engineering and Applied Science

Amir Etemadi
Assistant Professor of Engineering and Applied Science

The School of Engineering and Applied Science of The George Washington University certifies that Raymond C. Reese II has passed the Final Examination for the degree of Doctor of Engineering as of November 13, 2018. This is the final and approved form of the Praxis.

Improved Federal IT Project Schedule Performance Prediction Model

Raymond C. Reese II

Praxis Research Committee:

Ebrahim Malalla, Visiting Associate Professor of Engineering and Applied Science, Praxis Co-Director

Amir Etemadi, Assistant Professor of Engineering and Applied Science, Praxis Co-Director

Timothy Blackburn, Professorial Lecturer of Engineering Management and Systems Engineering, Committee Member

© Copyright 2019 by Raymond C. Reese II
All rights reserved

Dedication

I would like to dedicate this to my supportive parents, Raymond and Wendy Reese, my encouraging sisters, Raymonda and Wynndee Reese, my awesome and lovely wife, Tiara Reese, and the greatest gift that I have ever received, my daughter, Reagan.

Acknowledgements

I would like to first acknowledge God and Jesus for making this all possible, for without them none of this would have happened. I would also like to especially thank Dr. Malalla for continually supporting, advising, challenging, and transforming my thinking during this most rewarding journey. I would also like to thank Dr. Etemadi for the helpful feedback and additional support and Dr. Blackburn for the helpful feedback and immense data analytics support and insight throughout this research voyage. I would also like to acknowledge the friends, family, and loved ones that supported me on this journey, for without your prayers and support, none of this would have been possible.

Abstract of Praxis

Improved Federal IT Project Schedule Performance Prediction Model

The federal government currently collects information for the Information Technology (IT) Projects in which it invests, but it does not apply this information to more accurately predict the performance of these projects. As a result, project managers within the federal government are not effectively planning for and managing projects in an optimal manner. This produces low project performance and ineffective resource utilization. A tool that accurately predicts the performance of IT projects would allow project managers to more closely monitor at-risk projects and plan accordingly to optimize project performance. This research project originally attempted to develop a model that more effectively predicted cost and schedule performance of IT projects by examining both continuous and categorical predictors, such as planned project schedule duration, planned project lifecycle cost, project manager experience level, software development lifecycle methodologies, and product release frequency. Ultimately, the research was merely able to achieve better results for schedule performance with a misclassification rate of 31.5% relative to 45% for a comparable prediction model. Understanding the potential impact of predictors on schedule performance of IT projects before they begin enables organizations to strategically reallocate resources before and during projects to minimize downtime and the number of potentially unused resources.

Table of Contents

Dedication ..... **iv**

Acknowledgements ..... **v**

Abstract of Praxis ..... **vi**

List of Figures ..... **x**

List of Tables ..... **xii**

List of Acronyms ..... **xiii**

Chapter 1—Introduction 1

1.1 Background 1

1.2 Research Motivation 3

1.3 Problem Statement 6

1.4 Thesis Statement 6

1.5 Research Objectives 7

1.6 Research Questions and Hypotheses 8

1.7 Research Scope and Limitations 10

1.8 Organization of Praxis 11

Chapter 2—Literature Review 12

2.1 Introduction 12

2.2 Risk and Success Factors 13

2.3 Management Methods 18

2.4 IT Project Types 24

2.5 Performance Prediction Models and Methods 27

2.6 Research Opportunities in IT Project Management 32

2.7 Summary and Conclusion 33
Chapter 3—Methodology 35
3.1 Introduction 35
3.2 Data 36
3.2.1 Data Sources and Analysis Tools 36
3.2.2 Data Inclusion and Outliers 37
3.3 Analysis Methods 38
3.3.1 Response Variables 39
3.3.2 Predictor Variables 39
3.3.3 Model Development 41
3.3.4 Model Accuracy and Validation 42
3.3.5 Model Overview 44
3.3.6 Model Comparison Procedure 45
Chapter 4—Results 47
4.1 Introduction 47
4.2 Analysis Results for Continuous Predictor Variables 47
4.2.1 Effect on Cost Variance and Continuous Predictors 47
4.2.2 Effect on Schedule Variance 51
4.3 Analysis Results for Categorical Predictor Variables 53
4.3.1 Effect on Cost Variance 54
4.3.2 Effect on Schedule Variance 56
4.3.3 Two-Sample T-Test Analysis 57
4.3.4 Spearman Correlation Analysis 59

4.4 Regression Analysis Results 60
4.4.1 Cost Variance Regression Analysis 65
4.4.2 Schedule Variance Regression Analysis 66
4.5 The Schedule Variance Prediction Model 67
4.6 Validation and Comparison Results 70
4.7 Results of Hypothesis Testing 72
4.7.1 Cost-Related Hypotheses Discussion 72
4.7.2 Schedule-Related Hypotheses Discussion 76
4.7.3 Prediction Model-Related Hypotheses Discussion 79
Chapter 5—Conclusion 82
5.1 Discussion and Conclusions 82
5.2 Contributions to the Body of Knowledge 83
5.3 Recommendations for Future Research 84
References 86
Appendix A-Raw Data Plots 93
Appendix B-Model Variants Normality Graphs and P-values 95
Appendix C-Two Sample T-Test Box Plots of Sample Distribution and Anderson-Darling Test Results 98

List of Figures

Figure 1:Agile Development Model 19

Figure 2:Waterfall Model 20

Figure 3:Cost Variance (after Johnson Transformation) vs Planned Project Duration 48

Figure 4:Cost Variance (after Johnson Transformation) vs Planned Project Lifecycle Cost 49

Figure 5: Cost Variance (after Natural Log Transformation) vs Planned Project Duration 49

Figure 6: Cost Variance (after Natural Log Transformation) vs Planned Project Lifecycle Cost 50

Figure 7: Schedule Variance (after Johnson Transformation) vs Planned Project Duration 51

Figure 8: Schedule Variance (after Johnson Transformation) vs Planned Project Lifecycle Cost 52

Figure 9: Schedule Variance (after Natural Log Transformation) vs Planned Project Duration 52

Figure 10: Schedule Variance (after Natural Log Transformation) vs Planned Project Lifecycle Cost 53

Figure 11:Boxplot of Cost Variance vs PM Experience Level 54

Figure 12: Boxplot of Cost Variance vs SDLC Methodology 55

Figure 13:Boxplot of Cost Variance vs Release Frequency 55

Figure 14: Boxplot of Schedule Variance vs PM Experience Level 56

Figure 15: Boxplot of Schedule Variance vs SDLC Methodology 56

Figure 16: Boxplot of Schedule Variance vs Release Frequency 57

Figure 17: Residual Plots of Cost Variance (after Johnson Transformation)

vs Planned Project Lifecycle Cost 63

Figure 18:Residual Plots of Cost Variance (after Johnson Transformation)

vs Planned Project Duration..... 64

Figure 19: Residual Plots of Schedule Variance (after Natural Log Transformation)

vs Planned Project Duration..... 64

Figure 20: Residual Plots of Schedule Variance (after Natural Log Transformation)

vs Planned Project Lifecycle Cost 65

Figure 21: Proposed Prediction Model Framework..... 80

List of Tables

Table 1:Summary of Analysis Methods for Hypothesis Testing 38

Table 2: PM Level Definitions 40

Table 3:Independent Predictors Studied for the GRS and Proposed Model 45

Table 4: Summary of Cost Variance Correlations Statistics 50

Table 5: Summary of Schedule Variance Correlation Statistics 53

Table 6: Two-Sample T-Test Results 59

Table 7: Spearman Correlation Analysis Results of Categorical Predictors 60

Table 8: Ryan-Joiner Normality Results of Categorical Predictors 65

Table 9: Cost Variance Predictor P-values 66

Table 10: Schedule Variance Predictor P-values 66

Table 11: Accuracy Results of the 4 Variants of the Schedule Variance Prediction Model 68

Table 12: Schedule Variance Prediction Model Statistical Results 68

Table 13: Equations of the Categorical or Continuous Schedule Variance Prediction Model 70

Table 14: Subset of the Best Model Validation Results Relative to Historical Data 71

List of Acronyms

ANOVA:	Analysis of Variance
CIO:	Chief Information Officer
CPIC:	Capital Planning and Investment Control
DOI:	Degree of Interaction
EVM:	Earned Value Management
FITARA:	Federal Information Technology Acquisition Reform Act
HVA:	Hazard and Vulnerability Assessment
IT:	Information Technology
PLS:	Partial Least Squares
PM:	Project Manager
SDLC:	Software Development Lifecycle
SEMMA:	Sample Explore Modify Model Assess
VIF:	Variance Inflation Factor

Chapter 1—Introduction

1.1 Background

Many factors can impact the performance of IT projects. These include quantitative factors, such as project costs and duration, and qualitative factors, such as inadequate abilities and competencies related to project management experience and management methods (Junior and Carvalho, 2013). Furthermore, specifically within the federal government, the lack of management and project oversight has caused IT projects to fail to deliver the proper functionality (Government Accountability Office, 2013). Other factors which may influence mismanagement of IT projects that negatively affects performance include incomplete sets of criteria, diversity, basing projects on unrealistic targets, and applying ambiguous criteria (Hussein, 2013). Although the federal government has spent over $600 billion on IT during the past decade, these factors can stifle project performance and result in minimal productivity improvements relative to those that private industry has realized from IT (Government Accountability Office, 2013). Key findings also indicate that between 40% and 84% of information technology (IT) projects fail (Fisher, 2006). The data and findings demonstrate that the lack of effective project management is a root cause of federal IT projects' poor performance.

The size and complexity of projects can also affect performance. Elements related to the size and complexity of a project include budget, cost, man-hours, resources, project duration, technology level, impact, and cross-cultural nature. Specifically, within the federal government, projects vary in terms of size and complexity based on certain guidelines; these include small projects that cost less than $25k and require less than 6

months, midrange projects that cost between $25k and $200k and require up to a year to complete, and large projects that cost over $200k and require over a year for completion (Wilson, 2014).

Measuring project success also requires robust criteria to accurately track and monitor effectiveness. Various types of criteria are used to assess project success, including quantitative measures such as schedule variance, cost variance, and system uptime, as well as qualitative measures such as customer satisfaction levels and team effectiveness. According to Crumrine (2013), the federal government currently measures project success based on cost and schedule variance and calculates the variances as follows:

$$Cost\ Variance(Percentage) = \frac{Actual\ Cost - Planned\ Cost}{Planned\ Cost} X\ 100$$

Equation 1-1

$$Project\ Schedule\ Variance\ (Percentage) = \frac{Planned\ Duration - Actual\ Project\ Duration}{Planned\ Duration} X\ 100$$

Equation 1-2

The federal government also collects additional data that could potentially impact project success, such as project manager experience level, software development lifecycle methodology (SDLC), and product release frequency. Project manager experience level accounts for both the project manager's number of years of experience and their possession of a project management certification. Level 1 project managers are the most experienced, while those at level 9 are the least experienced. The SDLC methodology of projects is documented as either *waterfall* or *iterative. Waterfall* describes the process when all requirements are determined before a project begins Waterfall project

development occurs in phases which include the following: analysis, design, implementation, testing, deployment, and maintenance. The *iterative* process involves dividing a project into many smaller releases and delivers incrementally to the customer. Lastly, the product release frequency component of the federal government IT data relates to whether a project was structured for six-month product release frequencies.

1.2 Research Motivation

Currently, the federal government predominantly uses earned value management (EVM) techniques to predict cost and schedule variance; its limitations in predicting schedule variance have been widely recognized. One limitation of EVM is that the current metrics used to predict schedule variance are not effective throughout the entire life of a program (Crumrine, 2013). This method uses data collected as the program progresses; it is therefore not as effective before and during the early stages of project implementation. Additionally, ambiguity affects the measurement of schedule performance using EVM; this may potentially confuse program managers and thereby cause inconsistent data reporting and collection (Crumrine, 2013).

Recent research conducted by the Office of Management and Budget (2018) and the Government Accountability Office (2013) has demonstrated that current data collection practices related to IT projects consist of inconsistent metrics, unclear definitions, or metrics; This creates a compliance culture that ultimately inhibits performance. The research also determined that “many reporting requirements are interpreted as mere compliance exercises that are not consistently used to improve IT management and oversight” (Office of Management and Budget, 2018).

Furthermore, the Office of Budget and Management (OMB) (2018) proposes that to more effectively "utilize federal IT spending data, the federal government must integrate data collection efforts, standardize reporting data, and seek new opportunities to simplify, automate, and consolidate reporting". Although the OMB proposes some potentially useful measures that can be implemented to improve IT project spending and reporting, it neglects to specifically mention using the data to more accurately predict project performance, which would enable project managers to not only effectively report data but also to manage IT projects more effectively. Using Federal IT data, Moy (2016) studies the effects that earned value management (EVM), cost variance (CV), schedule variance (SV), and actual cost (AC) have on predicting IT project success. In this study, Moy (2016) also investigates potential outcomes (cost variance and schedule variance) as predictors of project success as opposed to inputs or indicators of project success.

Although Moy (2016) discovered a link between SV and IT program success, the EVM method, by design, becomes more accurate as a project continues to be executed and as more data becomes available, which limits its usability as an effective prediction method before a project is initiated (Lipke, Zwikael, and Anbari, 2009). In another study, Gemino, Sauer, and Reich (2010) examine the effects of budget, duration, effort, project management experience, relative size, requirements uncertainty, requirements difficulty, and technical complexity on predicting the success of an IT project. They use a regression model with a misclassification rate of 45%, a neural network model with a misclassification rate of 39%, and a classification tree model with a misclassification rate of 35%. The misclassification rate is the rate at which projects are incorrectly classified (as *performing* or *underperforming*) in the model. For example, if the actual classification

of a project is performing, and the model classifies the project as underperforming, then the model has misclassified the project.

Considering the rates at which the predictor models inaccurately classified a project, prediction models can be improved to classify projects at a significantly more accurate rate by examining the predictors that are most impactful. Another study conducted by Adoko (2016) also considers predictors such as system performance, technology maturity, schedule, reliability, and risk to determine the cost overrun of IT projects. Adoko (2016) cites schedule and reliability as the predictors that exert the greatest impact on cost overrun; he achieved a prediction accuracy of 62.1% and 83.3% for projects with and without cost overruns, respectively. Although this model yields favorable accuracy results, it does not address the impact of the predictors described by the federal government related to schedule variance.

To enhance project management within the federal government, it is necessary to implement a project performance prediction model driven by the IT project data that the federal government has garnered to achieve higher productivity. The proposed research and federal IT prediction model in this praxis expands upon previous research by examining predictors such as the scheduled duration of projects, types of project management methods, and years of experience of project managers to assess their impact on IT project performance.

1.3 Problem Statement

The federal government lacks the means for effective management oversight, which produces both cost and schedule variance issues and results in IT project failures (Government Accountability Office, 2013).

An effective early detection mechanism for IT projects would enable project managers to be more effective in planning for and providing the oversight necessary to ensure project success. Furthermore, ineffective collection and use of IT data and tools within the federal government has caused many projects to fail. Haji-Kazemi, Andersen, and Krane (2013) contend that despite recent improvements in the utilization of project management tools and techniques, a significant number of projects fail; this indicates that the current tools can be improved. Implementing a prediction tool across all agencies that incorporates federal government IT project data would establish a standardized method within the federal government that leverages historical IT project data to predict IT project performance.

1.4 Thesis Statement

The model proposed in this praxis to predict IT project performance will provide project managers with a more accurate and proactive method to detect predictors that could precipitate high schedule (and hence costs) variances before IT projects begin.

Such a model would allow organizations to plan accordingly and reallocate resources to minimize project inefficiencies and enhance the overall quality of the project (Alshamrani and Bahattab, 2015). Exploring the impacts of specific predictors on project performance will also allow project managers to positively impact team-building by

creating synergy and ultimately improving customer satisfaction (Ilincic, 2008). The implementation of this model will not only produce positive benefits related to quantitative components of IT projects, such as schedule variance, but this may also positively impact qualitative components, such as communication and customer satisfaction (Beleiu, Crisan, and Nistor, 2013).

1.5 Research Objectives

This research project aims to analyze the relationship between predictors that the federal government currently tracks to document details of its IT project investments on the federal IT dashboard and the impact of these parameters on IT project performance. In the first phase of the research project, potential predictors such as planned duration of the project, planned project lifecycle cost, experience of the project manager, SDLC methodology, and product release frequency have been identified as factors that could potentially impact IT project performance (cost and schedule variance). The second phase of the research project explores the impact of these parameters on federal IT project performance. The third phase of this research project entails inputting the parameters into a model to predict federal IT project performance. The final phase compares the proposed model to a similar prediction model, namely the Gemino, Sauer, and Reich (GRS) model (Gemino, Sauer, and Reich, 2010).

Based on the research phases, the research objectives of this study are as follows:

1. Based on research, identify the current project performance predictors that are being tracked along with additional project performance predictors that are not tracked for federal government IT projects;
2. Measure the effect of each identified predictor on federal government IT project performance;

3. Develop a model that could be used by the federal government to predict IT project performance based on the predictors that are currently being tracked;
4. Compare the model to a comparable prediction model that uses both continuous and categorical predictors.

1.6 Research Questions and Hypotheses

To adequately examine the foundation of predicting the performance of federal government IT projects based on the collected data and information related to cost and schedule, focused research questions and hypotheses must be developed to guide the research. The following questions and hypotheses concern predictors that could affect cost and schedule variance.

Research Questions:

1. Which parameters are effective predictors of IT project performance? (aligns with research objective 1)
2. Which predictors negatively affect IT project performance? (aligns with research objective 2)
3. Which predictors positively impact project performance? (aligns with research objective 2)
4. To what degree (in terms of percentage) does each predictor affect the accuracy of the IT project performance prediction model? (aligns with research objectives 2 and 3)
5. What is the accuracy of the proposed prediction model relative to historical results? (aligns with research objective 3)

6. How effective is the proposed performance prediction model relative to other comparable performance prediction models? (aligns with research objective 4)

Hypotheses:

Cost-related hypotheses:

- H_1: Projects with lengthier planned schedule durations tend to have higher cost variances.
- H_2: Projects with higher planned lifecycle costs tend to have higher cost variances.
- H_3: Projects overseen by managers with less experience tend to have higher cost variances.
- H_4: The type of SDLC (software development lifecycle) methodology affects cost variances.
- H_5: The frequency of project releases every six months affects cost variances.

Schedule-related hypotheses:

- H_6: Projects with lengthier schedule durations tend to have lower schedule variances.
- H_7: Projects with higher planned lifecycle costs tend to have higher schedule variances.
- H_8: Projects overseen by managers with less experience tend to have higher schedule variances.

- H_9: The type of SDLC (software development lifecycle) methodology affects schedule variances.
- H_{10}: The frequency of project releases every six months affects schedule variances.

Prediction model-related hypotheses:

- H_{11}: The prediction results of the model proposed in this praxis will be within a statistically relevant percentage of historical data results for cost and schedule variances.
- H_{12}: The prediction model proposed in this praxis will be more accurate in predicting performance relative to the GRS model.

1.7 Research Scope and Limitations

This research and the relevant data primarily relate to the performance of IT projects within the federal government. It encompasses information and data regarding IT projects collected from the federal IT dashboard. For instance, broader projects related to industries as finance, manufacturing, and marketing have been deemed beyond the scope of this research. The potential predictors used in this research project are from the 2017 IT dashboard. The dataset contains values for cost variance, schedule variance, PM (project management) experience level, SDLC methodology, planned project duration, and planned project lifecycle cost. During this research project, time constraints precluded interviews with the individuals responsible for inputting the data into the federal IT dashboard. Speaking with these individuals would have contributed context and insight to the data to improve comprehension. The years of the availability of data

have imposed an additional limitation. Since the IT dashboard was a recent initiative, more data is available for 2013 and later. Missing data also limited the amount of data that could be used within this research project.

1.8 Organization of Praxis

The praxis consists of five key chapters that comprehensively delineate the important elements of this research project. Chapter 2 presents a literature review which addresses the key areas of IT project risks, IT project management methods, the IT portfolio management process, IT project types, and IT project prediction methods. Chapter 3 discusses the research methodology, including data collection and the analytics process and methods used in this research project. Chapter 4 presents and highlights key findings and observations while linking these results to the hypotheses and research questions. The last chapter discusses final implications and conclusions and suggests recommendations for future research which can potentially contribute to the body of research.

Chapter 2—Literature Review

2.1 Introduction

Information technology (IT) projects within the federal government have become increasingly complex and more expensive, which has continuously necessitated more effective project management methods and tools. While the private sector, driven by profits and shareholder interests, continues to innovate in terms of IT project management, the federal government continues to lag. This research specifically addresses the federal government because, relative to the private sector, the federal government has a higher need for IT project management methods that enhance efficiency and productivity, minimize failed projects, and reduce costs (Government Accountability Office, 2013). The potential factors that can influence the outcomes of IT projects exert specific impacts and vary immensely. Additionally, the factors which potentially affect federal IT projects can influence those external to the federal government.

Therefore, by applying a holistic and comprehensive research approach, the literature review encompasses (both within and beyond the federal government) IT project risk and success factors, management methods, types of IT projects, and performance prediction methods. The information that is initially presented in the literature review discusses the details of the internal IT project environment. A discussion of the risk and success factors provides the foundation for the important elements specific to a project that ultimately affect its outcome. After comprehensively addressing the potential risk and success factors of projects, the literature review then discusses methods of managing the various risks that can negatively impact an IT project. IT projects can generally be categorized as

cloud, software, or hardware projects; the nuances of each are subsequently discussed to account for both unique and shared risks. This culminates in an exploration of prediction models that consider various factors to predict the outcomes of IT projects.

2.2 Risk and Success Factors

To determine what can impact IT project's performance, the factors that contribute to positive results of cost, schedule and performance must be explored. One factor that contributes to positive results is the effective modernization of technology. This can only be achieved with the appropriate management and organizational structure and with a program to modernize business practices (Hardy, 2010). Competent management and leadership can mitigate unforeseen and unanticipated problems that may arise in complex IT projects.

According to Donovan (2016), risk management tools are very important to the federal government when it comes to protecting its IT systems. One such risk management tool, the Hazard and Vulnerability assessment (HVA) is a comprehensive tool that enhances the security of the federal government. Furthermore, it is important that management drive risk management efforts and the implementation of risk management tools to ensure that employees realize their importance. (Laisasikorn and Rompho, 2014). In addition, the precedent established by management can also diminish the risks that may affect both cost and schedule performance. This point further reinforces the importance of establishing a risk management culture within an organization that effectively originates among leadership and filters below to subordinates. Conversely, not having executives who support the importance of risk management in an organization can potentially precipitate consistent IT project failure.

Furthermore, some behavioral factors must be regarded as risk factors; for instance, this applies to the Persaud Paradox, which describes when a false sense of confidence arises because peers all apply the same approach (Mefford et al., 2017). A *group think* mentality within an organization can be highly detrimental, which is why IT project success should be interpreted in terms of both quantitative measures and qualitative factors. Junior and Carvalho (2013) demonstrate that several factors can contribute to project performance. These factors include schedules (timeframes and deadlines), costs, scarcity of resources, inadequate abilities, competencies, etc. They also determined that care with uncertainties, individual knowledge of the business, and utilization of processes, techniques, and tools exert a significant impact on project success (Junior and Carvalho, 2013). Organizations tend to evade the unknown rather than applying the knowledge accumulated over time to resolve challenges. By adopting this previously established internal business knowledge to tackle uncertainties, organizations can overcome IT project difficulties and succeed. Junior and Carvalho (2013) also illustrate that when projects have a risk manager, they experience greater success. These results indicate that it is essential for project managers to invest in effective management practices to improve project outcomes. In many ways, highly ineffective project management practices can be even more detrimental than applying minimal or no robust project management practices. The risk of continuing to deplete time and money through ineffective management practices can cause an organization to become gradually less profitable and perhaps become incapable of conducting business.

Furthermore, Junior and Carvalho (2013) also determine that *simulation of timeframes and costs,* which are more quantitatively based factors, did not significantly

affect reliability; this suggests that assessing whether a project will ultimately be successful must extend beyond both cost and schedule considerations. In his research, Hosford (2017) demonstrates that competent and experienced project managers contribute positively to the success of projects. The competence, professional experience, education, and IT experience of project managers should be examined and measured as a component of the IT project monitoring process to enhance project implementation effectiveness. Bloom (2014) also notes that the time required for project preparations should be accounted for. Complex projects naturally require more time to prepare, but if less complex projects require more time than usual, this could indicate resource shortages and a lack of commitment among project personnel; lower preparation costs and shorter preparation times typically correlate with smaller project size (Bloom, 2014).

As organizations undertake more IT projects, it is also pivotal for preparation time to be standardized based on highly specific project characteristics. Developing an IT project performance prediction model (as proposed in this research paper) and a supporting framework is an effective means of preparing for a project based on its specific characteristics. Complex IT projects naturally require more preparation and therefore more robust prediction models and frameworks relative to less complex projects; however, if preparation extends beyond these well-established models and frameworks, that indicates that a crucial factor has been omitted during the preparation phase of the IT project. Hussein (2013) identifies four statistically-relevant risk factors which could negatively impact project success if they occur during the initiation phases of a project. These factors include incomplete criteria, diversity, unrealistic targets, and ambiguous or

non-measurable criteria (Hussein, 2013). Rather than focusing on the characteristics of the project itself, these factors relate more to the management approach of the project.

Incomplete criteria can be detrimental because this prevents project managers from executing common goals and objectives for a product that will effectively satisfy the needs of the customer; instead, during project implementation, substantial unnecessary time, energy, and resources may be expended completing tasks that will ultimately not be fruitful. Diversity is also intriguing because it is often overlooked as a factor that contributes to project success. As described in the study of Hussein (2013), diversity encompasses variation in stakeholder interests, power, and influence. The amount of interest among individuals who hold stakes in projects can be both beneficial and detrimental to projects. While the *mass backing* of a project can result in adequate resource availability, conflicting interests of stakeholders can overcomplicate project execution and ultimately precipitate project failure. Unrealistic targets during IT project implementation, in terms of completion dates, budget, and performance, may be detrimental as well. Unfounded targets can cause organizations to both over- and under-plan for projects, which may result in squandering money and resources that could have otherwise been allocated more effectively. Lastly, applying immeasurable criteria can hinder organizations from accurately tracking progress against baselines and organizational goals. For an organization to target and mitigate potential IT performance risks, it must first understand the potential sources of these IT performance risks; immeasurable criteria render this nearly impossible.

Based on their research and data analysis, Beleiu, Crisan, and Nistor (2013) conclude that the five factors which contribute the most to project success are as follows: 1) clearly

defined goal and directions, 2) competent project team members, 3) clearly defined roles and responsibilities, 4) communication and consultation with stakeholders, and 5) compliance with the planned budget, timeframe, and performance criteria. These factors incorporate both categorical, or qualitative, factors (competent project team members, clearly defined roles and responsibilities, and communication and consultation with stakeholders) and continuous, or quantitative, factors; this further confirms the importance of considering both types of factors when measuring and monitoring the success of IT projects. Developing criteria to measure qualitative factors is challenging; however, this should not hinder organizations from exploring creative means of measuring these factors to promote the success of projects.

Conversely, this study demonstrates that owners' and sponsors' involvement in projects and the provision of timely data to key players exert the lowest impact on project success (Beleiu, Crisan, and Nistor, 2013). Based on these findings, more qualitative and management-based factors are accounted for relative to quantitative and project characteristic-based factors. Additionally, the factors are categorized into primary (leading) and secondary (resulting) factors. Although all factors were determined to be statistically relevant, the qualitative factors had more impact than the quantitative factors. For example, defining roles and responsibilities and communication protocol during project preparation can undoubtedly result in budget and timeframe compliance during the initial stages of project execution. Furthermore, sustained effects such as synergistic relationships and a collaborative culture can be established within the organization and provide a foundation that continuously promotes successful IT projects.

Both quantitative and qualitative factors contribute significantly to project success. Quantitative factors such as cost, scheduling, and performance are generally easy to measure and monitor. While qualitative factors such as communication, culture, and competence are not as seamless to measure, they can exert an equally formidable impact on projects. Successfully monitoring both types of factors can ultimately facilitate successful project performance.

2.3 Management Methods

The IT project management methods utilized during the execution of an IT project can be pivotal in determining the types of issues, risks and success factors which ultimately determine project success or failure. This section examines the differences, advantages, and disadvantages of key project management methods that are applied throughout project execution.

Pederson (2013) suggests that under the agile model, user involvement is a component of the repeated iteration of the process to continually improve the product based on feedback and updated requirements from the client. Pederson (2013) implies that projects which are more dynamic in nature and are more sensitive to changing requirements and desires of both stakeholders and environments are more likely to be successful when the agile method of IT project management is applied. However, due to the inherent changes that occur during projects that apply this method, managers also risk experiencing uncontrolled growth in the project's scope and implementing unnecessary changes during the project.

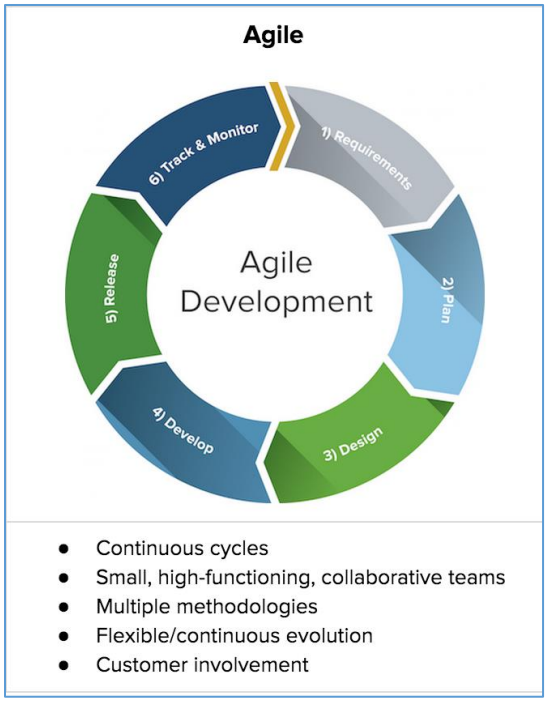


Figure 1:Agile Development Model

Adapted from "Everything You Need to Know About Agile Project Management." Retrieved October 29, 2018. from https://www.smartsheet.com/everything-you-need-to-know-about-agile-project-management.

Conversely, the waterfall model is recommended for highly regulated environments or industries (Pederson, 2013). Since IT projects in these industries are typically more rigid and *gate process*-oriented, it is regarded as more advantageous to adopt a step-by-step project management methodology to ensure that steps are not omitted, which could ultimately result in compliance issues and industry violations. These projects also tend to have relatively lengthier delivery times relative to agile products; therefore, they are not as time-sensitive in terms of project and product delivery. In recent years, as IT project technologies have become more complex, with shorter project durations and stricter cost and schedule restrictions, increasingly more organizations have utilized more agility-oriented methods. However, organizations should not succumb to the popularity of these methods but should instead examine the characteristics of the organization and the nature

of the IT project to determine which method will produce optimal results, whether this may be a strictly agile method, a strictly waterfall method, or a customized combination of these.

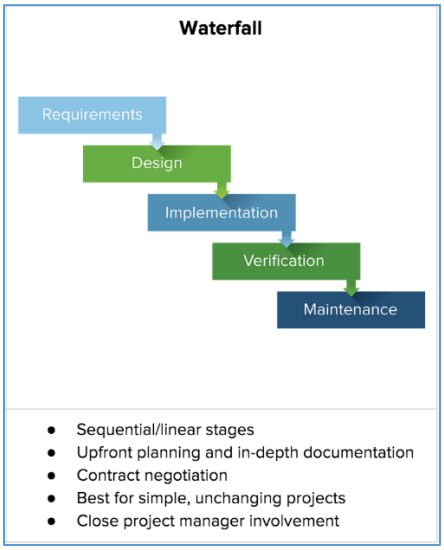


Figure 2:Waterfall Model

Adapted from "Everything You Need to Know About Agile Project Management." Retrieved October 29, 2018. from https://www.smartsheet.com/everything-you-need-to-know-about-agile-project-management.

Chow (2016)'s study, which examines the relationship between project management methods and project success, schedule, and quality, concludes that the relationships between these do not differ based on software development methods in medium-to-high-complexity projects. Chow (2016) also elucidates the various elements that affect projects based on complexity level. These can be interpreted through three perspectives: 1) Projects that are long-term and less complex are typically less sensitive to schedule changes and cost changes; 2) conversely, short-term, less complex projects tend to be relatively more schedule- and cost-sensitive relative to the former; 3) however, more complex projects with shorter durations are the most sensitive to cost and schedule changes and should be handled with the greatest caution.

Applying the appropriate methodology is crucial to ensure that IT projects are released successfully, on-time, and within budget (Geambasu et al., 2011). Furthermore, the appropriate methodology can mitigate risks and enable realistic estimations of costs and development. Alshamrani and Bahattab (2015) suggest that the Waterfall IT Project Management Model ensures that design flaws are apparent before the development of a project. Since the waterfall model allows formal design reviews at various stages of the process as a *gate check* mechanism, issues with design are addressed before a project can proceed to the next phase. This model is effective for projects that place high emphasis on quality control and intensive documentation and planning (Alshamrani and Bahattab, 2015). These findings also support the effectiveness of this model for IT projects within highly regulated industries which require meticulousness and compliance. The trade-off of using a waterfall model is apparent in the avoidance of fines, levies, and sometimes loss of life, which can be detrimental for both the reputation and the bottom line of an organization.

Furthermore, the spiral IT project management model focuses on risk assessment and minimization; it entails breaking projects into smaller segments, which allows for seamless change and continuous improvement throughout the project development cycle (Alshamrani and Bahattab, 2015). Lastly, the iterative and incremental model integrates elements of the waterfall model and incrementally delivers products to the customer. Both the spiral and iterative or incremental models are subsets of the Agile IT Project Management Models, which are typically applied to projects that are more sensitive to time constraints and both customer and environmental changes. However, each of these models present unique shortcomings as well. For example, the waterfall method is

typically ineffective when applied to complex projects; customers may be unable to preview projects until later stages, when it may be too late to implement necessary changes. The spiral IT project management method requires an extensive investment in and knowledge of risk management and can be both time-consuming and expensive in terms of the time devoted to evaluating risks, planning, and prototyping. For the iterative management model to be effective, the operational requirements of the final system must be completed to plan the incremental stages of the project. Regardless of the IT project management model that is applied, it is essential that the information known about a project is finalized before it is initiated; this is because it is increasingly difficult to execute a project successfully when these details are unveiled later during the project lifecycle.

To effectively execute a project, it is not only important to ensure that the appropriate management model is implemented based on the characteristics of a project; the managerial style that is used during project execution is also important (Kocherla, 2012). In the execution of projects, the managers' focus and style should align with the significant elements of the management model. The waterfall model, which is highly *gate*- and linear process-focused, is most effectively executed by a manager whose leadership style is highly linear, rigid, and process-oriented. Being strict regarding the criteria that projects must satisfy to enter the next phases is a crucial element of the management style to effectively execute this model. More agile models such as iterative, incremental, and spiral models, which are more flexible in nature, require a management style that is more adaptable to constant change and managing simultaneous process flows, which results in incremental delivery of products. Effective managerial styles

which apply these types of models require constant communication and interactions with customers that enable continual product improvement based on customer feedback throughout the lifecycle of the project. In addition, it is equally important for the managers to drive this change into subordinates so that the appropriate mentality that is necessary to execute the IT project management method is promulgated throughout the culture of the organization. The lack of *buy-in* from subordinates can hinder an organization similarly to that of those in management positions.

In addition to unique advantages and disadvantages in applying certain IT management models, there are also elements that impact both agile and non-agile project management models. When discussing *IT management model agnostic* elements that impact projects, Ilincic (2008) states "less frequent changes in requirements for projects, providing appropriate project tools to project team members, and building cultural synergy among the globally distributed team positively affects the outcomes of projects, particularly in terms of time, scope, and customer satisfaction". The author highlights that cultural synergy refers to both the culture related to language and customs and to the culture within an organization as well. The cultural implications that influence the effectiveness of an IT management model must also be accounted for. Organizations that have primarily implemented and executed projects that traditionally required non-agile methods and must now use agile methods to execute newly explored projects must experience a cultural shift before they can effectively implement new IT management models. It is pivotal for a robust cultural change and implementation plan to be executed prior to the execution plan for the actual IT project. Regardless of how thorough and

robust the IT implementation plan is, if the cultural change does not occur, then ineffectiveness is highly probable.

2.4 IT Project Types

Depending upon the type of IT project being implemented and executed, an array of benefits and challenges ultimately have implications for project success. This section of the literature review explores the details related to cloud, hardware, and software-based IT projects.

Cloud solutions are widely adopted in enterprises because they can reduce capital expenditures and transform this into operational costs, while simultaneously offering a platform for modular and reusable web services throughout an organization (Andrikopoulos et al., 2012). Cloud IT projects also enable the automated deployment of applications while maintaining control of IT management costs. Suo (2013) notes that in relation to implementing cloud solutions in organizations, critical success factors should be categorized into strategic and tactical factors (Suo, 2013). Project schedules and plans, as Suo (2013) posits, should be categorized as strategic factors of cloud implementation, while personnel and communication should be categorized as tactical factors of communication. The potential risks that may be experienced during cloud implementation can be categorized in this manner as well. In contrast with the more seamlessly quantifiable elements such as budget and schedule, the more challenging elements of cloud implementation, such as personnel experience levels and management communication, should not be overlooked because all factors ultimately impact cloud project success. Twill (2012) contends that cloud facilitates improved communication with customers, higher profits, low investment, and operational costs. Furthermore, Twill

(2012) also suggests that some of the most important considerations of management when determining whether to venture into cloud adoption include compatibility, top management support, and relative advantage.

It is important that organizations develop mechanisms to measure these crucial factors before and after project implementation, while continually monitoring these during the cloud implementation process. Khan (2016) demonstrated that the primary elements of strategies to adopt cloud include security and cost. Security is a pivotal element because of the uniqueness of cloud and the environments in which it is implemented. Historically, since most applications were stored internally, security was less of a concern among organizations because it was relatively easy to secure applications using *in-house* security measures. Conversely, since cloud-based applications typically reside in an environment that is external to organizations, they have much greater security implications; if these are not addressed early, these can cause both cost and schedule issues during implementation and after deployment.

There is a strong relationship between managers' decisions to adopt cloud and the perceived cost-effectiveness of cloud computing and cloud computing need, reliability, and security effectiveness (Ross, 2010). These findings further emphasize the potential complexity of cloud adoption decisions. In addition, the data collection efforts and tools to support these decisions by managers must be equally complex and account for all the factors that ultimately impact successful cloud implementation. Specifically, an IT project performance prediction model can be an effective tool that helps IT managers determine key decisions before implementing such complex projects. According to Mink (2015), the lack of a common decision framework within the federal government

regarding whether legacy IT systems should migrate to the cloud presents another cloud-related challenge. As previously stated, standardization within the federal government's decision-making process is critical because it allows organizations to apply identical decision-making criteria.

When organizations consider implementing the cloud, they must also make decisions regarding legacy systems. Many issues related to system compatibility and data transfer can pose problems for organizations transitioning to the cloud. Project cost and schedule implications are also associated with this decision-making process. Therefore, it is essential to implement a robust framework to support legacy IT system migration, particularly within the federal government, which is not generally on the cutting edge of technological advancements. Specifically related to hardware considerations in IT projects, Ryan (2010) suggests that standard measures evaluating IT hardware should be implemented within organizations to render hardware decisions more transparent. Implementing a standardized method for agencies within the federal government to predict project success can render IT project decisions both transparent and easily comprehensible.

The success of software projects beyond the commercial and federal sectors in academia exhibits trends that highlight that participants within the IT project must regard communication, collaboration, and problem-solving as key elements in successful software project implementation. One study emphasizes the importance of the social process in relation to software development and implementation projects (Guinan and Sawyer, 1998). This further confirms that shared norms among teams involved in software project implementation are pivotal in their impact on IT project performance.

The requirements that are provided and managed throughout the development process also contribute to the success of software projects (Hussain, Mkpojiogu, and Kamal, 2016). Initially, requirements should be both effectively communicated by the customer and well understood by the individuals involved in the software project's implementation. Secondly, these requirements should be closely monitored throughout project development and implementation to prevent scope creep, which may result in cost overruns, schedule overruns, and low quality of the product delivered.

2.5 Performance Prediction Models and Methods

As technology has advanced and IT projects have become more complex in recent years, many efforts have aimed to develop tools, methodologies, frameworks, and models that can provide insight into the potential success or failure of various projects based on a myriad of inputs and variables in many industries and organizations. This section examines a variety of these, which are quantitively, qualitatively, and hybrid-based; it also addresses their effectiveness relative to the data collection platform that currently exists within the federal government.

PoCock, Liu, and Tang (1997) explore the relationship between project designers and builder interactions in terms of project performance; they present a predictive model for project performance based on levels of interaction, which they measure through a questionnaire that surveys managers about communication points and stakeholders. The performance indicators encompass both quantitative and qualitative metrics, including cost growth, schedule growth, number of modifications, modifications due to design deficiencies, and user satisfaction. This model produces a Degree of Interaction (DOI) score; this indicates the extent of interaction among stakeholders; project performance is

then measured relative to this. This demonstrated a modest correlation between a DOI score above .4 and a decline in project performance; it could therefore not be accurately predicted. In addition, regarding the data collection platform that exists for a federal IT project, levels of interaction are not currently assessed.

Lee et.al (2004) examines the prospect of predicting cost and schedule performance of construction projects using discriminant function analysis (DFA). This data analysis uses Fisher's coefficients to quantify the statistical impact of construction project-related (excluding personnel-related variables such as manager experience) variables. When analyzing these results, projects were divided into quartiles based on actual cost and schedule and then predicted based on these quartiles; this yielded a pure chance accuracy of 25%. Despite the relatively limited accuracy of this model and although the variables and information differ across IT projects and construction projects, this model was examined because it also included a suggested improvement mechanism to increase the potential of project success.

Lipke et.al (2009) examines the use of earned value management and earned schedule and statistical methods to predict both cost and schedule indices for IT projects. Their study concludes that this method is broadly applicable despite the type of project and industry where cost and schedule forecasting are performed. The study examined 12 projects that were characterized as having low to moderate costs and moderate project duration. The results indicate that reliable estimates with 90% confidence were applied to the data when the project was 60% complete (Lipke et. al, 2009*).* This method also exclusively examines project-related inputs as opposed to personnel (such as project manager experience and manpower) and management-related inputs (such as software

development methodology). However, in lieu of its applicability to a wider variety of project types (since it considers cost and schedule-related metrics that are common across most projects), this method lacks specificity to IT projects. Furthermore, as indicated by the previously mentioned reliability estimates, this method becomes more useful and accurate as the project continues; this diminishes its effectiveness as an actual project performance prediction model to be used before the actual start and implementation of IT projects.

In related research, implementing change point analysis has been considered to further enhance the accuracy and effectiveness of the EVM model (Mortaji et. al, 2015). Change point analysis refers to when project managers observe a shift in either a cost- or schedule-related element while monitoring a project; based on predefined thresholds, they then implement corrective actions to either course-correct the factor that has shifted out of tolerance or monitor this element with more rigor. Change point analysis within EVM allows managers to be more flexible in adapting to changes during projects.

Moy (2016) utilizes a subset of the dataset that was examined for this praxis and specifically investigates cost variance, schedule variance, and actual cost of software projects within the federal government and their potential impact on software project success. Moy (2016) also examines the quantitative correlations between earned value management (EVM), cost variance (CV), schedule variance (SV), and actual cost (AC); they conclude that SV has a relationship with IT program success, but that CV and AC do not correlate with IT program success. In addition to applying the EVM method, which becomes more accurate as more data becomes available while the IT project is executed, Moy (2016) also investigates potential outcomes (cost variance and schedule variances)

as predictors of project success as opposed to inputs or indicators of project success. Moy (2016) also considers expanding the model to encompass additional variables in her future research to potentially improve the accuracy of her model. To supplement the foundational work and analysis conducted by Moy (2016), this research examines additional predictors and their effects on cost and schedule variances as measurable outcomes of project success.

Adoko's (2016) research investigates the following prediction factors: system performance, technology maturity, schedule, reliability, and risk. Of these factors, schedule and reliability are identified as the key factors that determine whether a project experienced cost overrun. The final regression model predictability accuracy is 62.1% in terms of accurately predicting projects with cost overruns and 83.3% in terms of accurately predicting projects with no cost overruns. The predictability of this model is relatively accurate; however, as previously mentioned, models do not address the effects of personnel-related prediction factors involved in IT project management (i.e., project management experience and software lifecycle development cost) on project performance. Analyzing the impacts of these factors on project performance is pivotal because this directly relates to the information that the federal government currently collects for its IT projects. This model also only aims to predict cost overruns and not success factors related to schedule overruns. Although schedule is accounted for as a prediction factor, the impact of the other prediction factors (system performance, technology maturity, reliability, and risk) on schedule overruns is not discussed in this research.

Darnell (2015) examines explicit and implicit measures of coordination as predictors of project success. Explicit coordination occurs when coordination mechanisms such as plans and rules, objects and representations, defined roles (e.g., project manager), and prescribed routines are overtly used to coordinate work. Implicit coordination is a component of coordination effectiveness wherein coordination occurs when interdependent groups and individuals anticipate the actions and requirements of one another and adjust their behaviors appropriately without engaging in purposeful communication (Darnell, 2015). A survey with three questions was administered to determine the level of explicit coordination, and a survey with eight questions was administered to determine the level of implicit coordination using the Likert scale. Success was measured through interview responses to questions based on the quality, scope, timeliness, and cost of the project.

Based on the results and regression analysis, Darnell (2015) determined that higher scores for explicit coordination were associated with higher scores for project success, and higher scores for implicit coordination were moderately associated with higher scores for project success. Although this research provides adequate evidence regarding predictors of project success, it does not incorporate these findings into a reliable model that can be used to predict project success. A model that could be used by IT project managers to provide insights into the potential failure or success of projects using the discovered information would be highly useful, particularly for federal IT projects.

In their research, Gemino, Sauer, and Reich (2010) compare classification tree, regression, and neural network prediction models. Furthermore, they study the effects that continuous and class variables have on these models. The continuous variables

include budget, duration, effort, and project management experience, while the class variables encompass relative size, requirement uncertainty, requirement difficulty, and technical complexity. From the study, they determine that the level of requirement uncertainty was the most influential variable (Gemino, Sauer, and Reich, 2010). This suggests that a qualitative measure is the most likely to affect the outcome of a project and that IT managers should provide more pre-initial phase resources and devote effort to the development of requirements long before IT project implementation begins if they wish to influence the success of IT projects. Furthermore, the study investigates the accuracy of the various prediction techniques in relation to a group of projects that had underperformed and a group that had performed. Outputs of the prediction model yielded a percentage of the model performing as opposed to not performing; based on this comparison, a misclassification rate was produced which statistically indicated whether the model incorrectly classified the project as underperforming or performing.

Based on the results of the validation dataset, the misclassification rates of the regression model, neural network model, and classification tree model were 45%, 39%, and 35%, respectively (Gemino, Sauer, and Reich, 2010). As a basis for model comparison, the variables and predictors in this study along with the models, specifically the regression model, closely align with the elements examined in this research paper in relation to federal IT projects. A more accurate prediction model can be developed for use by IT project managers.

2.6 Research Opportunities in IT Project Management

As previously mentioned, IT projects within the federal government have become increasingly complex and more expensive; additionally, continuously failing IT projects

have cost the federal government millions of dollars each year (Government Accountability Office, 2013). If project management within the federal government continues to operate in this manner, then the amount of wasted money, time, and resources will continue to increase. This dilemma has necessitated more effective project management methods and tools to promote the efficiency of IT projects within the federal government. A prediction tool should be implemented across all agencies that use federal government IT project data to assist managers. This tool would establish a standardized method within the federal government that leverages historical IT project data to predict IT project performance that can be used prior to and during the initial stages of a project; this would allow organizations to plan accordingly and strategically reallocate resources before and during a project to minimize downtime and the number of unused resources. Furthermore, opportunities that have not been addressed in this research, specifically those related to federal government IT project predictors, should be explored as well. One such topic that will be explored in this study is the specific impact of management methods on federal government IT projects. Another area of research that has not been fully addressed and will be explored further in this research involves the effects of the categorical predictors collected by the federal government on cost and schedule variance.

2.7 Summary and Conclusion

In relation to IT project management and the various factors that influence the success of IT projects, a plethora of inputs and outputs, both internal to the project and in the external environment, are influential. It is crucial to consider a variety of measures and metrics, both quantitative and qualitative, that can affect project outcomes. As overwhelming as it may appear, understanding the nature of the IT project when

considering the various factors allows project managers to accurately determine which factors should be scrutinized. It also pivotal to consider the intangibles of IT projects, such as stakeholder interactions and communications that are not as seamlessly measurable as those from the research. IT projects are range immensely in terms of complexity, cost, technology, and industry and a *one-size fits-all* approach to managing them can be detrimental. IT project managers should instead perform due diligence and research existing methods and either use those based on the specifics of the IT project, use a combination or hybrid approach, or develop a unique method specific to the nuances of the project to achieve optimal performance. Although no method, factor, model, or process is perfect in terms of managing a project and predicting the success of the project, more effective tools and methods applied before and during IT project implementation certain maximizes the probability of a project's success relative to ineffective methods or no methods or tools.

Chapter 3—Methodology

3.1 Introduction

This research explores the relationship between predictors that the federal government currently tracks and the impact of these predictors on cost and schedule variances. Based on research, the potential predictors were identified in the dataset. Once these predictors were identified, they were categorized as either continuous predictors or categorical predictors. Continuous predictors are numeric variables that have an infinite number of values between any two values. In this research, the continuous predictors were demonstrated to include *planned duration of project* and *planned project lifecycle cost*. Categorical predictors typically encompass a limited and fixed number of possible values. The categorical predictors from the IT Federal dashboard data were identified as "project manager experience," "SDLC methodology," and "product release frequency." Once categorized, the predictors were analysed using statistical methods and hypothesis testing to achieve research objective 1 (see Chapter 1) and to determine how each impacted both cost and schedule variance. The categorical predictors were evaluated using boxplots, while the continuous predictors were evaluated using linear regression. To achieve research objectives 2 and 3, the effectiveness of each predictor in the prediction model for both cost and schedule variance was then evaluated using P-values. Using step-wise regression at an alpha value of .1, each predictor was evaluated for statistical significance. An alpha value of .1 (confidence level of 90%) was deemed to be acceptable during the data analysis due to the relatively smaller sample size and the fact that the specific hypotheses in this research have not previously been tested. It should also be noted that other prediction models have used this confidence interval as well to

evaluate effectiveness. Debray et.al (2012) in their research to develop a clinical prediction model and Christoffersen and Goncalves (2005) in their research to evaluate financial risks models both use 90% confidence levels. The alpha value of .1 provides an acceptable baseline for the statistical significance level which has the potential to be improved upon as more data becomes available.

Predictors with a P-value of less than .1 were deemed significant and were used in the final prediction model. The predictor variables hypotheses tests (see Chapter 1) were also evaluated using these results; P-values of over .1 resulted in failure to accept the null hypotheses. Conversely, P-values of less than .1 resulted in the rejection of the null hypotheses. The prediction model was then executed to obtain prediction results for cost and schedule variance, which were compared with actual results. The accuracy of the prediction model was obtained by observing what percentage of predicted values were within 10% of the actual values. Finally, to achieve research objective 4 (see Chapter 1), the misclassification rates, or the rates at which projects are incorrectly classified in the model, of the proposed model and the GRS model were compared.

3.2 Data

3.2.1 Data Sources and Analysis Tools

The data used during the research originated from publicly available federal government IT data from agencies within the U.S. government; this data was derived from the following websites:

- **https://www.itdashboard.gov/drupal/archived-data**
- **https://www.itdashboard.gov/drupal/e-gov-act-reports**

The *2017 projects* folder data from the federal IT dashboard downloadable data is provided in Excel format, which is easily viewable. This folder was used because it was the most recent and consisted of project data from projects initiated between 1992 and 2017. The project durations range between 2 and 11,321 days. Costs of the projects in the dataset range from $0 to 3.95B. In addition to the predictors, other information captured in the dataset includes government identifiers such as agency name, investment title, project name, project outcome, and unique investment identifier. The data analysis tools used include Microsoft Excel and Minitab 18.

3.2.2 Data Inclusion and Outliers

Data included in this research satisfied the following criteria:

- Data points included contained all values for cost variance, schedule variance, PM experience level, SDLC methodology, planned project duration) and planned project lifecycle cost;

- Data points included were positive values for both cost and schedule variance. Negative values of cost and schedule variance were excluded from this study to facilitate examination of the linear relationships of the predictor variables with cost and schedule variance;

- Fifteen outliers were eliminated based on Minitab analysis of large residuals and unusual X's, which yielded N = 78 projects studied.

3.3 Analysis Methods

When evaluating the data, several analysis methods were used to verify each hypothesis as it related to the predictor and response variables. The following Table 1 summarizes the analysis methods that were used in this research study.

Table 1:Summary of Analysis Methods for Hypothesis Testing

Hypothesis	Response Variable (Y)	Predictor Variables (X)	Predictor Variable Type	Analysis Method
H1	Cost Variance	Project Length (Days)	Continuous	Linear Regression Analysis
H2		Project Lifecycle cost (Dollars)		Linear Regression Analysis
H3		Project Manager Experience (Level)	Categorical	1) Linear Regression Analysis 2) Univariate Analysis: Boxplot 3) Spearman correlation
H4		Type of Software Development Lifecycle Method (Iterative, Waterfall)		1) Linear Regression Analysis 2) Univariate Analysis: Boxplot 3) Spearman correlation 4) 2-Sample T-Test
H5		6 month Frequency of Release		1) Linear Regression Analysis 2) Univariate Analysis: Boxplot 3) Spearman correlation
H6	Schedule Variance	Project Length (Days)	Continuous	Linear Regression Analysis
H7		Project Lifecycle Cost (Dollars)		Linear Regression Analysis
H8		Project Manager Experience (Years)	Categorical	1) Linear Regression Analysis 2) Univariate Analysis: Boxplot 3) Spearman correlation
H9		Type of Software Development Lifecycle Method (Iterative, Waterfall)		1) Linear Regression Analysis 2) Univariate Analysis: Boxplot 3) Spearman correlation 4) 2-Sample T-Test
H10		6 month Frequency of Release		1) Linear Regression Analysis 2) Univariate Analysis:Boxplot 3) Spearman correlation
H11	Schedule Cost Variance	Statistically Significant Predictors	Categorical Continuous	Accuracy Check
H12				Misclassification Rate

Note: Predictor variables validated at P<.1 for statistical tests

3.3.1 Response Variables

The response variables examined include the cost and schedule variance percentages. Despite the raw data consisting of both cost variance in dollar amounts and schedule variance in days, the percentages of both response variables were selected for consistency in comparison.

Linear Regression Analysis: Continuous Predictors

The linear regression data analysis approach was applied to address the relationships of continuous predictor variables, *planned project duration,* and *planned project lifecycle cost* with cost and schedule variance. However, before this analysis method was employed, the required assumptions were assessed as illustrated in Chapter 4.

3.3.2 Predictor Variables

After studying the provided data from the *IT Dashboard* and the literature related to predictor variables, we identified five predictor variables that we were able to extract and closely analyse:

1) *Planned project duration (continuous predictor variable)*

   The following basic calculations were performed to transition some of the predictor variables into a *final* state to be analyzed based on data provided in the downloaded spreadsheet. The planned project duration was calculated based on the following:

Actual project duration = Length of start date to completion date

***Planned project duration** = Actual project duration - schedule variance (in days)*

Equation 3-1

2) *Planned project lifecycle cost (continuous predictor variable)*

The planned project lifecycle cost was also calculated in a similar manner as follows:

Planned project lifecycle cost *= Project lifecycle cost - cost variance (in dollars)*

Equation 3-2

3) *Project manager experience (categorical predictor variable)*

Project manager experience, which is a categorical predictor variable, was based on a scoring value from 1 to 9; 1 corresponds with the highest level of experience, and 9 corresponds with the lowest level of experience. The following Table 2 explains each project management level in greater detail:

Table 2: PM Level Definitions

PM Level	Definition
1	Individual possesses a Senior Defense Acquisition Workforce Improvement Act (DAWIA) Level 3 Certfication
2	Individual possesses a Senior Defense Acquisition Workforce Improvement Act (DAWIA) Level 2 Certfication
3	Individual possesses a Senior Defense Acquisition Workforce Improvement Act (DAWIA) Level 1 Certfication
4	Individual possesses other certification with 4 or more years PM experience (within the last five years)
5	Individual possesses other certification with between 2 and 4 years PM experience (within the last five years)
6	Individual possesss other certification with less than two years PM experience (within the last five years)
7	Individual possesses no certification, but with 4 or more years PM experience (within the last five years)
8	Individual possesses no certification, but with between 2 and 4 years PM experience (within the last five years)
9	Individual possesses no certification, but with less than two years PM experience (within the last five years)

4) *Type of software development lifecycle methodology (categorical predictor variable)*

Type of SDLC methodology is a categorical predictor variable that was separated into two categories, namely 1) *waterfall* and 2) *iterative*. These terms are further explained in chapters 1 and 2.

5) *Six-month frequency of release (categorical predictor variable)*

The final predictor variable observed whether a project implemented a six-month release schedule. Project releases are periodic project reviews or demos released to the customer which demonstrate development progress every six months (in this case). The six-month frequency release variable is binary: It is coded as 1 if and only if the project has a six-month release frequency.

To ensure that a consistent approach was applied to analyze the data with the same attributes, the data points examined were strictly limited to the points which had data populated for all five predictor variables and the two response variables mentioned above.

3.3.3 Model Development

The model was developed by first inputting the data into a Minitab worksheet. Before transferring the data to Minitab, the cost variance values were transformed using Johnson transformation, and the schedule variance values were transformed using natural log transformation. Johnson transformation was chosen for cost variance because given the original data distribution, this provided the optimal transformation to normal distribution for the data. Contrarily, natural log transformation was chosen to transform the schedule variance data as this provided the optimal transformation to normal distribution for the data. These transformations are discussed in more detail in section 4.4. In the Minitab

worksheet, values for each of the predictor and response variables were placed in separate columns. The response variables included the values for the Johnson transformation of cost variance and the values for the natural log of the schedule variance. The predictor variables included planned project duration, planned lifecycle cost, SDLC methodology, PM experience level, and product release frequency. After applying the *fit regression* model feature in Minitab, the variable columns were selected to be either *responses, categorical predictors*, or *continuous predictors*. When developing the model to predict the schedule variance based on the data, the response columns selected included the *Natural Log of the Schedule Variance* column and the *Johnson Transformation of Cost Variance* column. The categorical predictor columns selected included the *Planned Project Duration* and *Planned Lifecycle Cost* columns. Lastly, the categorical columns selected included the *SDLC Methodology, PM Experience Level, and Product Release Frequency* columns. Once this data was properly segmented, the step-wise regression feature with an alpha value of 0.1 was selected. Selecting a step-wise regression with this alpha value eliminated any predictors with a P-value over .1, which yielded only the statistically relevant predictors in the model.

3.3.4 Model Accuracy and Validation

Once the final model was developed using exclusively the predictors that were statistically significant, the model was then used to predict cost and schedule variance results using the *Predict* feature in Minitab. Once the *Predict* feature has been selected, the categorical and continuous predictor columns are individually selected as inputs to the prediction model. After running the prediction model, the prediction values for the cost and schedule variances are output to the Minitab worksheet as *PFITs* values. The *PFITS*

values are then converted from transformed values into cost and schedule variance values that can be compared with the actual cost and schedule variance values. The data was then divided into four categories, namely actual cost variance, predicted cost variance, actual schedule variance, and predicted schedule variance. When measuring the accuracy of the model, the percentage difference between the actual and predicted cost variances was compared along with the percentage difference between the actual and predicted schedule variance. Once the percentage differences had been calculated, the accuracy rate of the model was calculated using the number of prediction values within 10% of the actual values:

$$\text{Accuracy Rate} = \frac{Number\ of\ Predictions\ within\ 10\%\ of\ Actual\ Values}{Total\ Number\ of\ Predictions}\ X\ 100$$

Equation 3-3

For comparison purposes, the partial least squares regression method was applied; this used all predictor variables in the model regardless of significance to observe whether results were better relative to the step-wise linear regression method.

Partial least squares (PLS) regression emphasizes developing the predictive model as opposed to being used in explaining responses. To compare the accuracy of the two methods, the predicted values of the response variables from each model (PLS and step-wise regression) were compared to observe which model had the highest accuracy of predicted values within 10% of the actual values. The model with the highest accuracy was then selected to be compared with the GRS model; the details of this are explained in Chapter 2.

3.3.5 Model Overview

During model comparison, the proposed model was compared with the GRS regression model, which applied similar predictors to examine the effects on similar response variables. In the GRS model, the continuous predictor variables that were used included budget, duration, effort, and PM experience. The categorical predictors used, which were rated using a Likert scale of 1 (low) to 5 (high), included relative size, requirement uncertainty, requirement difficulty, and technical complexity.

The performance measures used included budget variance, schedule variance, and scope variance; if a project deviated in terms of one or more of the project performance measures by 10% or more, it was deemed an underperforming project. After the collection of results based on project manager questionnaires, a dataset with 198 (45%) underperforming projects and 242 (55%) performing projects was developed. The predictive variables were input into the data analysis tool used for the GRS model, which is called SAS Enterprise Miner. Within SAS, a framework for analysis called SEMMA (sample, explore, modify, model, and assess) was developed; this is used to transform raw data into a validated predictive model (Gemino, Sauer, and Reich, 2014). For the data analysis of the proposed model in this praxis, the linear regression model analysis features of Minitab were used.

To compare the models in the most consistent manner possible, several assumptions and modifications were applied. The general functionality of the data analysis tools of *SAS* and *Minitab* were deemed comparable and as possessing highly similar data analysis capabilities. For the GRS model, the researcher determined that all predictors, whether they were deemed statistically significant or not, were used in the GRS model since it

was not stated otherwise; conversely, only the predictors that were deemed statistically significant were used in our proposed model. Table 3 presents a comparison of the predictors for each model.

Table 3:Independent Predictors Studied for the GRS and Proposed Model

Predictor Variables studied in the GRS Model

	Predictor Variable
Continuous Variables	Budget ($000 US)
	Duration
	Effort (Person Months)
	PM Experience (Years)
Class Variables	Relative Size (1-low, 5-high)
	Requirements Uncertainty (1-low, 5-high)
	Requirements Difficulty (1-low, 5-high)
	Technical Complexity (1-low, 5-high)

Predictor Variables studied in Proposed Model

	Predictor Variable
Continuous Variables	Planned Project Length
	Planned Project Lifecycle Cost (Budget)
	Project Manager Experience Level
Discrete Variables	Type of Lifecycle Development Project Method
	6 month Release Frequency

Considering the range of 10% variance allowances for the underperforming and performing ranges, an alpha value of .1 (confidence level of 90%) was used during data analysis. It was also presumed that the GRS model merely considered positive variance values for the performance measures because when reading the GRS research, the researcher could not find any contrary evidence. In addition, because the data used for the research in the proposed model was restricted to measuring cost and schedule variances, only the cost and schedule variance performance measures from the GRS model were compared. Within the GRS model, three datasets of the regression prediction model were observed, including the testing dataset, training dataset, and validation dataset.

3.3.6 Model Comparison Procedure

Since the dataset used for the proposed model had a relatively lower number of data points which could not be adequately divided into three sets, only the final validation dataset results of the GRS model were compared with the results of the proposed model. To consistently compare the results of the GRS model with the proposed model, a misclassification rate was calculated for the results of the proposed model and compared

with the misclassification rate of the GRS model. The misclassification rate is the rate at which the prediction model incorrectly classifies a project as either performing or underperforming based on the predicted variances. The misclassification rate was calculated as follows:

$$Misclassification\ Rate = \frac{Incorrectly\ Predicted\ Performance\ Project\ Levels}{Total\ Number\ of\ Projects}$$

Equation 3-4

To effectively illustrate the proposed model, both the actual and predicted variance values with variances of less than 10% were denoted with green, while those with variances of over 10% were denoted with red. The misclassification rates of the proposed model and GRS model were then compared.

Chapter 4—Results

4.1 Introduction

This chapters present the results of the data analysis and hypotheses testing through which the relationships between the predictor and response variables were examined. In subsections 4.2-4.6, results and analysis for the predictors and response variables are presented; the performance prediction model testing and validation results are discussed; and the comparison results of the proposed model and the GRS model are discussed. Section 4.7 discusses the results and the specific implications that they had on each hypothesis presented in the research.

4.2 Analysis Results for Continuous Predictor Variables

As previously mentioned in Chapter 3, the methods of linearization of the Johnson transformation method were compared with that of the logarithmic method and the impact of the continuous predictors of planned project duration and planned project lifecycle cost on both cost and schedule variance. This section presents the results of our analysis.

4.2.1 Effect on Cost Variance and Continuous Predictors

The relationship between the response variable cost variance (after Johnson transformation) and the predictor variable planned project duration was examined using Minitab, through which fitted line plots were generated for the response and predictor variables. The results are illustrated graphically in Figure 3. There is no correlation between the cost variance and the planned project duration based on the S-value of 1.0752 and the R-sq and R-sq (adj) values of 0. It should be noted that the R-sq and R-sq

(adj) values indicate the percentage of variation in the response that is explained by the model. The higher the R-sq and R-sq(adj) values, the better the model fits the data.

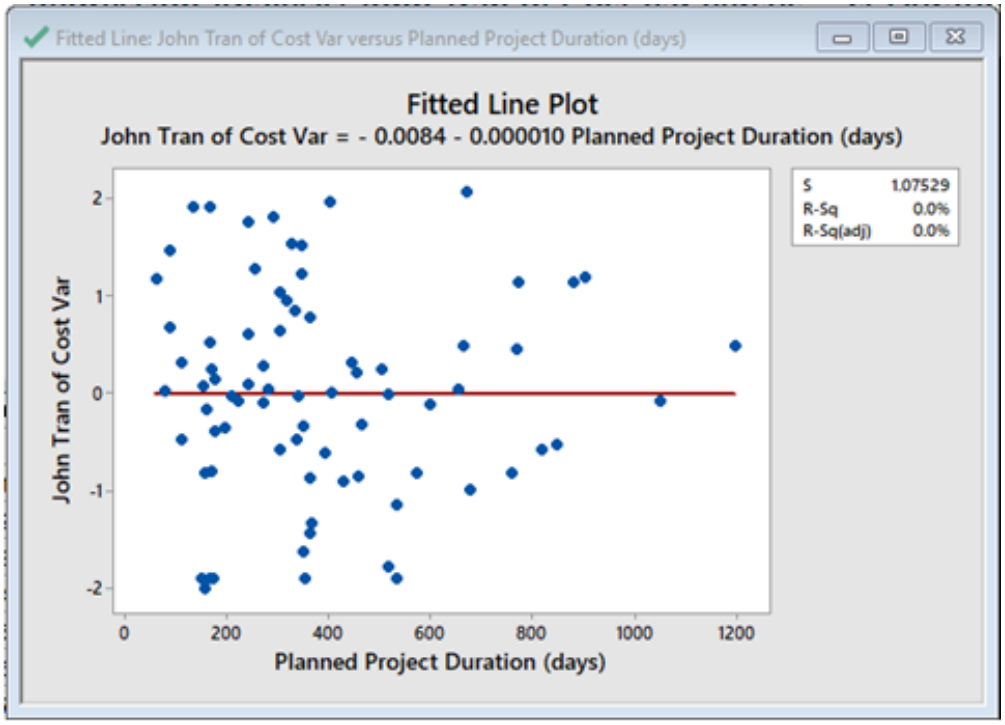


Figure 3:Cost Variance (after Johnson Transformation) vs Planned Project Duration

The relationship between cost variance (after Johnson transformation) and planned project lifecycle cost was examined through Minitab by generating fitted line plots for the response and predictor variables, which are illustrated graphically in Figure 4. There is minimal correlation between cost variance and planned project cost based on the S-value of 1.06735, the R-sq value of 1.5%, and the R-sq (adj) value of 0.2%.

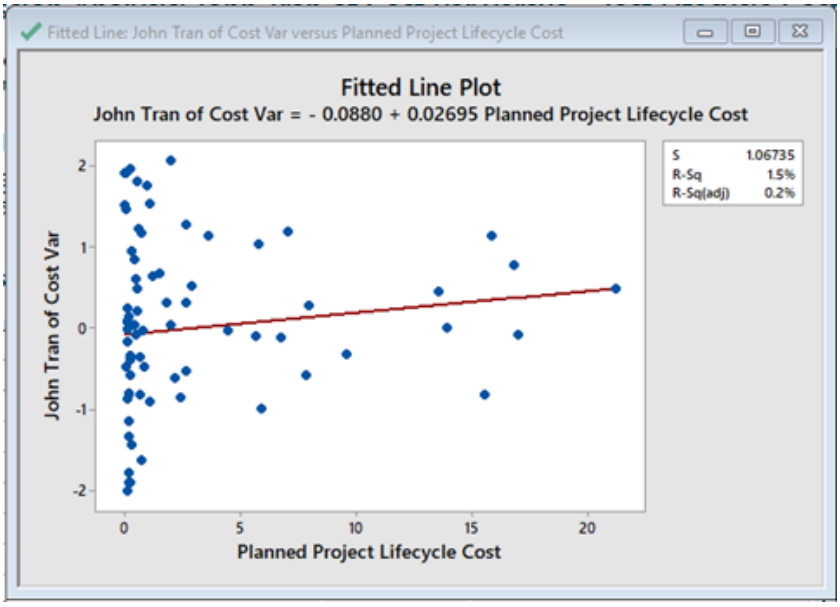


Figure 4:Cost Variance (after Johnson Transformation) vs Planned Project Lifecycle Cost

The relationship between cost variance (after natural log transformation) and planned project duration was examined; this is illustrated graphically in Figure 5. There is no correlation between cost variance and planned project duration based on the S-value of 1.59662 and the R-sq and R-sq (adj) values of 0.

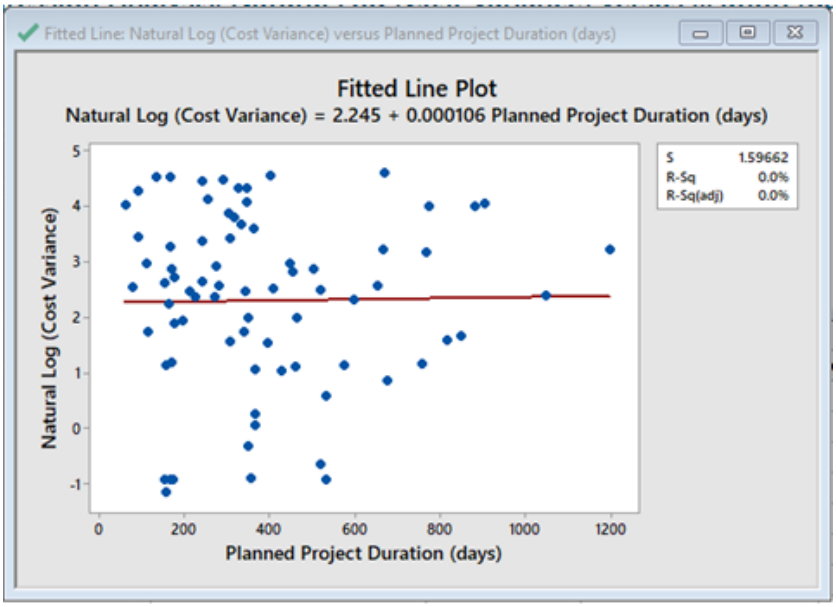


Figure 5: Cost Variance (after Natural Log Transformation) vs Planned Project Duration

The relationship between cost variance (natural log transformation) and planned project lifecycle cost (in millions of dollars) was examined; this is displayed graphically in Figure 6. Although this relationship is slightly greater relative to the Johnson transformation, there is only a minimal correlation between cost variance and planned project lifecycle cost based on the S-value of 1.57545, R-sq value of 2.7%, and R-sq (adj) value of 1.4%.

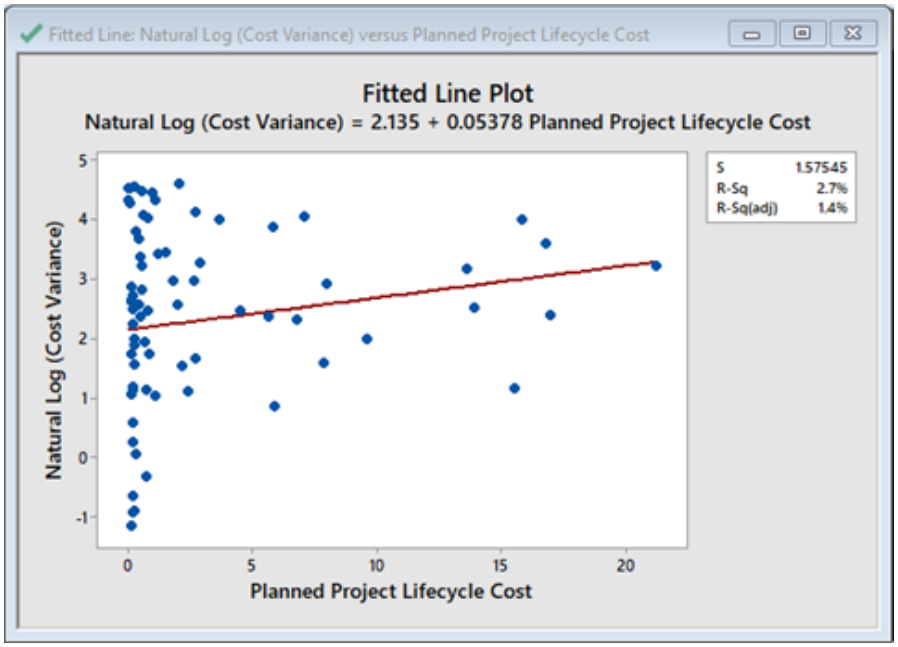


Figure 6: Cost Variance (after Natural Log Transformation) vs Planned Project Lifecycle Cost

Below is a summary of the cost variance correlation statistics:

Table 4: Summary of Cost Variance Correlations Statistics

Response Variable	Predictor Variable	S-value	R-value	R(sq) value
Cost Variance (After Johnson Transformation)	Planned Project Duration	1.0752	0.0%	0.0%
	Planned Project Lifecycle Cost	1.06735	1.5%	0.2%
Cost Variance (After Natural Log Transformation)	Planned Project Duration	1.59662	0.0%	0.0%
	Planned Project Lifecycle Cost	1.57545	2.7%	1.4%

4.2.2 Effect on Schedule Variance

Figure 7 displays a negative relationship between schedule variance (after Johnson transformation) and planned project duration. As the schedule variance (after Johnson transformation) decreases, the planned project duration increases. The calculated statistical values include the S-value of .917640, R-sq of 12.3%, and R-sq (adj) of 11.1%.

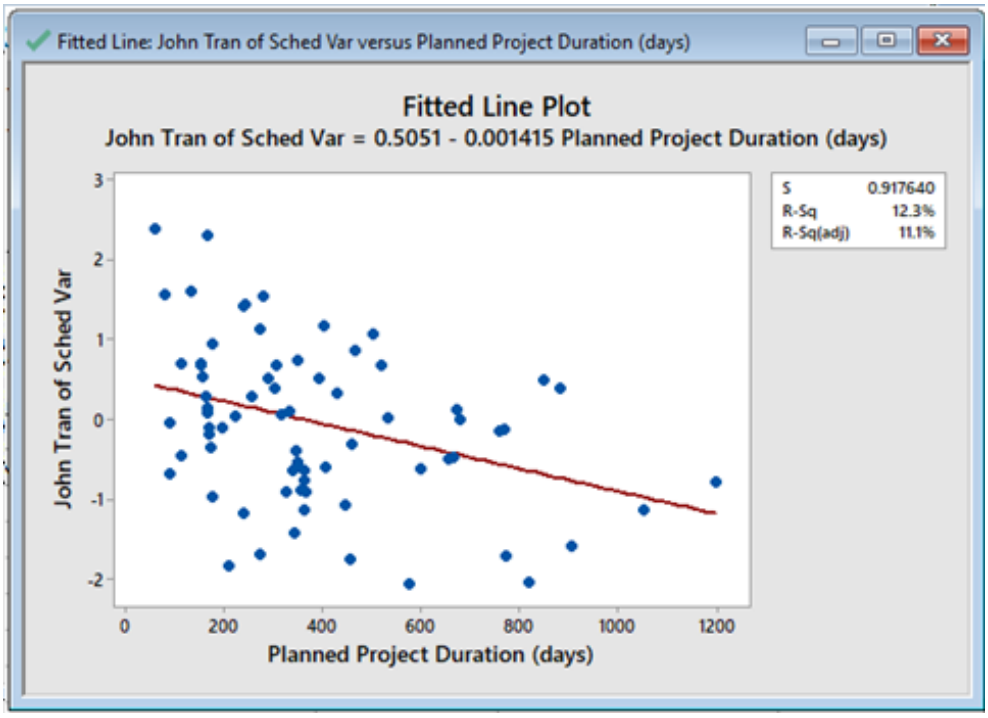


Figure 7: Schedule Variance (after Johnson Transformation) vs Planned Project Duration

The graph in Figure 8 indicates a slightly negative relationship; the schedule variance (Johnson transformation) generally decreases as the planned project cost (in millions of dollars) increases. The calculated statistical values include the S-value of 0.946272, R-sq of 6.7%, and R-sq (adj) of 5.5 %.

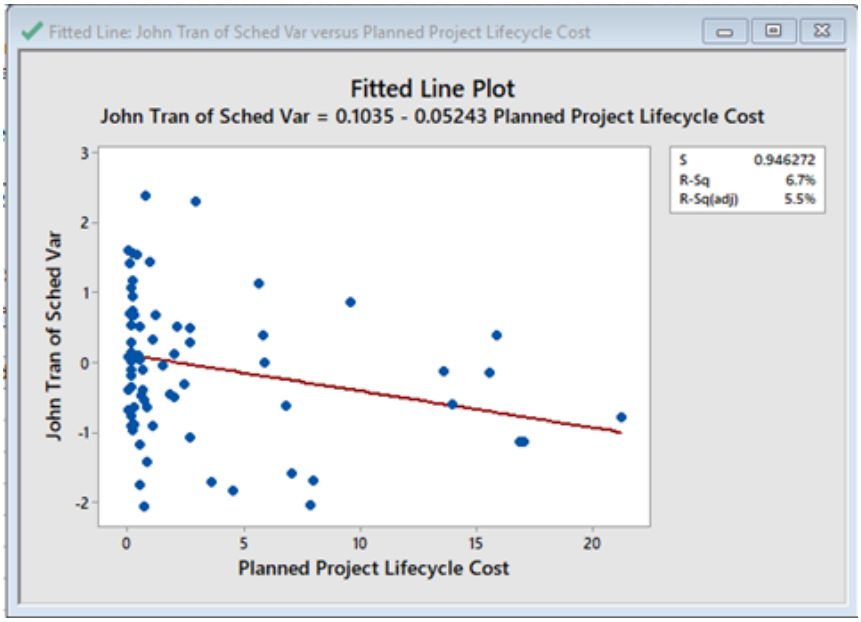


Figure 8: Schedule Variance (after Johnson Transformation) vs Planned Project Lifecycle Cost

The graph in Figure 9 displays a negative relationship; the schedule variance (natural log transformation) generally decreases as the planned project cost (in millions of dollars) increases. The calculated statistical values are S-1.11109, R-sq of 11.8%, and R-sq (adj) of 10.7%.

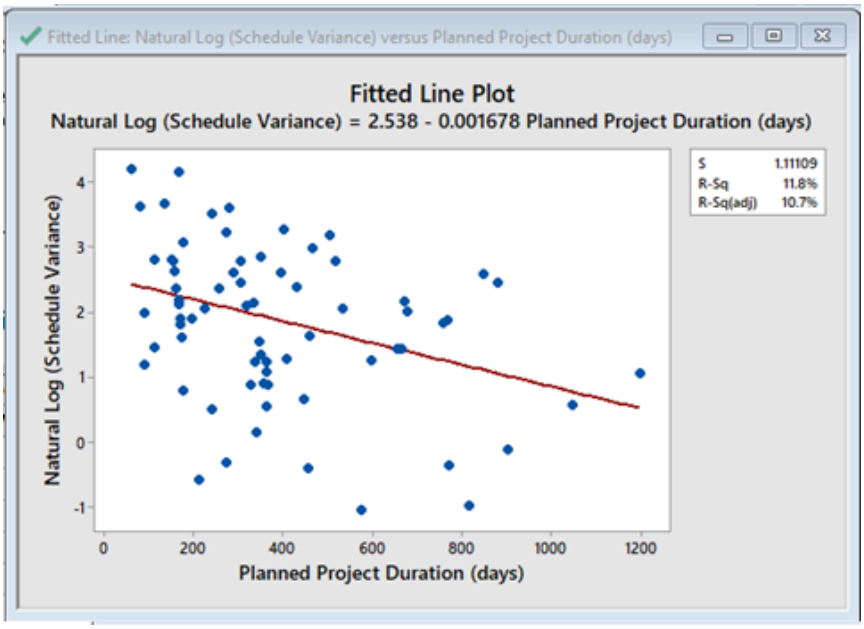


Figure 9: Schedule Variance (after Natural Log Transformation) vs Planned Project Duration

The graph in Figure 10 indicates a slightly negative relationship because the schedule variance (natural log transformation) generally decreases as the planned project cost (in millions of dollars) increases. The calculated statistical values include S-.1.14928, R-sq of 6.7%, and R-sq (adj) of 5.5 %.

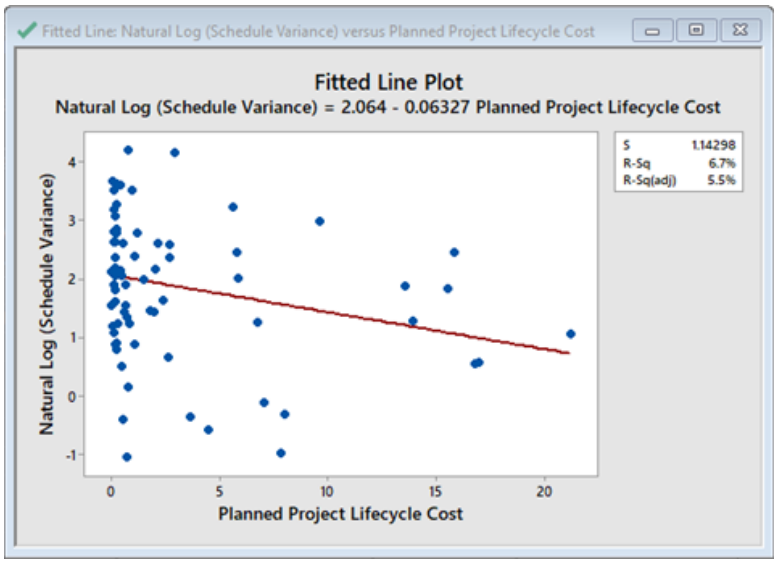


Figure 10: Schedule Variance (after Natural Log Transformation) vs Planned Project Lifecycle Cost

Below is a summary of the schedule variance correlation statistics:

Table 5: Summary of Schedule Variance Correlation Statistics

Response Variable	Predictor Variable	S-value	R-value	R(sq) value
Schedule Variance (After Johnson Transformation)	Planned Project Duration	0.91764	12.3%	11.1%
	Planned Project Lifecycle Cost	0.946272	6.7%	5.5%
Schedule Variance (After Natural Log Transformation)	Planned Project Duration	1.11109	11.8%	10.7%
	Planned Project Lifecycle Cost	1.14928	6.7%	5.5%

4.3 Analysis Results for Categorical Predictor Variables

After observing and analyzing the relationships between the categorical predictor variables and the response variables, boxplots were developed which illustrate the effect

of each predictor on both the cost and schedule variances of the IT projects. The results are presented below.

4.3.1 Effect on Cost Variance

Figure 11 illustrates the relationship between the cost variance and PM experience levels (as explained in Section 3.3.2). The lowest levels of PM experience (levels 7 or 8) correspond with the highest cost variances, while the highest levels of PM experience (levels 1 or 2) correspond with lower cost variance. Level 5 is an anomaly because it corresponds with the lowest cost variance. Since this level is in the middle of the PM experience range, one would expect it to correlate with an average cost variance. This level corresponds with few data points in relation to the other levels, which indicates that perhaps more data points are necessary at this level for a more accurate evaluation.

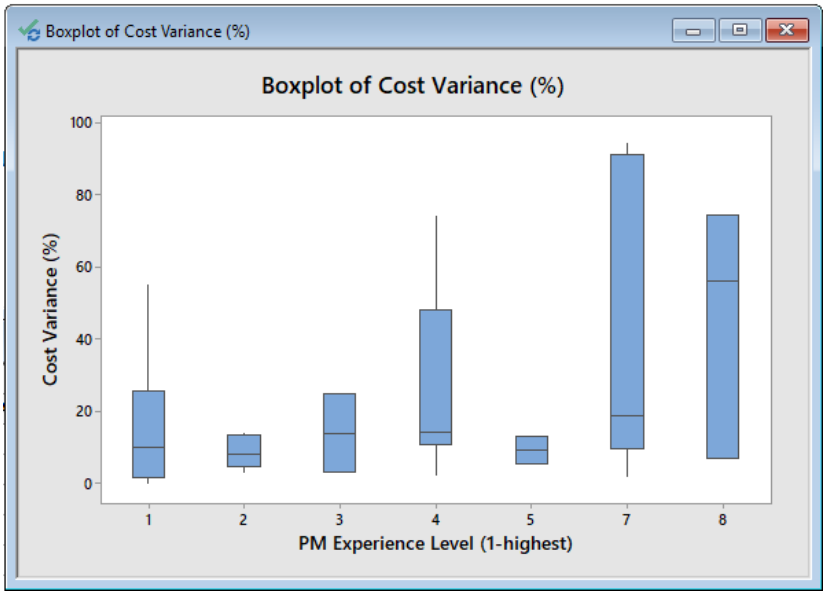


Figure 11:Boxplot of Cost Variance vs PM Experience Level

Figure 12 indicates that projects that applied SDLC methodology 2 had considerably less cost variance than projects that applied SDLC methodology 1. As previously stated, SLDC methodology 1 corresponds with the waterfall methodology, while methodology 2 corresponds with the iterative methodology.

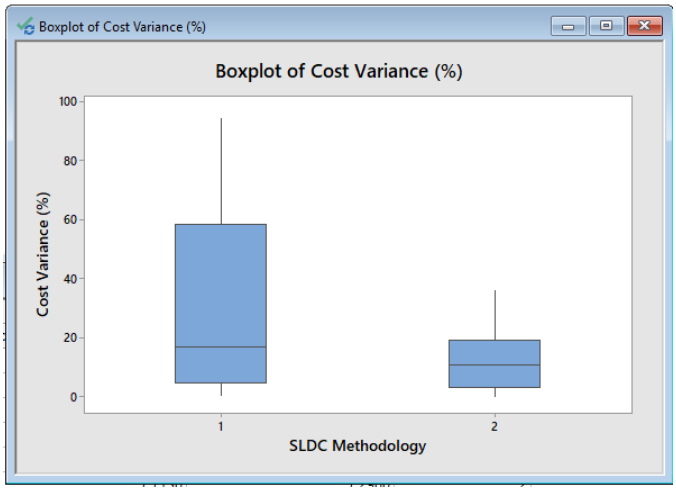


Figure 12: Boxplot of Cost Variance vs SDLC Methodology

Figure 13 demonstrates no significant difference between the cost variance for release frequency method 1-*Product Release Every Six Months* and release frequency method 2-*No Product Release Every Six Months.*

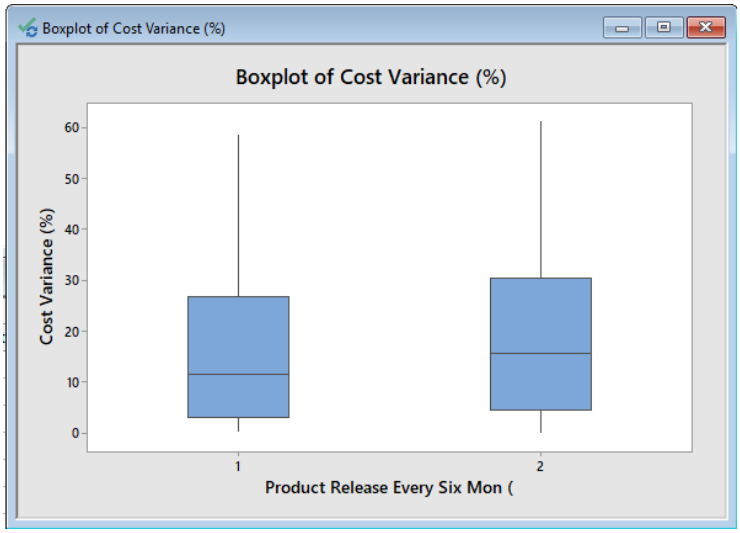


Figure 13:Boxplot of Cost Variance vs Release Frequency

4.3.2 Effect on Schedule Variance

Figure 14, in comparison with the cost variance boxplot, displays that PM experience and schedule variance are not correlated.

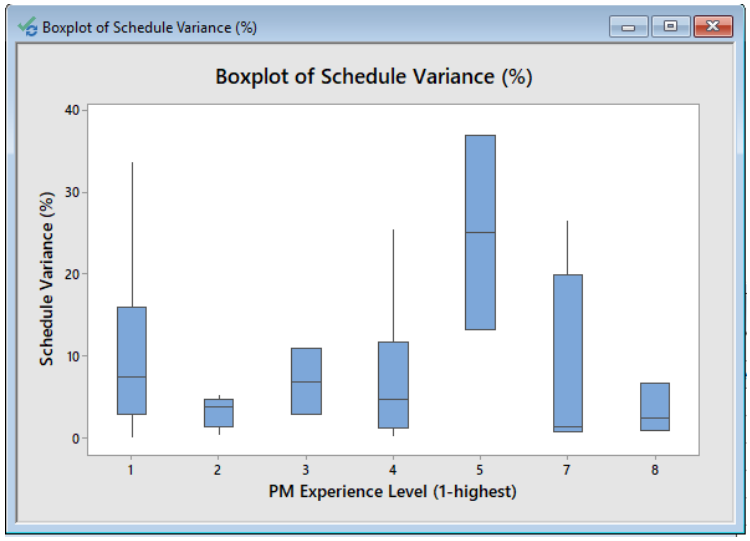


Figure 14: Boxplot of Schedule Variance vs PM Experience Level

Figure 15, when compared with the boxplot of cost variance, suggest that the SDLC methodology does not exert a significant impact on schedule variance.

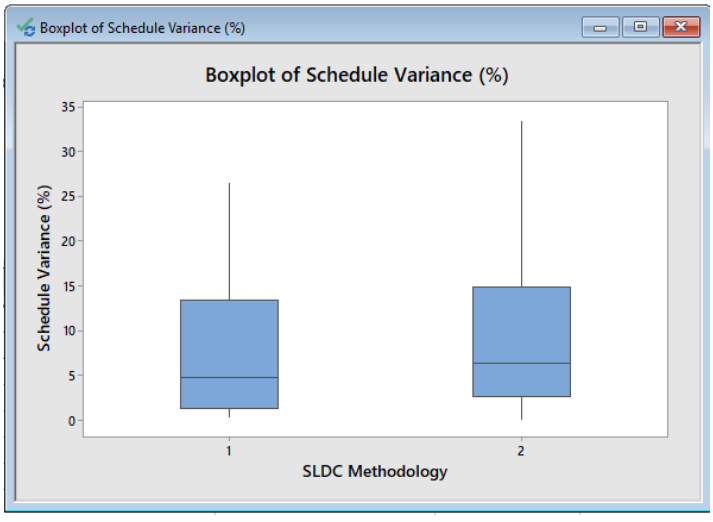


Figure 15: Boxplot of Schedule Variance vs SDLC Methodology

Figure 16 indicates that the product release frequency exerts a slightly greater effect on the degree of schedule variance when the product is released every six months; this results in slightly lower schedule variance relative to when the project has no product release frequency.

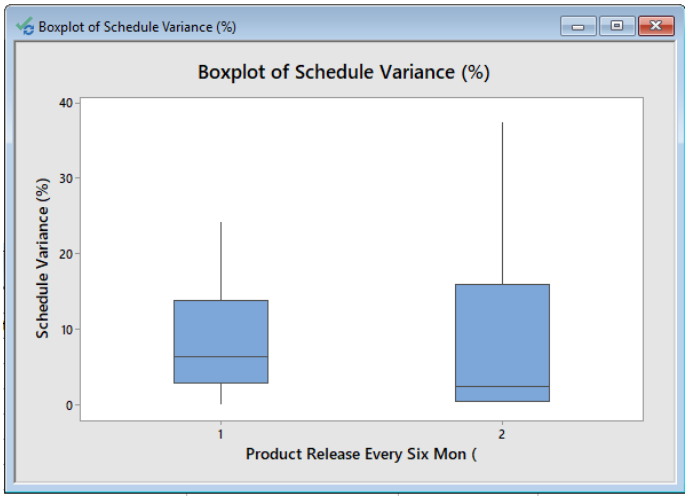


Figure 16: Boxplot of Schedule Variance vs Release Frequency

4.3.3 Two-Sample T-Test Analysis

A two-sample t-test was also conducted for the SDLC methodologies to determine if there is a significant difference in the means of both response variables, cost variance (after Johnson transformation) and schedule variance (after natural log transformation). Before conducting the two-sample t-test analysis, the following assumptions were examined:

1. *Data follows a continuous or ordinal scale*
   The response variables are continuous and the predictor variable is ordinal.

2. *Data is normally distributed*
   An Anderson-Darling normality test was conducted. The results of the test can be seen in Appendix C in Figures C-3 and C-4. In the Anderson-Darling test for normality, the better the normal distribution, the smaller the Anderson-Darling (AD) statistic will be. These values are .604 and .422 for schedule variance (after

natural log transformation) and cost variance (after Johnson transformation), respectively. Additionally, if the P-value is less than .05, then the null hypothesis that the data come from a normal distribution is rejected. The results show that the P-values for schedule variance (after natural log transformation) and cost variance (after Johnson transformation) are .113 and .314, respectively, confirming this assumption.

3. *The sample data from both populations are unbiased and representative*
   The data from both populations of the SDLC methodologies are unbiased and reflect the entire population of data.

4. *The variances of both populations are unknown*
   The variances of the both populations of the SDLC methodologies and the cost variance (after Johnson transformation) and Schedule Variance (after Johnson transformation) are not known.

5. *The sample size is relatively large (sample size >30)*
   The sample size of the datasets that were analyzed was greater than n=30.

In the two-sample t-test, an alpha of 0.1 was used to evaluate the statistical significance in the difference of the means. A P-value of less than 0.1 indicates that there is a statistical significance in the difference between the means, while a P-value of more than 0.1 indicates that there is not a statistical difference between the means. The results of the two-sample t-test shown in Table 6 indicate that there is a significant difference between the means of the cost variance for the two types of SDLC Methodologies because the p-value is $0.063 < 0.1$. However, there is no significant difference between the means of the schedule variance for the two types of SDLC Methodologies because the p-value is $0.459 > 0.1$.

Table 6: Two-Sample T-Test Results

SDLC Methodology	Response Variable	Mean	Difference	P-Value
1-Waterfall	Cost Variance (after Johnson Transformation)	0.28	0.484	0.063
2-Iterative		-0.205		
1-Waterfall	Schedule Variance (after Log Transformation)	1.76	-0.214	0.459
2-Iterative		1.97		

4.3.4 Spearman Correlation Analysis

A Spearman correlation analysis was also conducted in relation to the categorical predictors to determine their relationships with cost and schedule variances. Before conducting the Spearman correlation analysis, the following assumptions were examined:

1. *The variables must be continuous (interval or ratio) or ordinal measurements*
   The response variables, cost and schedule variance are continuous variables. The predictor variables, SDLC methodology, PM experience, and Release Frequency are ordinal measurements.

2. *Monotonic relationship exists between variables*
   From Figures 17-20 it is seen that the response variables are linearly transformed to create monotonic relationships.

Positive Spearman rho values indicate that variables increase together, while negative values indicate that the values are inversely related. An alpha of 0.1 was used to evaluate the statistical significance of the test. P-values lower than 0.1 indicate high statistical significance, while P-values higher than 0.1 indicate low statistical significance. Based on the results in Table 7, the categorical predictor with the strongest positive correlation with schedule variance (after natural log transformation) is Product release frequency; while PM experience level has the strongest negative correlation. The categorical predictor with the strongest negative correlation with cost variance (after Johnson transformation) is SDLC methodology; and conversely, PM experience level has the strongest positive

correlation. However, the only statistically significant correlations with P-values < 0.1 are between PM Experience and Cost Variance, PM Experience and Schedule Variance, and SDLC and Cost Variance.

Table 7: Spearman Correlation Analysis Results of Categorical Predictors

	SDLC Methodology	PM Experience	Product Release Frequency	Schedule Variance (Natural Log Transformation)
PM Experience Level	*Spearman rho:* -0.363			
	P-Value: 0.001			
Product Release Frequency	*Spearman rho:* -0.199	*Spearman rho:* 0.264		
	P-Value: 0.080	*P-value:* 0.019		
Schedule Variance (Natural Log Transformation)	*Spearman rho:* 0.058	*Spearman rho:* -0.191	*Spearman rho:* 0.092	
	P-Value: 0.616	*P-Value:* 0.094	*P-Value:* 0.421	
Cost Var (John Transformation)	*Spearman rho:* -0.189	*Spearman rho:* 0.274	*Spearman rho:* 0.160	*Spearman rho:* 0.053
	P-Value: 0.098	*P-Value:* 0.015	*P-Value:* 0.161	*P-Value:* 0.646

4.4 Regression Analysis Results

To further examine the relationships and the impacts of the selected predictor variables on the response variables, a regression analysis was also performed for both the cost and schedule variance with transformations. Before conducting the regression analysis, the following assumptions were also examined:

1) Linear relationship: Observation of the raw data suggested that the relationship between planned project duration and schedule and cost variance along with the relationship between planned project lifecycle and schedule and cost variance are not linear but instead exhibit an exponential relationship (see graphs A-1 through A-4 in Appendix A). Therefore, to linearize the relationship between both planned project lifecycle and planned project duration to schedule and cost variance, the cost variance data was transformed using the *Johnson transformation method*, which is a transformation feature

available in Minitab that uses the original data and displays the parameters of the Johnson transformation function that produces the ideal fit.

For example, suppose the Johnson transformation function is as follows:

$$\boldsymbol{T(X) = 0.762475 + 0.870902 \times \ln\left(\frac{X - 46.3174}{59.6770 - X}\right)}$$

If the original data value for X is 50, then the transformed data value of 50 is calculated as follows:

$$\boldsymbol{T(X) = 0.762475 + 0.870902 \times ln\left(\frac{50 - 46.3174}{59.6770 - 50}\right) = -0.07893.}$$

In addition, a less accurate *logarithmic transformation* was performed using the Excel function for both the schedule and cost variance in these cases. For example, suppose the natural log transformation function is as follows:

$$\boldsymbol{y = ln\,(X)}$$

If the original data value for X is 50, then the transformed data value of 50 is calculated as follows:

$$\boldsymbol{y = ln(50) = 3.91.}$$

In the case of the logarithmic transformation for the continuous values relationship with cost variance, less normal results were produced relative to the Johnson transformation results, while the logarithmic transformation results for the continuous values relationship with schedule variance produced comparative normal results relative to the Johnson transformation results.

As is discussed in greater detail later, the model accuracy of both the logarithmic transformation and the Johnson transformation was compared as

well. For cost variance, the Johnson transformation yielded better results, while for schedule variance, the natural logarithmic transformation yielded better results. Figures 17 through 20 display the results of the optimum transformations for cost and schedule variances.

2) Multi-variate normality (normal residuals): The test for normal residuals was assessed by observing the residual histogram produced in Minitab. Figures 17 through 20 illustrate the optimal results of the normal residual tests for cost variance (using Johnson transformation) and schedule variance (using natural log transformation).

3) Multi-collinearity: A test was also conducted to assess whether the continuous predictor variables strongly correlate with each other. The VIF (variance inflation factor) from the regression summary was assessed. The VIF values of the predictors ranged between 1.14 and 1.81, which indicates no substantial multi-collinearity among the variables.

4) Auto-correlation: An auto-correlation check was also performed to ensure that the variables were independent of each other. This was investigated by examining the scatterplot of the data, and no auto-correlation was detected. Figures 17 through 20 display the evidence of no auto-correlation.

5) Homoscedasticity: Lastly, homoscedasticity was assessed to ensure that the residuals were equal across the regression line. This was also checked using the residual observation scatterplot and was demonstrated to be equal across the regression line. Figures 17 through 20 illustrate homoscedasticity.

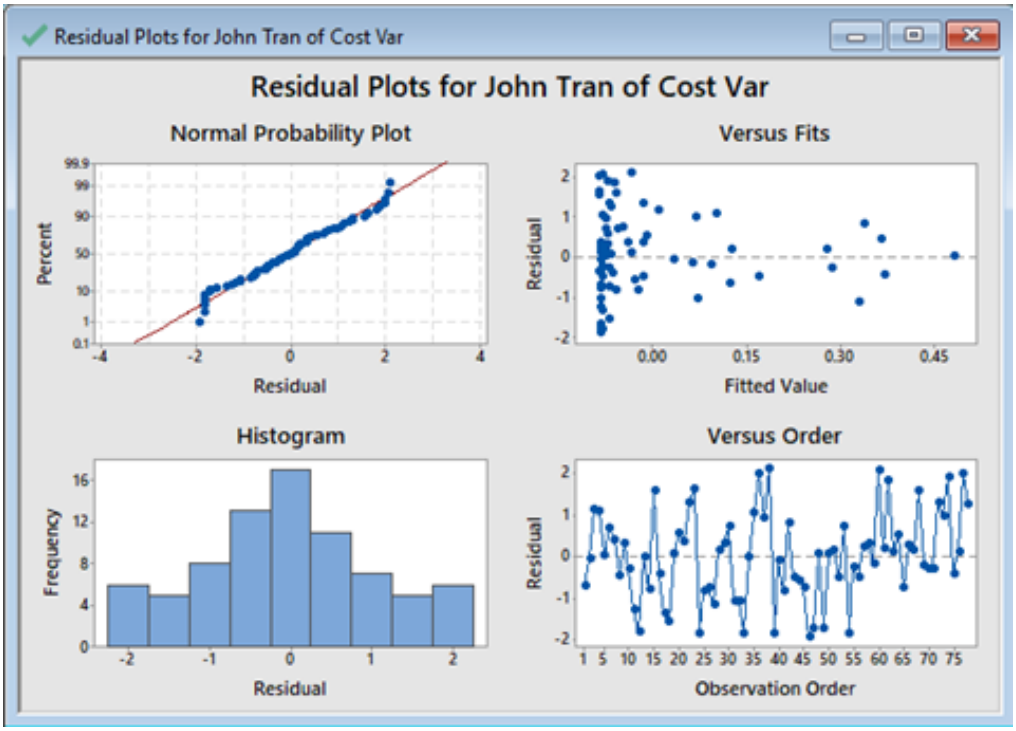


Figure 17: Residual Plots of Cost Variance (after Johnson Transformation) vs Planned Project Lifecycle Cost

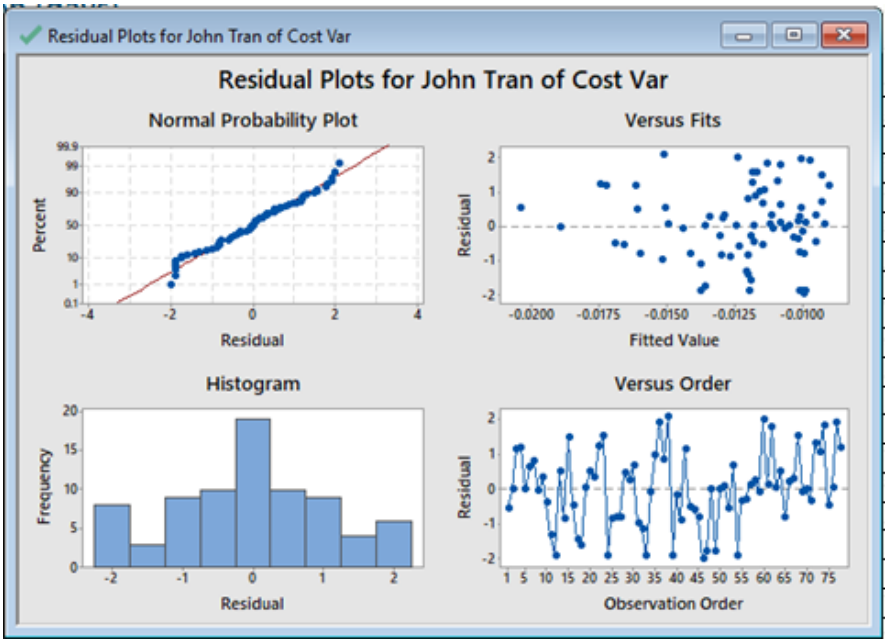


Figure 18:Residual Plots of Cost Variance (after Johnson Transformation) vs Planned Project Duration

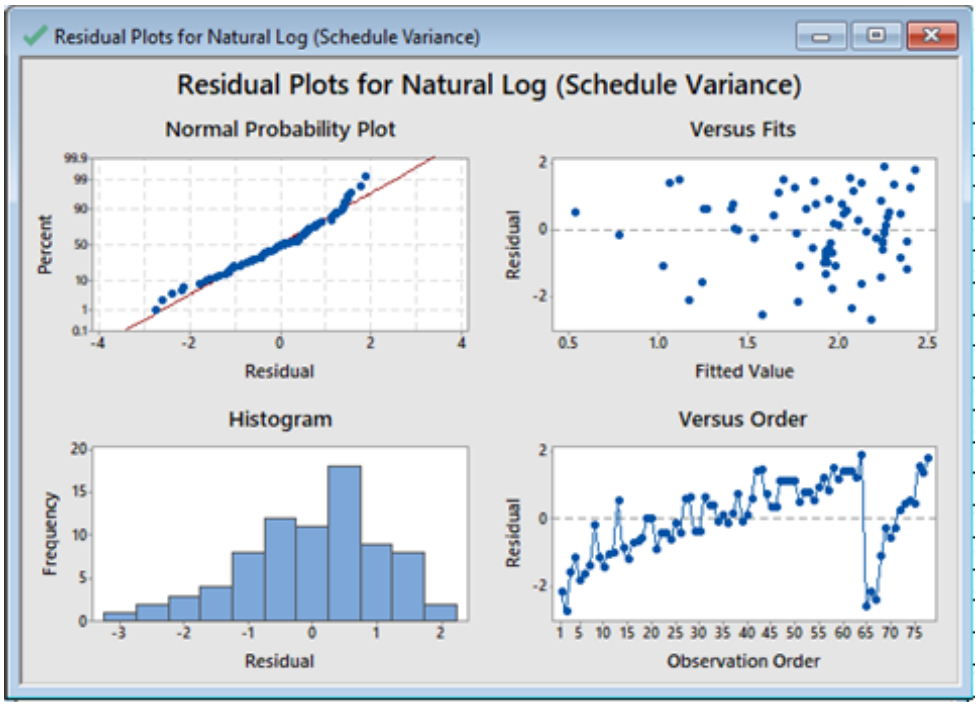


Figure 19: Residual Plots of Schedule Variance (after Natural Log Transformation) vs Planned Project Duration

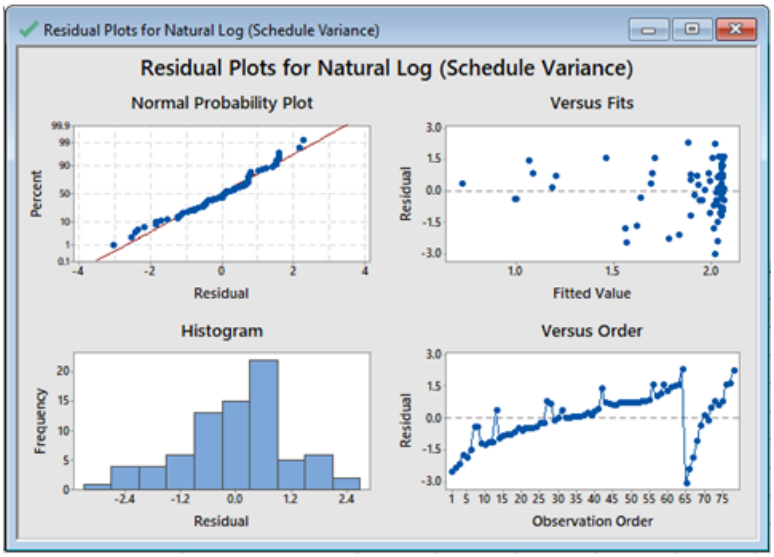


Figure 20: Residual Plots of Schedule Variance (after Natural Log Transformation) vs Planned Project Lifecycle Cost

A normality assessment of the categorical predictors was also conducted using the Ryan-Joiner statistics test which calculates the correlation between your data and the normal scores of your data. A correlation coefficient near 1 indicates that the population is likely to be normal; if it is less than the appropriate critical value, it results in a rejection of the null hypothesis of population normality. The results of the Ryan-joiner tests are as displayed in Table 8:

Table 8: Ryan-Joiner Normality Results of Categorical Predictors

Categorical Predictor	Ryan-Joiner Coefficient	P-Value
PM Experience	0.989	>.1
Software Development Lifecycle Methodology	1	>.1
Product Release Frequency	1	>.1

4.4.1 Cost Variance Regression Analysis

After fulfilling the assumptions previously discussed in section 4.4, an initial regression analysis between cost variance and the predictors was performed to determine

which predictors were significant and which predictors were not based on their determined P-values at an alpha value of .1 (at a confidence level of 90%). Predictor variables with a P-value of less than .1 were considered statistically significant (further explanation of chosen alpha value is in Chapter 3). Table 9 indicates the P-value results of the regression analysis:

Table 9: Cost Variance Predictor P-values

Predictor Variable	P-Value
Planned Project Duration	0.664
Planned Project Lifecycle Cost	0.29
PM Experience	0.124
Software Development Lifecycle Methodology	0.05
Product Release Frequency	0.115

Table 9 demonstrates that the only significant predictor variable for the model is the software development lifecycle methodology, with a P-value of .05. With only one predictor variable deemed significant, the effectiveness of a prediction model for cost variance would most likely be substantially low.

4.4.2 Schedule Variance Regression Analysis

After evaluating the predictor variables in a regression model at the same confidence level, P-values were also calculated for schedule variance regression in Table 10.

Table 10: Schedule Variance Predictor P-values

Predictor Variable	P-value
Planned Project Duration	0.001
Planned Project Lifecycle Cost	0.906
PM Experience	0.097
Software Development Lifecycle Methodology	0.952
Product Release Frequency	0.131

Table 10 illustrates that the significant predictor variables include planned project duration, with a P-value of 0.001, and PM experience, with a P-value of 0.097.

4.5 The Schedule Variance Prediction Model

After the regression analysis was performed for the cost and schedule variance predictors, the effectiveness level of the cost variance prediction model was observed to be low; therefore, considering the time limit of this research project, we focused our efforts on the schedule variance prediction model. As discussed in Chapter 3, once the significant predictor variables were determined, four variants of the schedule prediction regression model were developed, and an accuracy comparison was performed between four variations of the schedule variance prediction model to determine which variant is the most accurate in predicting the schedule variance within 10% of the actual values. Normality assumptions and statistical significance tests were also performed on the model variants. Each model met the normality assumption checks and the significant P-value tests (see Appendix B for results). The results of this comparison are indicated in Table 11. The accuracy rate of the model was calculated using the number of prediction values that were within 10% of the actual values. The prediction values percentage difference was calculated using the actual and predicted schedule variance values shown in equation 4-1. Once the percentage differences were calculated, equation 4-2 demonstrates how the accuracy rates were calculated for each model.

$$Predictions\ Percent\ Difference(Absolute\ Value) = Actual\ Variance\ Percentage - Predicted\ Variance\ Percentage$$

Equation 4-1

$$\text{Accuracy Rate} = \frac{Number\ of\ Predictions\ within\ 10\%\ of\ Actual\ Values}{Total\ Number\ of\ Predictions}\ X\ 100$$

Equation 4-2

Table 11: Accuracy Results of the 4 Variants of the Schedule Variance Prediction Model

Model Variant	Regression Method	Transformation Method	Accuracy Rate
1	Step-wise Fit Regression	Johnson Transformation	79.45%
2		Natural Logarithmic	83.56%
3	PLS Regression	Johnson Transformation	78.08%
4		Natural Logarithmic	79.45%

A normality check was performed on the model variants and each exhibited normal results (see Appendix B for results). Table 11 indicates that the model with the highest accuracy was the *step-wise fit regression model with natural logarithmic transformation* of the schedule variance, using planned project duration and PM experience as predictors, with a prediction accuracy of 83.56%. Table 12 presents the statistical results of this model.

Table 12: Schedule Variance Prediction Model Statistical Results

Model Type	S-value	R-sq	R-sq (adj)	R-sq (pred)
Categorical/Continuous	1.07457	24.05%	16.45%	8.87%
Continuous	1.0895	16.29%	14.06%	9.08%

The model type is denoted by continuous predictors which are numeric variables that have an infinite number of values between any two values and categorical predictors that correspond with one of a limited and typically fixed number of possible values. The first model type demonstrates the statistical results when the planned project duration is input as a continuous predictor, while PM experience is input as a categorical predictor. The second model displays the statistical results when both predictors are input as continuous predictors. The S-value indicates how the model describes the response. The lower the

value of S, the more accurately the model describes the response. In this case, the first model type is slightly better. The R-sq and R-sq (adj) values indicate the percentage of variation in the response that is explained by the model. The higher the R-sq value, the better the model fits the data. Based on the results, the first model type fits the data better. The R-sq (pred) value is calculated using a formula that is equivalent to systematically removing each observation from the dataset, estimating the regression equation, and determining how accurately the model predicts the removed observation. Models with larger predicted R-sq (pred) values have superior predictive ability. Based on the results, the second model type has slightly better predictability. These results suggest that both models are comparable.

Furthermore, Equation 4-3 displays the continuous regression equation (the use of this is discussed further in the *Discussion and Conclusion* section) which inputs both predictors as continuous variables into the model.

$$\ln(\text{Schedule Variance}) = 2.789 - 0.001580\,\text{Planned Project Duration} - 0.1202\,\text{PM Experience Level}$$

Equation 4-3

The combined *categorical and continuous* model, which inputs planned project duration as a continuous variable and PM experience level as a categorical variable, corresponds with the following equations, which are presented in Table 13.

Table 13: Equations of the Categorical or Continuous Schedule Variance Prediction Model

PM Experience Level	Regression Equation		
1	Natural Log (Schedule Variance)	=	2.765 - 0.001859 Planned Project Duration (days)
2	Natural Log (Schedule Variance)	=	2.559 - 0.001859 Planned Project Duration (days)
3	Natural Log (Schedule Variance)	=	3.227- 0.001859 Planned Project Duration (days)
4	Natural Log (Schedule Variance)	=	2.219- 0.001859 Planned Project Duration (days)
5	Natural Log (Schedule Variance)	=	4.143 - 0.001859 Planned Project Duration (days)
7	Natural Log (Schedule Variance)	=	1.945- 0.001859 Planned Project Duration (days)
8	Natural Log (Schedule Variance)	=	1.765- 0.001859 Planned Project Duration (days)

4.6 Validation and Comparison Results

In the final phase of the data analysis, the accuracy of the ideal model (*which is the step-wise regression fit or natural log transformation schedule variance categorical and continuous model*) was calculated as a misclassification rate (as explained in Chapter 3) and compared with the GRS model. A sample of the data results, from the dataset of 78 projects is presented in Table 14. In table 14, green represents projects with schedule variances below a 10% threshold, and red represents project schedule variances above a 10% threshold of the proposed model, as indicated in Table 14. The first column indicates the actual schedule variance percentage from the data. The second column displays the predicted schedule variance calculated from the model. The third column illustrates the color that corresponds to the actual schedule variance. The fourth column indicates the color that corresponds to the predicted schedule variance. The last column displays the matching result between the actual and predicted schedule variance. An “0”

indicates that the variances thresholds did not match, while a “1” indicates that the variances thresholds did match.

Table 14: Subset of the Best Model Validation Results Relative to Historical Data

Actual Schedule Variance (%)	Predicted Schedule Variance (%)	Actual Color	Predicted Color	Match
2.47	8.21	Green	Green	1
2.83	2.72	Green	Green	1
2.93	8.08	Green	Green	1
3.23	7.78	Green	Green	1
3.43	8.48	Green	Green	1
3.45	8.08	Green	Green	1
4.27	12.90	Green	Red	0
4.67	4.83	Green	Green	1
4.67	4.83	Green	Green	1
4.95	11.52	Green	Red	0
5.12	5.51	Green	Green	1
6.08	11.58	Green	Red	0
6.27	3.88	Green	Green	1
6.46	3.81	Green	Green	1
6.59	11.58	Green	Red	0
7.22	13.44	Green	Red	0
7.41	2.61	Green	Green	1
8.62	4.57	Green	Green	1
8.79	11.67	Green	Red	0
10.5	11.76	Red	Red	1
10.9	11.35	Red	Red	1
11.65	3.09	Red	Green	0
13.22	13.03	Red	Red	1
13.41	7.64	Red	Green	0
13.81	11.89	Red	Red	1
15.93	11.95	Red	Red	1
16.48	11.98	Red	Red	1
33.33	10.15	Red	Red	1
33.58	10.13	Red	Red	1
37.3	13.72	Red	Red	1
63.13	11.65	Red	Red	1
36.84	37.38	Red	Red	1
38.64	12.36	Red	Red	1
66.12	14.16	Red	Red	1

The final misclassification rate achieved by our ideal model on the historical data was demonstrated to be 31.5%, which is clearly better than 45%, the misclassification rate reported for the GRS Model.

4.7 Results of Hypothesis Testing

After extensive research and data analysis, some key discoveries were made in relation to the distinct hypotheses that guided the research and data analysis. Below is a discussion of each research hypothesis in relation to cost variance, schedule variance, and the proposed prediction model, along with the results that support the conclusions related to them.

4.7.1 Cost-Related Hypotheses Discussion

- **H_1: Projects with lengthier planned schedule durations tend to have higher cost variances.**

Based on the data analysis of the provided dataset, this hypothesis is not supported and cannot be accepted. The P-value of .664 in Table 9 from the regression analysis along with the R-value from the fitted plot graphs in Figures 3 and 5 suggests that no correlation exists between this predictor and the response variable. This may be because several shorter, more expensive projects (as well as lengthier planned duration, low-cost projects) potentially exist, which could affect the linear correlation between project duration and cost variance. Within IT projects, this phenomenon is highly apparent because organizations constantly require quick-turnaround, high technology implementations that are generally expensive to execute. It is pivotal that IT managers attend closely to these types of projects that are more sensitive to both significant variances in both cost and schedule.

- **H_2: Projects with higher planned lifecycle costs tend to have higher cost variances.**

Based on the data analysis of the provided dataset, this hypothesis is not supported sufficiently convincingly and cannot be accepted. Although the fitted plot graphs in Figures 4 and 6 indicate a slight correlation between higher cost variances and higher planned lifecycle costs, the P-value of .29 from Table 9 does not render this predictor variable statistically significant. As mentioned previously in H1, although evidence does not suggest a correlation of these two variables, it is important that IT managers pay close attention to these projects that have shorter planned lifecycle costs and high planned lifecycle costs because these project types are more sensitive to significant variances in both cost and schedule. Intuitively, the more projects cost over time, the more one would expect the cost variance to be higher.

One reason that it is potentially not evident in this study is the low number of projects that have been considered. Because the projects considered must include all parameters in the data points mentioned in the methodology section, several projects with additional cost variance data points that could be measured have been excluded. Inclusion of these data points could have potentially supported the evidence demonstrating the positive correlation between cost variance and planned project lifecycle costs.

- **H_3: Projects overseen by managers with less experience tend to have higher cost variances.**

Based on the analysis of the dataset provided, this hypothesis cannot be accepted. Based on the data analysis, although the P-value of .124 from Table 9 is not sufficient to deem this predictor significantly significant, Figure 11 indicates a general increase in cost

variance among project managers with less experience; conversely, this evidence suggests that the more years of experience a project manager has, the less cost variance a project will experience. Table 7 also displays a slight Spearman correlation of .274, which indicates that as the PM experience level increases (which corresponds with less experience), cost variance increases.

Based on the supplemental data, this predictor should be accounted for when managing the cost variance of IT projects. This makes sense because it is expected that project managers with more experience and training can plan for and anticipate issues that could adversely affect IT projects, which is supported in the literature *(Lipke et al., 2009* and Gemino, Sauer, and Reich, 2010). Furthermore, depending upon the confidence level that is adopted for this predictor (i.e., 85% confidence level), it may still be significant. This finding also suggests that categorical predictors, which are at times more difficult to quantify than continuous predictors, are equally as important in affecting project performance and should be accounted for during pre-project planning activities.

- **H_4: The type of SDLC (software development lifecycle) methodology affects cost variance**

Based on the analysis of the given dataset, the null hypothesis is rejected. The P-value of .05 from Table 9 indicates that software development lifecycle methodology is a significant predictor variable when considering its effect on cost variance. Results of the two samples t-test shown in Table 6 as well as those of the Spearman Correlation test shown in Table 7 also indicate that there is a significant difference between the means of cost variance for the two groups of projects that use Iterative versus Waterfall SDLC methodologies. Figure 12 also illustrates a considerable difference in cost variance

depending upon which SDLC methodology is adopted. Based on the data, SDLC methodology 2, a more iterative and agile approach, results in lower cost variance for the IT project.

This is also supported by the literature (Pederson, 2013) because iterative and agile projects diminish the risks experienced in projects due their *smaller incremental* approach, wherein risks are mitigated during shorter project cycles well before the final product is released. An agile approach allows phenomena that can ultimately affect cost, such as requirement changes, resource changes, and schedule changes to be addressed in brief periodic segments, which prevents vast cost changes over the course of a project; this approach instead allows for tighter cost control, wherein the IT project management team and the customer the product is delivered to constantly remain in sync during IT project implementation.

- **H_5: The frequency of project releases every six months affects cost variance.**

Based on the analysis of the dataset, this hypothesis cannot be accepted. The P-value of .115 from Table 9 is insufficient to deem this predictor statistically significant, and no substantial difference in cost variance is demonstrated by Figure 13. However, although this predictor is related to the level of agility of a project, this is an important factor that should be attended to. The literature (Geambasu et al., 2011) indicates that release frequency can impact the success of an IT project; projects with more frequent releases experience higher success rates. Related to agile project management methods, higher product release frequencies can allow IT projects to have lower cost variances by enabling IT project teams to gradually monitor and control cost more tightly. However, the results of the analysis of this predictor align with the research of Chow (2016), who

discovered no visible difference in the performance of IT projects based on the method. Therefore, it is important to determine the optimum balance of product release frequencies to ensure that the IT project remains as agile as possible while continuing to deliver value.

4.7.2 Schedule-Related Hypotheses Discussion

- **H_6: Projects with lengthier planned schedule durations tend to have lower schedule variance.**

Based on the analysis of the dataset, the null hypothesis is rejected. The P-value of .001 from Table 10 deems planned schedule durations a significant predictor variable considering its effect on schedule variance. The graphs in Figures 7 and 9 also illustrate that as the planned project duration increases, schedule variance decreases. This is a sensible observation because the overall effect of schedule variance of a project decreases as the length of the project increases.

Conversely, shorter projects are more susceptible to the effects of schedule changes since there is significantly less room for error with these types of projects. However, it is pivotal for IT project managers to perform more pre-planning and risk mitigation upfront and more closely attend to elements that can affect schedule variance for projects with shorter durations, since the same quantity of days of schedule variance correspond with a significantly higher percentage change effect on shorter projects relative to those that are lengthier.

- **H_7: Projects with higher planned lifecycle costs tend to have higher schedule variances.**

Based on the analysis of the dataset, this hypothesis cannot be accepted. The P-value of .906 from Table 10 is insufficient to deem this predictor statistically significant. The graphs in Figures 8 and 10 illustrate a slight correlation between project lifecycle cost and schedule variance and suggests that projects with higher planned lifecycle costs have somewhat lower schedule variances. This demonstrates that lengthier planned projects typically have higher costs, which precipitates lower schedule variances (based on the previous H6 discussion).

The relationship is not as apparent here because there is not a direct correlation between planned project lifecycle cost and planned project duration, which means that in the data, several short duration projects that are expensive to execute may substantially affect this. Nevertheless, IT project managers should consider both the predictor variables of planned lifecycle cost and planned project duration simultaneously during IT project planning.

- **H_8: Projects overseen by managers with less experience tend to have higher schedule variances.**

Based on the analysis of the dataset provided, the null hypothesis is rejected. The P-value of .097 from Table 10 deems project manager experience level a significant predictor variable in terms of its effect on schedule variance. The graph in Figure 14 illustrates that PM managers with more experience typically experience less schedule variation. As mentioned previously in relation to cost variance, project managers with more experience, knowledge of tools, and training can plan for and anticipate issues that

could adversely affect IT projects schedule; this is supported by the literature as well *(Lipk et al., 2009* and Gemino, Sauer, and Reich 2010). In terms of the relevance of project manager experience as it relates to cost variance, it is also evident that this predictor influences schedule variance, which further demonstrates that this predictor, although categorical and not easily quantifiable, can affect overall IT project performance and should not be overlooked.

- **H_9: The type of SDLC (software development lifecycle) methodology affects schedule variances.**

Based on the analysis of the dataset, this hypothesis cannot be accepted. The P-value of .952 from Table 10 is insufficient to deem this predictor statistically significant. The graph in Figure 15 also does not indicate a significant disparity between the types of SDLC methodology. This may be because of the restricted number of data points used during this study. A more inclusive investigation which specifically examines the effects of SDLC methodology and its effect on schedule variance may yield additional insight.

- **H_{10}: The frequency of project releases every six months affects schedule variance**

Based on the analysis of the dataset, this hypothesis cannot be accepted. From the data analysis, the P-value of .131 from Table 10 is insufficient to deem this predictor statistically significant. However, the graph in Figure 16 illustrates a slight correlation between higher variance and a product not being released every six months. Release frequency can impact the success of an IT project in terms of schedule variance because the P-values related to cost and schedule variance are .115 and .131, respectively. This suggests that release frequency exerts a highly similar impact on both cost and schedule

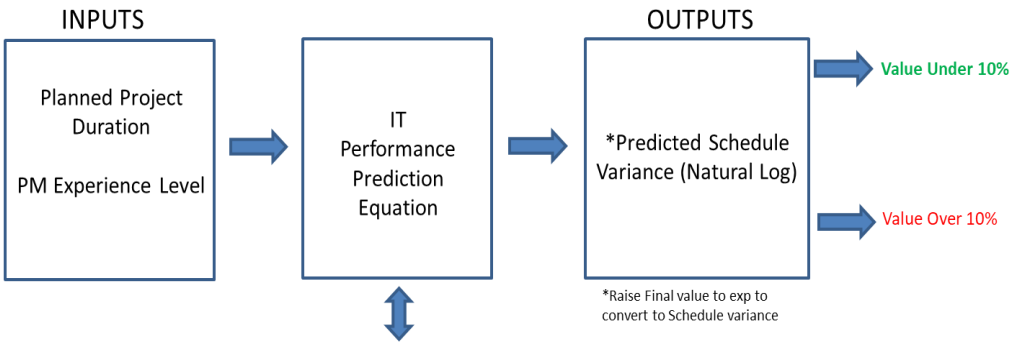
variance and should be regarded as equally important when examining both response variables.

4.7.3 Prediction Model-Related Hypotheses Discussion

- **H_{11}: The prediction results of the model proposed in this praxis will be within a statistically relevant percentage of historical data results for cost and schedule variances.**

Based on the analysis of the dataset, the null hypothesis is rejected as it specifically relates to the schedule variance. A robust prediction model related to cost variance could not be generated from the dataset. Since only one predictor was statistically significant, the cost prediction model would have been a highly inaccurate binary prediction model and was therefore not developed. Based on Table 11, the results of the most accurate model had a prediction accuracy rate within 10% of the actual schedule 83.56% of the time. Furthermore, the sample data in Table 14 suggest that the model tended to err on the side of caution when inaccurately predicting the categorization of project performance levels. These elements of the model render it a useful tool when predicting projects that are highly sensitive to minimal schedule variances. The color-coding of the project performance levels also allows for seamless visual management by IT project managers when efficient decision-making must occur. Figure 21 shows the framework of the final proposed prediction model.

Proposed Schedule Variance Prediction Model Framework



PM Experience Level	Regression Equation		
1	Natural Log (Schedule Variance)	=	2.765 - 0.001859 Planned Project Duration (days)
2	Natural Log (Schedule Variance)	=	2.559 - 0.001859 Planned Project Duration (days)
3	Natural Log (Schedule Variance)	=	3.227- 0.001859 Planned Project Duration (days)
4	Natural Log (Schedule Variance)	=	2.219- 0.001859 Planned Project Duration (days)
5	Natural Log (Schedule Variance)	=	4.143 - 0.001859 Planned Project Duration (days)
7	Natural Log (Schedule Variance)	=	1.945- 0.001859 Planned Project Duration (days)
8	Natural Log (Schedule Variance)	=	1.765- 0.001859 Planned Project Duration (days)

Figure 21: Proposed Schedule Variance Prediction Model Framework

- **H_{12}: The prediction model proposed in this praxis will be more accurate in predicting performance relative to the GRS model.**

Based on the analysis of the dataset and key assumptions, the null hypothesis is rejected. The schedule variance misclassification rates of the GRS model and the proposed prediction model were 45% and 31.5%, respectively. Two equations were generated for the proposed model. One equation uses both predictors as a continuous variable provides project managers with a *blanket* equation that can be used to conduct a quick prediction calculation; however, it does not provide the most accurate results. The second equation type, which is based on the PM management experience level, allows IT project managers to calculate more precise schedule variance based on the project managers' experience levels but can be time-consuming if an Enterprise IT manager has

a variety of project managers overseeing various sub-projects that they are responsible for.

The proposed model may be more accurate than the GRS model because of the use of all of the predictor variables, regardless of their significance, in the GRS model. When researching the GRS model, the statistical significance of both the continuous and categorical predictors was not evident; therefore, no conclusion of their statistical significance could be determined. As a component of the research, the accuracy of the proposed model was also evaluated using all the predictors, but the accuracy was nevertheless lower than the proposed model, which only used the statistically significant predictors. However, since the potential of predictors were included in the GRS model, this could adversely affect its prediction accuracy along with its misclassification rate; this may have compromised the accuracy of the model.

Chapter 5—Conclusion

5.1 Discussion and Conclusions

The results presented in the previous chapter suggest both categorical and continuous predictor variables can impact the cost and schedule variances of IT projects. For cost variance, the most statistically impactful predictor was discovered to be the type of Software Development Lifecycle Methodology; for schedule variance, the most statistically impactful predictor was demonstrated to be both planned project duration and project manager experience level. In both cases, the predictors that impacted IT project performance based on cost and schedule variance included a combination of categorical and continuous predictors. Generally, IT project managers monitor and track the obvious and easily quantifiable metrics that affect cost, quality, and performance, such as budget and schedule.

Based on the research and the results, assessing the non-quantifiable metrics such as project manager experience, project agility, and project management methodologies can also allow IT managers to more accurately predict IT project success. It may be fruitful to examine the federal IT dashboard IT project data more closely and adopt a comprehensive approach by considering both the easily quantifiable variables (continuous predictors) and the non-quantifiable variables (categorical predictors) concerning managing IT projects ; additionally, considering the full spectrum of currently collected factors and potential factors that are not currently collected in the data (such as project complexity and number or type of resources used) that can affect IT project

success will allow IT project managers in the federal government to more successfully predict and manage the outcomes of IT projects.

5.2 Contributions to the Body of Knowledge

Within the federal government, a vast amount of data is collected, but it is seldom deeply analyzed and utilized for better decision-making purposes, specifically as it relates to IT project management. This research project contributes to the body of knowledge in several impactful ways:

1. The research and data analysis that has been conducted contributes to the ongoing academic discourse concerning developing a more comprehensive framework that allows IT projects to be continuously managed more successfully as IT projects become increasingly more complex with shorter execution times.

2. The proposed federal IT project prediction model was developed by examining federal IT data (from the IT Dashboard) from a specific perspective that had not been previously explored (to the researcher's knowledge) using IT project performance data related to cost and schedule variance and investigating potential relationships with other variables. The results of this research provide the federal government with a framework to not only collect data concerning its federal IT projects but also implores the federal government to use the current and future data it collects to acquire valuable insights into how they potentially impact IT project performance.

3. The research and proposed prediction model contribute specifically to the body of knowledge concerning low schedule sensitivity projects. The proposed prediction

model is a tool that can be used as a tool specifically for IT projects that are sensitive to low and positive schedule variances. Complex and *high-dollar* IT projects that have low error margins for meeting schedules on time due to strict resource and budget planning can use this model to accurately predict schedule variance and adjust accordingly.

4. Government agencies that are adversely impacted by down time due to projects being completed overly ahead of schedule can also adopt this model for planning purposes. Being aware of high schedule variances allows organizations to plan accordingly and strategically reallocate resources before and during a project to diminish downtime and the number of potentially unused resources.

5. This study also contributes to the body of knowledge by stimulating the discussion regarding which additional predictors, which are not as apparent, can impact IT project performance within the federal government. It is crucial that both continuous and categorical predictors and their impacts continue to be investigated to continually optimize project performance.

5.3 Recommendations for Future Research

Several recommendations for future research should be pursued to continue to advance this body of knowledge forward, specifically as it relates to IT projects within the federal government. For instance, potential predictors that affect cost variance should be examined more thoroughly since a clear and significant relationship could not be determined during this research. Since cost variance is a pivotal element of project success (in some cases more important than schedule variances), knowledge of the key

predictors that affect this response is crucial for organizations to determine what factors should be investigated to decrease the overall bottom line.

A specific study which targets short duration and high cost projects should be conducted as well as an investigation concerning whether unique predictors potentially impact these types of federal IT projects. As previously mentioned, the types of IT projects that organizations continuously implement are becoming more complex, with shorter project durations and high costs. Investigating the unique factors and predictors that impact the success of these specific types of projects will be beneficial in determining which elements should be monitored closely during these projects that are highly cost- and schedule-sensitive.

Furthermore, future research should examine negative schedule and cost variances to observe whether there are unique predictors that affect these as well. Future investigations and studies concerning these variances can provide organizations with insights to more effectively and accurately allocate resources and budgets before projects begin. This model can also be used in non-federal environments as well as for non-IT projects. Since many similarities exist among organizations outside of the federal government, the results of this study can be leveraged in a variety of ways to facilitate improved project management across various industries.

References

Adoko, M. T. (2016). *Developing A cost overrun predictive model for complex systems development projects* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 3740133)

Alshamrani, A., & Bahattab, A. (2015). A comparison between three SDLC models: Waterfall model, Spiral model, and Incremental/Iterative model. *International Journal of Computer Science Issues, 12* (1), pp.106-111.

Andrikopoulos, V., Binz, T., Leymann, F., & Strauch, S. (2013). How to adapt applications for the cloud environment. *Computing. Archives for Informatics and Numerical Computation, 95*(6), pp.493-535.

Beleiu, I., Crisan, E., & Nistor, R. (2015). Main Factors Influencing Project Success. *Interdisciplinary Management Research*, *11*, pp. 59-72.

Blum, J. R. (2014). *What factors predict how public sector projects perform: A review of the world bank's public sector management portfolio.* Washington, D.C: World Bank Group.

Chow, P. (2016). *A success rate analysis of methodologies in software development: A correlational and comparative study* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 10143290)

Christoffersen, P. & Gonçalves, S. (2005). Estimation risk in financial risk management. *The Journal of Risk, 7*(3), pp.1-28.

Cramer, S. F., Tetewsky, S. J., & Marczynski, K. S. (2010). What facilitates and impedes collaborative work during higher education software implementation projects. *Journal of Educational Change*, *11*(4), pp.457-486.

Crumrine, Kevin T. (2013). *A Comparison of Earned Value Management and Earned Schedule as Schedule Predictors on DoD ACAT I Programs* (Master's Thesis). Air Force Institute of Technology, Ohio, United States.

Darnell, R. C. (2015). *Implicit and explicit measures of coordination effectiveness as predictors of agile software development project success: A regression approach* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 3745479)

Debray A, Koffijberg T.P., Lu, H., Vergouwe, D., Steyerberg, Y. E. W., & Moons, K.G. (2012). Incorporating published univariable associations in diagnostic and prognostic modeling. *BMC Medical Research Methodology*, *12*, pp. 1-9.

Donovan, Shaun (2016, November). Memorandum for heads of executive departments and agencies. Retrieved August 24, 2018 from https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2017/m-17-08.pdf

"Everything You Need to Know About Agile Project Management,". Retrieved October 29, 2018 from https://www.smartsheet.com/everything-you-need-to-know-about-agile-project-management.

Fisher, E. (2006). *Do eighty percent of IT projects Fail? An examination of IT industry studies, compared with the experience of software development and implementation*

projects in Vancouver, British Columbia (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. MR29002)

Geambasu, C. V., Jianu, I., Jianu, I., & Gavrila, A. (2011). Influence Factors for the Choice of a Software Development Methodology. *Accounting and Management Information Systems, 10*(4), pp.479-494.

Gemino, A., Sauer, C., & Reich, B. H. (2010). Using classification trees to predict performance in information technology projects. *Journal of Decision Systems, 19*(2), pp.201-223.

Haji-Kazemi, S., Andersen, B., & Krane, H. P. (2013). A review on possible approaches for detecting early warning signs in projects. *Project Management Journal, 44*(5), pp.55–69.

Hardy, K. (2010). Managing risk in government: An introduction to enterprise risk management. *IBM Center for the Business of Government, 2nd edition*. Retrieved on Aug 27, 2018 from https://www.rims.org/resources/ERM/Documents/Risk%20in%20Government.pdf

Hosford, B. (2017). *Correlational study of project managers' competence experience, education, and technology experience on project* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 10269745)

Hussain, A., C Mkpojiogu, E.O., & Kamal, F. M. (2016). The role of requirements in the success or failure of software projects. *International Review of Management and Marketing, 6*(7), pp. 306-311.

Hussein, B.A. (2013). Factors influencing project success criteria. *Proceedings of the 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), IEEE Conference Proceedings*, *1*(2), pp. 566-571.

Ilincic, R. (2008). *Examining agile management methods and non-agile management methods in global software development projects* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 1461032)

Khan, Z. (2016). *Exploring strategies that IT leaders use to adopt cloud computing* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 10244947)

Kocherla, S. (2012). *Evaluating managerial styles for system development life cycle stages to ensure software project success* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (Order No. 3527308)

Laisasikorn, K., & Rompho, N. (2014). A study of the relationship between a successful enterprise risk management system, a performance measurement system and the financial performance of Thai listed companies. *The Journal of Applied Business and Economics, 16*(2), pp.81-92.

Lee, S., Thomas, S. R. and Tucker, R.L. (2004). Effective practice utilization using performance prediction software. *Journal of Construction Engineering and Management, 130*(4), pp.576-85.

Lipke, W., Zwikael, O., Henderson, K., & Anbari, F. (2009). Prediction of project outcome: The application of statistical methods to earned value management and earned

schedule performance indexes. *International Journal of Project Management, 27*(4), pp.400.

Mefford, R. N., Tay, N. S. P., Doyle, B., & Ohara, F. (2017). Portfolio risk management using six sigma quality principles. *The Quality Management Journal, 24*(2), pp.6-30.

Minitab 18 Support. "Overview of Fit Regression Model". Retrieved on Aug 25, 2018 from http://support.minitab.com/en-us/minitab/18/

Mink,Allan L.,II. (2015). *US federal agencies and cloud: A common decision framework for determining which legacy IT systems should migrate to cloud* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 3707301)

Moy, M. N. (2016). *Evaluating federal information technology program success based on earned value management* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 10013915)

Pedersen, M. (2013). *A quantitative examination of critical success factors comparing agile and waterfall project management methodologies* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 3602588)

Pocock, J., Liu, L. and, Tang, W. (1997). Prediction of Project Performance Based on Degree of Interaction. *Journal of Management in Engineering*,*13*(2).

R. Rabechini J. and M.M. de Carvalho (2013). Understanding the impact of project risk management on project performance: An empirical study. *Journal of Technology Management & Innovation*, *8*(2), pp. 64–78.

Ross, V. W. (2010). *Factors influencing the adoption of cloud computing by decision making managers* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 3391308)

Ryan, D. L. (2010). *Towards the development of a model for hardware standards in information technology procurement: Factors for consideration* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 3398341)

Sawyer, S., & Guinan, P. J. (1998). Software development: Processes and performance. *IBM Systems Journal, 37*(4), pp. 552-569.

Suo, S. (2013). *Cloud implementation in organizations: Critical success factors, challenges, and impacts on the its function* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 3576584)

Tweel, A. (2012). *Examining the relationship between technological, organizational, and environmental factors and cloud computing adoption* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (Order No. 3529668)

U.S. Government Accountability Office. (2013, January). *Information technology: OMB and agencies need to more effectively implement major initiatives to save billions of dollars*. (Publication No. GAO-13-297T).

U.S. Office of Management and Budget (2018). *Analytical Perspectives: Budget of the U.S. Government*. Washington, D.C: U.S. Government Printing Office.

Urhahn, C., & Spieth, P. (2013). Governing portfolio management for innovative new product portfolios: A conceptual framework. *International Journal of Product Development, 18*(5), pp. 377.

Wilson, E. (2014). *Project Sizing Guidelines*. Retrieved on October 27, 2018 from https://intranet.ecu.edu.au/__data/assets/pdf_file/0017/670400/ITSC_Project_Sizing_Guidelines.pdf

Appendix A-Raw Data Plots

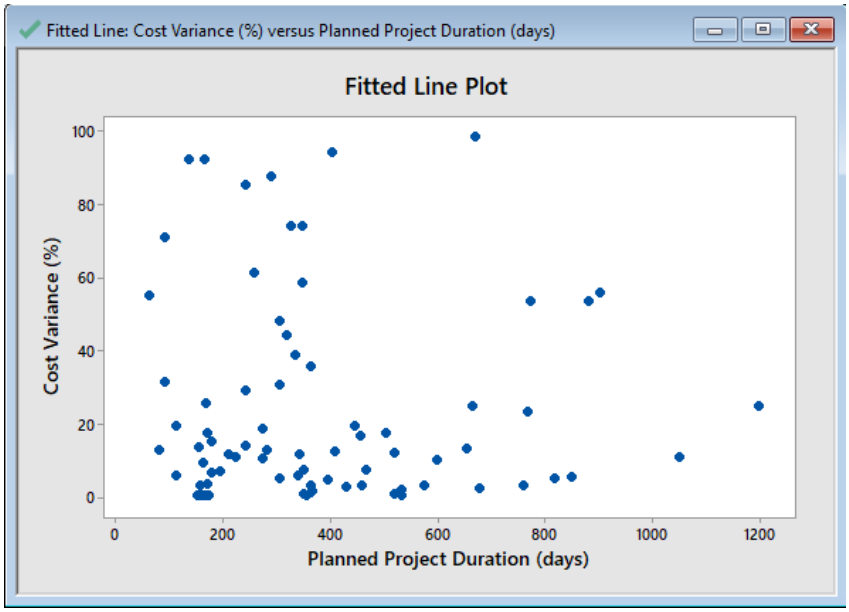


Figure A-1: Cost Variance (%) vs Planned Project Duration

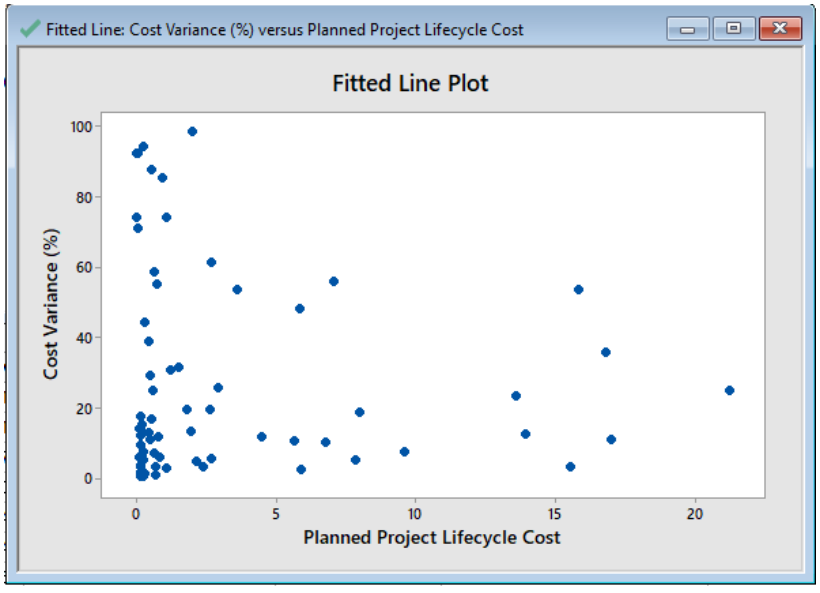


Figure A-2: Cost Variance (%) vs Planned Project Lifecycle Cost

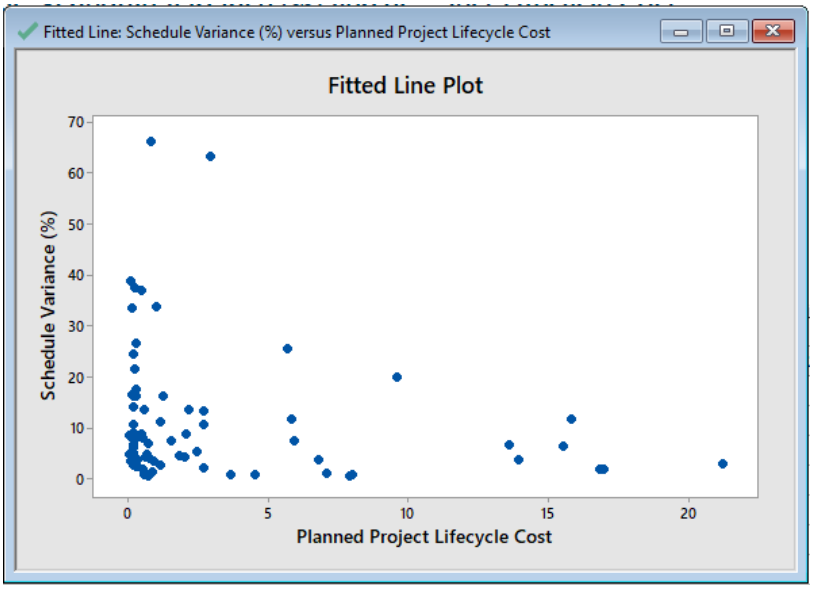


Figure A-3: Schedule Variance (%) vs Planned Project Lifecycle Cost

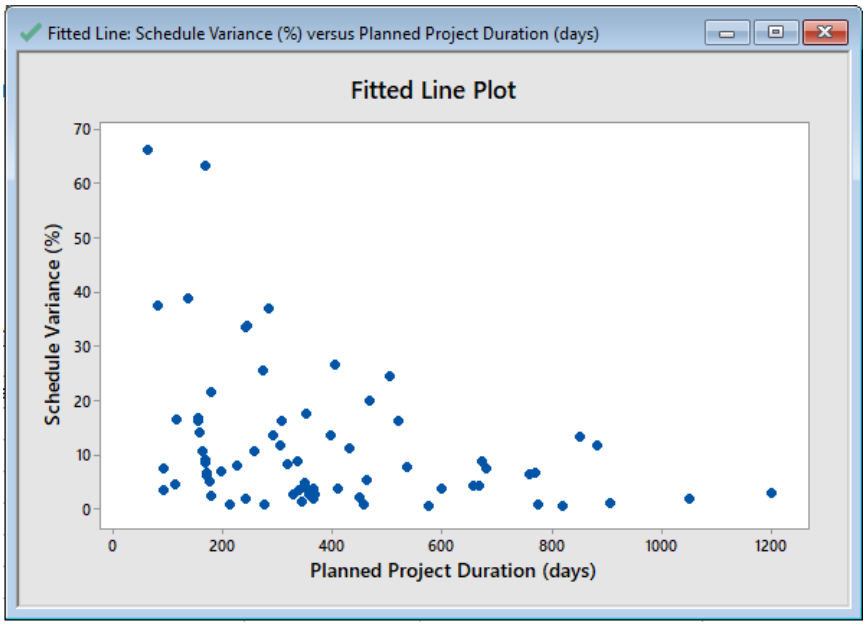


Figure A-4: Schedule Variance (%) vs Planned Project Duration

Appendix B-Model Variants Normality Graphs and P-values

	Response Variable	Transformation Method	Predictor	P-Values
Model Variant 1	Schedule Variance	Johnson Transformation	PM Experience	0.036
			Planned Project Duration	0.002
Model Variant 2		Natural Log Transformation	PM Experience	0.014
			Planned Project Duration	0
Model Variant 3		Johnson Transformation	PLS Model using PM Experience and Planned Project Duration	0.001
Model Variant 4		Natural Log Transformation		0.001

Table B-1: Model Variant P-values

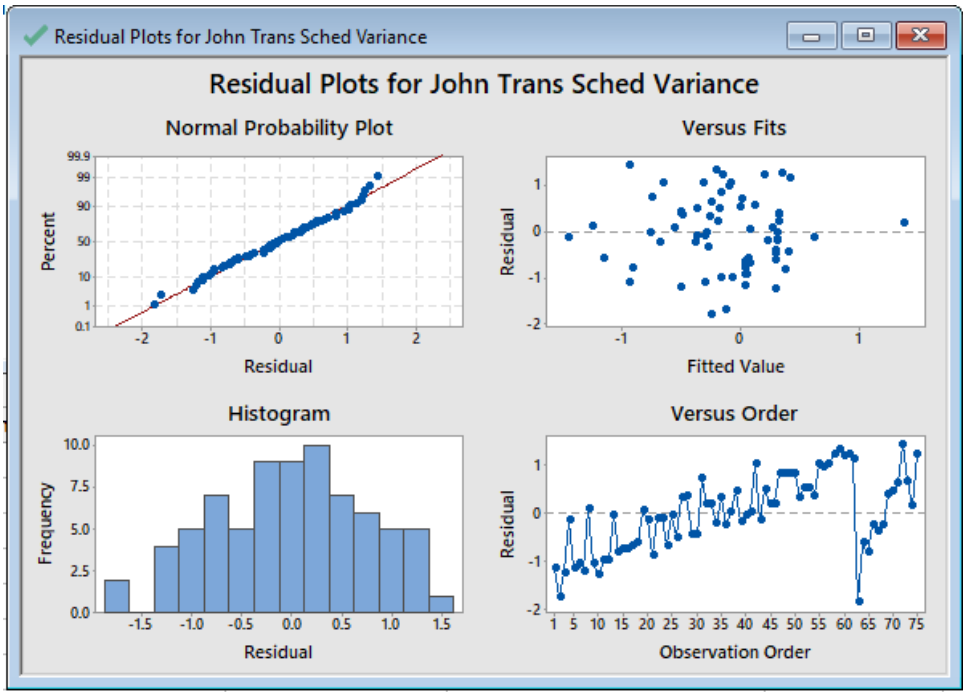


Figure B-1: Model Variant 1 Normality Graphs

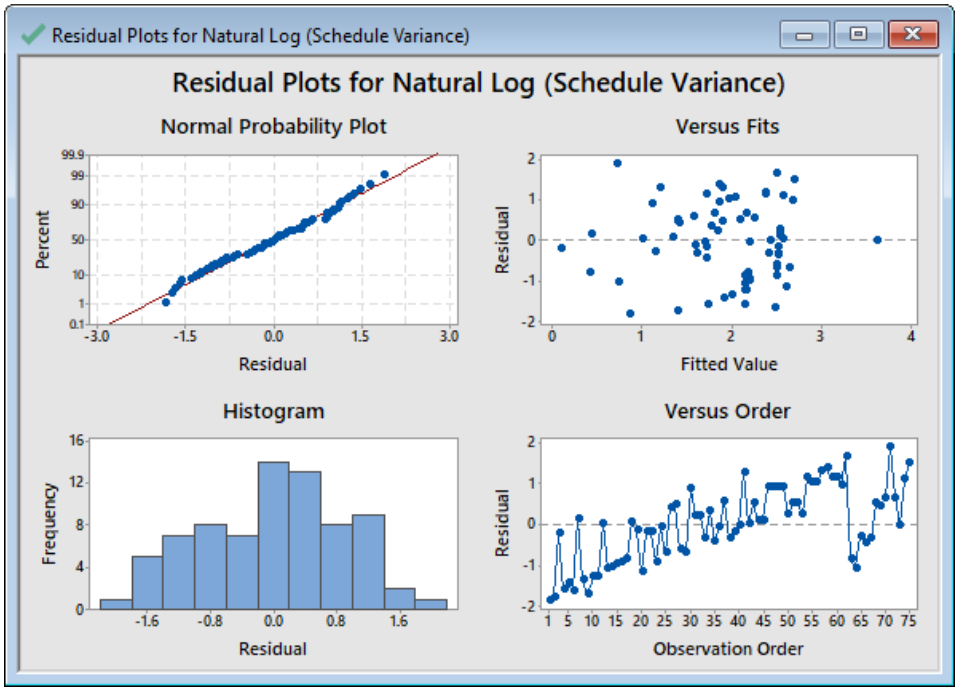


Figure B-2: Model Variant 2 Normality Graphs

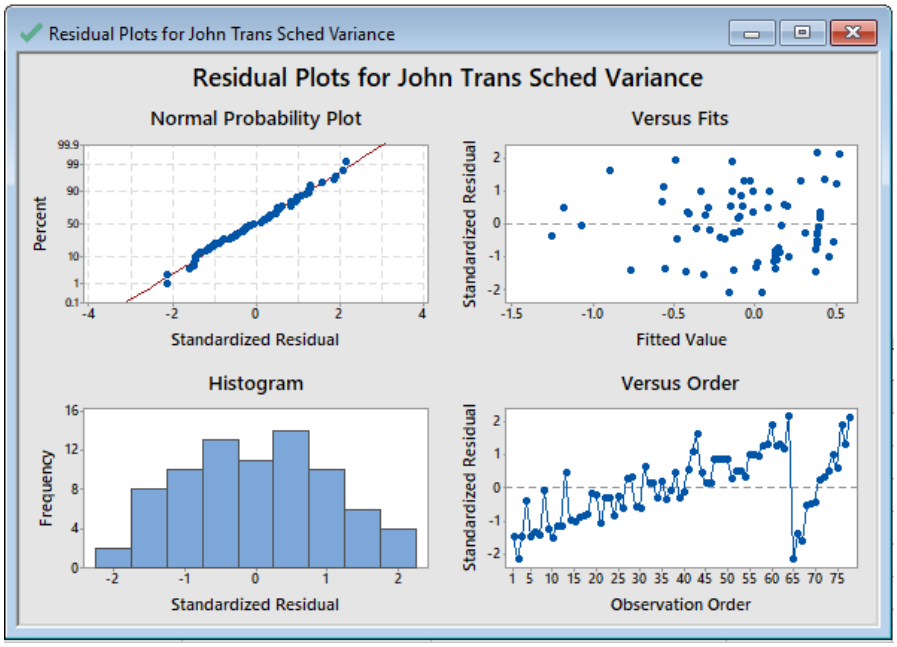


Figure B-3: Model Variant 3 Normality Graphs

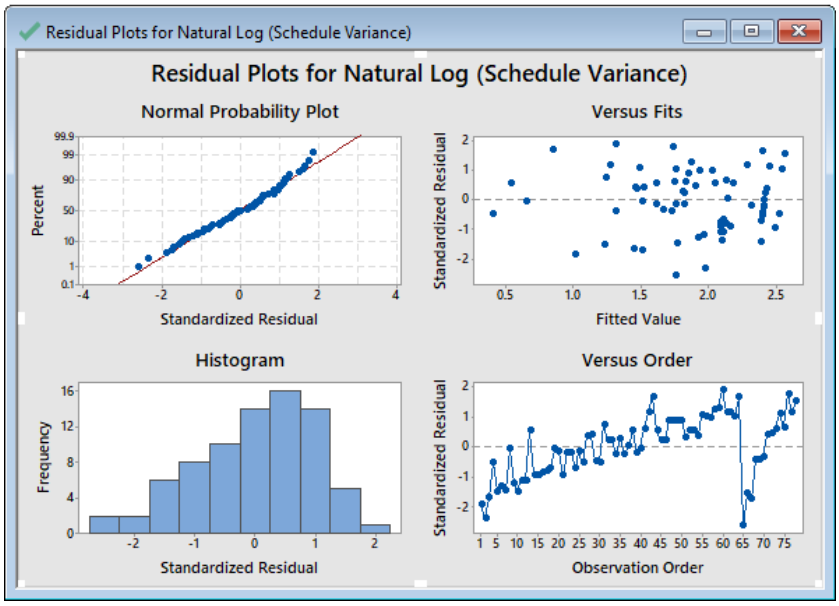


Figure B-4: Model Variant 4 Normality Graphs

Appendix C-Two Sample T-Test Box Plots of Sample Distribution and Anderson-Darling Test Results

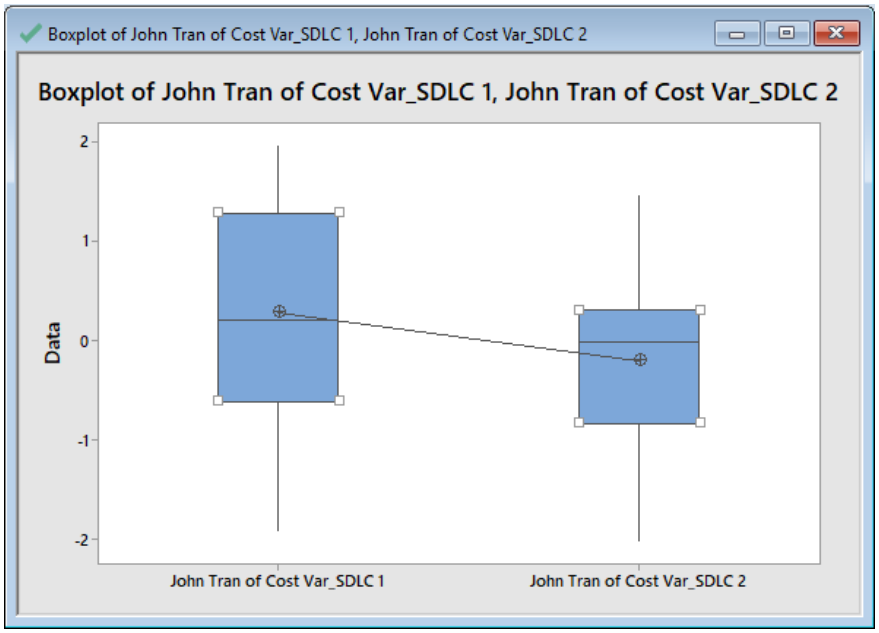


Figure C-1: Boxplot of SDLC Methodology and Cost Variance (after Johnson Transformation)

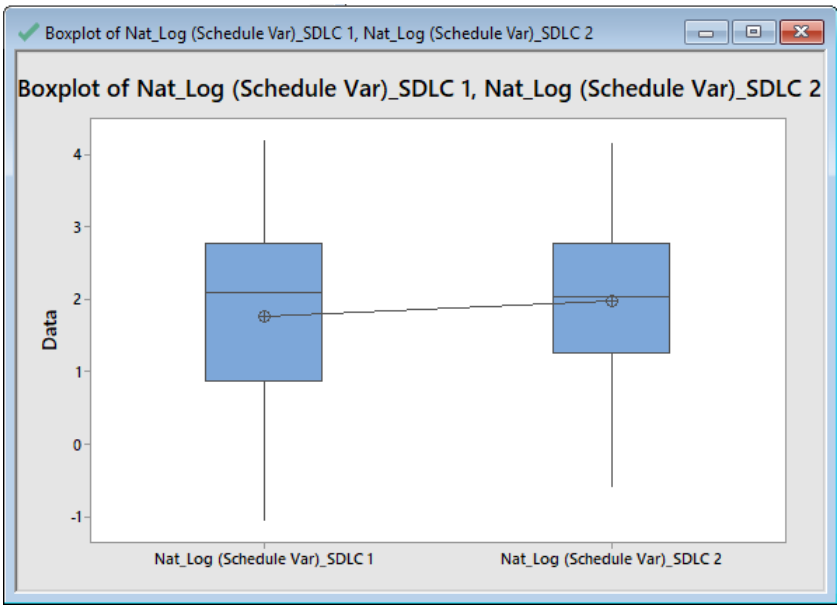


Figure C-2: Boxplot of SDLC Methodology and Schedule Variance (after Natural Log Transformation)

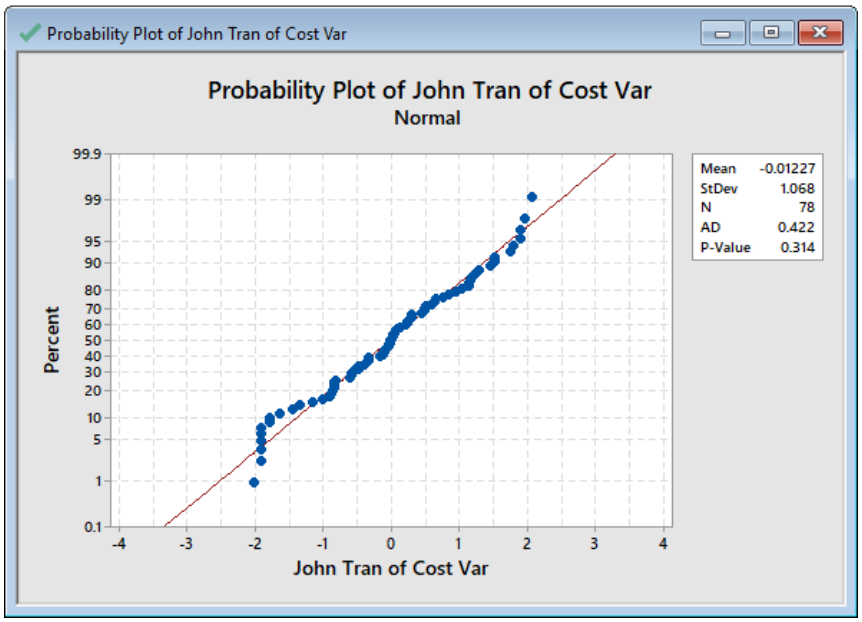


Figure C-3: Anderson-Darling Results of Cost Variance (after Johnson Transformation)

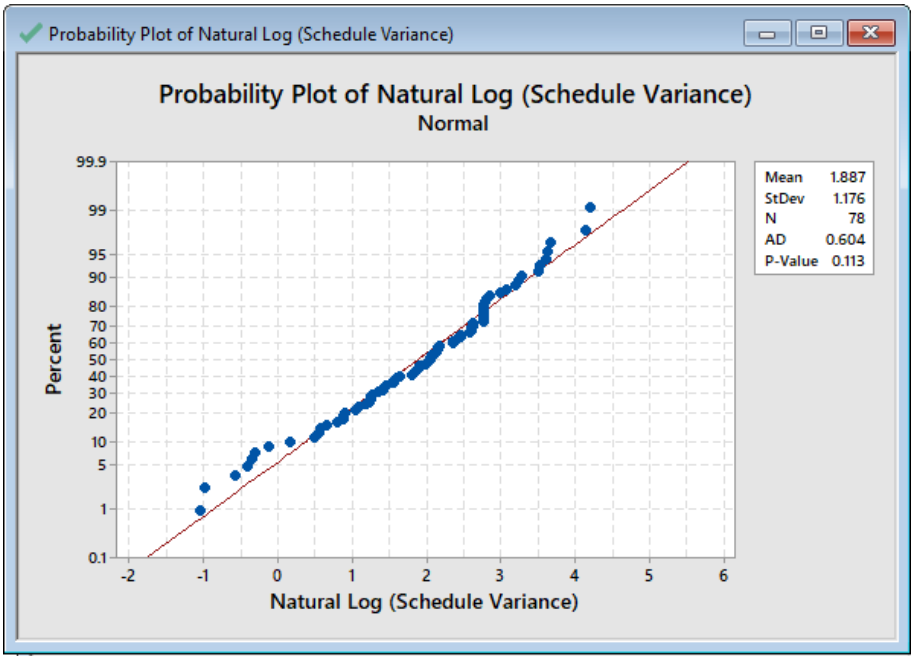


Figure C-4: Anderson-Darling Results of Schedule Variance (after Natural Log Transformation)