3-21-2019

# An Analysis of Estimate Variance in Program Office Estimates

Benjamin J. Bonenfant

**AN ANALYSIS OF ESTIMATE VARIANCE IN PROGRAM OFFICE ESTIMATES**

THESIS

Benjamin J. Bonenfant, Captain, USAF

AFIT-ENV-MS-19-M-163

**DEPARTMENT OF THE AIR FORCE**
**AIR UNIVERSITY**

# AIR FORCE INSTITUTE OF TECHNOLOGY

**Wright-Patterson Air Force Base, Ohio**

**DISTRIBUTION STATEMENT A.**
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENV-MS-19-M-163

# AN ANALYSIS OF ESTIMATE VARIANCE IN PROGRAM OFFICE ESTIMATES

THESIS

Presented to the Faculty

Department of Engineering and Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cost Analysis

Benjamin J. Bonenfant, B.S.

Captain, USAF

March 2019

**DISTRIBUTION STATEMENT A.**
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENV-MS-19-M-163

AN ANALYSIS OF ESTIMATE VARIANCE IN PROGRAM OFFICE ESTIMATES

Benjamin J. Bonenfant, B.S.

Captain, USAF

Committee Membership:

Lt Col Clay M. Koschnick, Ph.D
Chair

Jonathan D. Ritschel, Ph.D
Member

Edward D. White, Ph.D
Member

AFIT-ENV-MS-19-M-163


## Abstract

Past research has shown that predicting the cost growth within DoD systems is an important topic. Total program cost growth and predictors of program cost growth have been studied. Kozlak (2017) studied cost growth at four major reviews: Critical Design Review, First Flight, Development Test and Evaluation End, and Initial Operating Capability. This research attempts to assess cost growth and cost variance at similar points in a program life cycle. In the past the majority of studies have been done identifying programs as either: Acquisition Category (ACAT) I and non-ACAT I programs, or Major Defense Acquisition Programs (MDAP) and non-MDAP programs. This research has data that is able to highlight ACAT II and ACAT III programs. This research also attempts to create a CER for the relationship between Other Government Costs (OGC)-to-contract costs.

The research is not attempting to definitively evaluate or confirm the effects of program characteristics, but is rather trying to guide the bolstering of POE databases and POE research. This database and POE research should highlight cost growth and cost variance for ACAT II and ACAT III programs. Such programs are not highlighted in Selected Acquisition Reports (SAR) or the current cost growth literature.

**Acknowledgments**

I would like to thank my wife and my children for all of their support throughout this process. I would like to express my sincere appreciation to my thesis chair, Lt Col Koschnick, for his guidance and support throughout the course of this thesis effort. The insight and experience is certainly appreciated. I would, also, like to thank my sponsor, Shawn Valentine, from the Air Force LCMC for both the support and data provided to me in this endeavor.

Benjamin J. Bonenfant

## Table of Contents

Page

## List of Figures

Page

## List of Tables

**AN ANALYSIS OF ESTIMATE VARIANCE IN PROGRAM OFFICE ESTIMATES**

## I. Introduction

### General Issue

Historically Major Defense Acquisition Programs (MDAPs) experience 20 percent cost growth from the initial baseline estimate (Drezner, 1993). This cost growth has caused problems for both the DoD and Congress. Congress enacts laws and reforms that require the DoD to produce better cost estimates, while the DoD has to abide by the new laws and reforms to get its programs funded.

When a program experiences cost growth the Program Manager (PM) must request additional funding or reduce the scope of the program. Additional funding requests can impact other programs within the military service that is requesting the funding and it may impact joint programs or programs of other military services. When funds have to be moved from program to program, PMs scrutinize their programs to see if they can come up with cost savings or costs that can be postponed for future years.

If a program is projecting a cost underrun when additional funding is requested for another program, then the projected underrun could theoretically be diverted to the other program. However, projected cost underruns in of themselves raise the issue of the estimate being overly pessimistic. An overly pessimistic estimate can affect the funding of other programs. With the limited budget of the DoD, certain programs are cut due to lack of funding and others are constrained or have their timelines increased to accommodate the budget. Yet, if cost savings were discovered earlier and overly pessimistic estimates happened less, then more programs could be funded while other programs could be finished earlier.

This idea that cost growth as well as cost savings impact the DoD can be conceptualized as projected cost variance. The impact of cost growth is the topic of multiple studies yet this research has not found any that mention cost variance isn't there research in EVMS that studies cost variance. This research looks at projected cost variance and cost growth within Program Office Estimates (POE).

**Specific Issue**

Selected Acquisition Reports (SAR) are reported annually and contain the DoD Component's Current Estimate for each MDAP. A program estimated to have EMD costs greater than $480M or procurement costs greater than $2.79B, in FY2014 constant dollars, is considered an MDAP (Schwartz, 2016). DoD Component's Current Estimate is based off of the Service Cost Position (SCP).

The SCP is developed after a comparison and reconciliation between the POE and independent estimates, including the Component Cost Analysis (CCA). The CCA is developed by the Air Force Cost Analysis Agency (AFCAA). The SCP may be the POE, the CCA, the cost estimate presented in another independent estimate, or an estimate that incorporates all three. In other words the POE and the SAR may not match up and SAR's only look at programs that meet the MDAP dollar thresholds. POEs are conducted for all programs.

This research uses POEs instead of SARs. The research has found no prior literature using POEs. The POEs in this dataset can be grouped into ACAT I, ACAT II, and ACAT III programs. The research has found no prior literature grouping data beyond ACAT I and non-ACAT I programs. The POEs provided only estimates and no actuals. The research attempts to see if certain program characteristics are significantly predictive of estimated cost growth or cost variance.

This research looks at estimate variance and estimate growth using initial program estimates and estimates throughout the program life cycle. The current literature mostly looks at cost growth using estimates at MS B or MS II and completed program actuals throughout the program life cycle.

The research also seeks to use the estimates to create a Cost Estimating Relationship (CER) between Other Government Costs (OGC) and contract costs. Such a CER has not been found by this research and is not presented in the Air Force Cost Analysis Handbook (MCR Federal, 2007).

**Research Questions**

1. How do POEs change over the course of a program life cycle?
2. What are predictive characteristics of POE cost variance or cost growth?
3. What is the distribution of the ratio OGC-to-contract costs?

**Preview**

A review of the relevant Cost Growth literature and potential predictor variables is conducted in Chapter II. Chapter III explains the methodology of this research. Chapter IV is the analysis and results of this research. Chapter V goes through conclusions and thoughts for future research.

# II. Literature Review

## Chapter Overview

This chapter first reviews Program Office Estimates (POE) and briefly covers cost estimating techniques. The chapter then delves into the cost growth literature from RAND, IDA and AFIT. After that the chapter covers the predictor variables, the estimate factors, and the thresholds, while delving into the literature that discusses them. Then Other Government Costs (OGC) are discussed and the most common cost estimating techniques used for estimating OGCs: the analogy method, the engineering build-up method, and the parametric method (AFCAH, 2007).

## Program Office Estimates

POEs are acquisition cost estimates conducted within the program office. The DoD also conducts independent estimates, which are conducted by organizations other than the program office. A program estimated to have EMD costs greater than $480M or procurement costs greater than $2.79B, in FY2014 constant dollars, is considered a Major Defense Acquisition Program (MDAP) (Schwartz, 2016). Every MDAP is to be reported annually in the Selected Acquisition Reports (SARs). The estimates provided in the SARs may be based off the POE, an independent estimate or a combination of the POE and independent estimates.

The SAR estimate, also known as the Service Cost Position (SCP), is developed after a comparison and reconciliation between the POE and independent estimates. One of these independent estimates is the Component Cost Analysis (CCA). Which is developed by the Air Force Cost Analysis Agency (AFCAA).

POE and SAR estimates may not be the same dollar amounts and SARs only look at programs that meet the MDAP dollar thresholds. POEs, which are the focus of this research, are conducted for all programs. No prior literature has been found to use POEs.

Earned Value Management (EVM) and Estimates at Complete (EAC) both use estimates. EVM is not required for POEs or for ACAT II or ACAT III programs. This research seeks to match up POEs to SARs as much as possible which limits the applicability of EAC research. ACAT I, ACAT II, and ACAT III programs are within this dataset. No literature has been found that groups data beyond ACAT I and non-ACAT I programs or MDAP and non-MDAP programs. POEs within this research provides estimates, while the actuals are inferred by the percent complete data.

**Cost Estimating Techniques**

When developing a cost estimate the cost estimator should follow guidelines. The cost estimator may use the work breakdown structure (WBS) of the weapons system to organize and develop the estimate. Cost estimating is also known as cost modeling. Cost modeling highlights that when estimating the costs of a weapons system there are many individual pieces that make up the whole--how they tie together forms a cost model. The cost estimator needs to make a list of all assumptions related to the cost model.

The cost estimator must ensure that their estimate is in constant-year dollars to account for inflation. Inflation is the general rise in prices over time. The cost estimator should time-phase the estimate so that it can be broken out into portions that are to be completed year by year. The time-phasing should take into consideration how long it takes to complete the necessary tasks and how much money is going to be budgeted per year.

The cost estimator needs to validate the cost estimate. This is done by in-house proofreading of estimates and running simulations. This is also done by comparing the estimate to estimates from outside sources. Estimates performed by organizations not directly involved with the development of the weapons system are known as an independent cost estimate. Cost estimates are updated periodically as data—such as actuals—becomes available. (GAO, 2009).

**RAND and IDA Research**

Asher (1980) used SAR data and developed a method to predict weapon system cost growth. Asher (1980) divided the database into different categories: aircraft, missile, ships, and other systems. A six-step approach was developed to determine development and procurement cost growth. Estimator interpretation and subjective evaluation of the data is allowed for by their methodology. There is little mathematical backing to support the estimates, since the estimates were created subjectively. Asher (1980) state that as the DoD program database grows with future historical programs, cost estimating will improve.

128 weapons systems were studied by Drezner (1993). Drezner (1993) looked at development, procurement and total program duration for weapon system Cost Growth Factors (CGF). The two main factors that Drezner (1993) found to have an effect on CGFs were inflation and quantity. Inflation and quantity had such a large effect that Drezner (1993) accounted for the effects of inflation and quantity to see if there were any

other factors that could be influencing CGFs. After inflation and quantity were accounted for Drezner (1993) found that on the cost growth of an individual weapons system increases by 2.2% per year, on average. In other words cost growth increases by about 20% through the life of a program. Drezner (1993) found that Procurement CGFs were 7% less than Development CGFs. Drezner (1993) discovered that cost growth significantly correlates to longer program duration. Drezner (1993) also found that modification programs experienced less cost growth than new start programs (Drezner, 1993).

The research of Arena (2006) consisted of 68 completed programs which were similar to programs acquired by the Air Force. Arena (2006) used SAR data and defined completed programs as programs that had more than 90% of production completed. Arena (2006) found that there were three major categories affecting cost growth: schedule factors, acquisition strategy, and other factors. Schedule slip and program duration were the schedule factors that affected cost growth. Arena (2006) used CGFs and defined them as the ratios of the actual costs to the estimated costs. Contract incentives, competition in production, modification and prototyping were some of the acquisition strategies. Program management decisions, and poor cost estimates were other factors that were looked at (Arena, 2006).

Arena (2006) also split the data by funding category, milestones (MS) and commodity type. The funding categories that were looked at were Development and Procurement. The MS that were looked at were MS II and MS III. Aircraft and missile were some of the commodity types that were observed. Arena (2006) found that there was significant cost growth at MS II and MS III; 46% and 16% respectively.

Leonard (2013) focused on cost growth from commitment, typically MS B, to system development. System development is when a portion of production units, planned at MS B, are produced and delivered to the customer.  Programs at least 5 years past MS B but less than 80% funded were put into the continuing programs group.  Programs that were at least 80% funded were put into the completed programs group.  Three continuing space programs had extreme cost growth or cost growth greater than one standard deviation. The addition of the F-35 to the three space programs, makes up 95% of cost growth for all continuing programs.  Enhanced scrutiny is placed on the DoD's largest systems, since smaller programs were experiencing minimal cost growth. Leonard (2013) anticipate four programs, the F-35A, EELV, KC-46A, and the Long Range Strike Bomber will consume the majority of MDAP funding for the next 20 years.

**AFIT Research**

The Air Force Institute of Technology has multiple studies regarding cost growth. The majority of past studies focused on total cost growth.  The research of Kozlak in 2017 and this research focuses on cost growth at different points during the program lifecycle.

White (2004) used logistic and multiple regressions to predict cost growth for DoD weapon systems.  The cost growth that occurred during the Engineering and Manufacturing (EMD) phase of the acquisition lifecycle was the main focus. A logistic regression model predicted 70% of the validation data and identified schedule variables to have the most predictive ability, when focusing on Research and Development (RDT&E) dollars and limiting the study to engineering cost growth.  The research used the EMD portion of SARs. An objective approach to estimating cost growth was possible because the research used historical data and regression analysis. Cost growth was measured as a percent change from the development estimate to the final estimate. The logistic regression was used to identify programs sustaining cost growth. Then multiple regression was used to predict the cost growth in the identified programs. The multiple regression model contained time variables, length of program, funding variables, and weapon classification. A foundation for predictive cost growth research was ensured with the creation of an objective method for predicting cost growth by White (2004).

This work of White (2004) was built upon by Bielecki (2005) and Moore (2005). They generated logistic and multiple regression models to predict cost growth in different funding appropriations. Bielecki (2005) focused on cost growth in the RDT&E budget during the EMD phase of the program lifecycle. Moore (2005) focused on cost growth in the procurement budget during the EMD phase. The research of White (2004) was validated by their research. Their research also provides further detail into the predictive characteristics of program cost growth in RDT&E and procurement budget categories.

Birchler (2011) looked at 28 programs and used multiple regression techniques to determine if concurrency was correlated to cost growth. They found no relationship between cost growth and concurrency, yet their research into concurrency inspired Jimenez (2016) and Trudelle (2017) to research concurrency in their own research.

Jimenez (2016) conducted research developing a multiple regression model to predict the length of a program's schedule. The schedule would start with MS-B and go to IOC, using historical figures. The research included 56 programs and relied upon three separate statistically significant predictor variables. The predictor variables were whether a program was a modification to an existing platform, or a new start, the year MS-B occurred as it related to a specific change in defense acquisition policy, and the amount of funding prior to MS-B. This model is successful at explaining 42.9% of variability and predicting a realistic program schedule (Jimenez, 2016).

Trudelle (2017) added to the research of Jimenez (2016) by using the same database plus 17 programs. Trudelle's (2017) research had 73 programs instead of 56. Trudelle (2017) developed a regression model and logistic models that predicted the probability of overrunning cost and schedule growth thresholds. The regression model found Projected MS B to IOC (in months), RD&E dollars at MS B Start, Fixed Wing, Electronic System Program, ACAT I, Large Program, and Extra Large Program to be significant variables. Trudelle (2017) also found that MDAP or ACAT I programs were significant predictors in two of the logistic models. The thresholds used for the logistic models were the Nunn-McCurdy 15% and 25% increases from previous estimates (Trudelle, 2017).

Kozlak (2017) looked at 30 ACAT I programs. Kozlak (2017) used SAR data and categorized it by review: Critical Design Review (CDR), First Flight (FF), Development Test & Evaluation End (DT&E), Initial Operating Capability (IOC), and Last SAR (LS). Kozlak (2017) calculated the mean and median percent completes associated with each program review. Kozlak (2017) used Fisher's Exact Tests to determine the significance of program characteristics within his dataset. The variables that were found to be significant in Kozlak's (2017) Fisher's Exact Test results were: Months MS A to MS B greater than or equal to 50, Bomber, Aircraft, Prototype, MS B estimate after 1985, Modification, RD&E funding greater than 50%, Air Force, and Fighter (Kozlak, 2017).

**Predictor Variables and Relevant Research**

There are many variables that may be used by estimators or researchers to predict cost growth or variance. This research focuses on the variables ACAT I, ACAT II, ACAT III, prototyping and concurrency.

Birchler (2011) described concurrency as RDT&E appropriations being authorized within the same years that production appropriations are authorized. Jimenez (2016) and Trudelle (2017) used this definition. Drezner (1993) said that concurrency is designed to speed up the schedule so that there is less time for potential mishaps to occur but it also decreases the time available for early testing. Drezner (1993) measured concurrency as overlap in months between the completion of IO&E and MS-3. In other words if testing overlaps with initial production than concurrency is occurring. Jimenez (2016), Trudelle (2017) and Drezner (1993) found that concurrency was not significant. This research predicts that concurrency in its models will also be insignificant. This variable was looked at in the past and it is useful to be looked at again in this new database that includes ACAT II and ACAT III programs. Due to the lack of documents that analyze ACAT II and ACAT III programs there is a sparseness to ACAT II and ACAT III data and analysis. This sparseness warrants looking at variables that were looked at in the past regardless of their past significance.

Jimenez (2016) and Trudelle (2017) categorized programs that created a prototype, or prototypes, before the beginning of production as programs with prototyping. These authors used SAR data and SARs specifically mention prototypes. Jimenez (2016) and Trudelle (2017) found prototyping to be insignificant. This research predicts that prototyping will be insignificant in its models as well. Yet, the research is looking into this variable because of the sparseness of data on ACAT II and ACAT III programs. This variable was looked at in the past and it is useful to be looked at again in this new database.

Trudelle (2017) looked at ACAT I and non-ACAT I programs. Trudelle (2017) found that ACAT I programs saw more cost growth than non-ACAT I programs. Trudelle (2017) attributed this to ACAT I programs being more complicated than non-ACAT I programs. The research predicts that its models will agree with Trudelle's (2017) findings.

The research has not found literature mentioning ACAT II or ACAT III programs. The data also has ACAT II, and ACAT III programs. The research predicts that the lower the ACAT level the more cost growth will be seen. This is drawn from ACAT I programs showing more cost growth than non-ACAT I programs. The logic here is based on complexity of program being negatively correlated with ACAT level.

**Estimate Factors, Thresholds, and Relevant Research**

Porter (2009) used Program Acquisition Unit Cost (PAUC) in his Institute for Defense Analysis (IDA) paper on major causes of cost growth in defense acquisition. PAUC incorporates RDT&E, Procurement, and MILCON costs, and is adjusted for quantity. Porter (2009) also used constant year dollars to adjust for inflation. The POEs have constant year dollars. However, the POEs are sparse on MILCON data. This research will be using a modified PAUC that only takes into account RDT&E and Procurement data.

Acquisition Procurement Unit Cost (APUC) is a measurement that focuses on procurement costs. Leonard (2013) used PAUC and APUC in his RAND research report on Air Force MDAP cost growth drivers. PAUC is used as a metric for ACAT I programs. APUC is used for program to program comparison purposes. With this being seen as a valid portion of acquisition estimates the research uses APUC and PAUC in the analysis.

Cost growth is typically calculated in one of two ways. In 1981 McNichols used equation 2.1:

$$\frac{\text{Actual} - \text{Estimated}}{\text{Estimated}} \tag{2.1}$$

The estimated cost is subtracted from the actual cost and then divided by the estimated cost. If growth occurs it will be shown as a percentage above or below 0. Equation 2.2 was used by Drezner in 1993:

$$\frac{\text{Actual}}{\text{Estimated}} \tag{2.2}$$

The estimated cost is divided by the actual cost. . If growth occurs it will be shown as a percentage above or below 1. There is also the method of calculating cost growth as the difference from actuals to the most recent cost estimate. This latest method is not a factor but it was used by Arena in 2006. This research does not have actuals. It has estimates and percent complete. The higher a program's percent complete the more likely they will be using actuals to calculate their estimates.

This research will be using the McNichols (1981) equation for Estimate Growth Factor (EGF) and will be slightly modifying it for the Estimate Variance Factor (EVF). The EGF seeks to show what percentage of growth in cost is experienced by a program. If a program has a 0.20 EGF then it experienced a cost growth of 20%. The EVF seeks to show what the effects are regardless of the variance being negative or positive.

Kozlak (2017) used Fisher's Exact Tests to see the effects that characteristics had on cost growth. Fisher's Exact Tests are contingency tables that are adjusted for small sample sizes. Contingency tables require that the variables are binary. Kozlak (2017) found that when prototypes are included it is more likely that programs will experience cost growth. This research predicts that programs that have prototypes will be more likely to experience cost growth and cost variance. This research uses Fisher's Exact Tests to see what effects characteristics have on the estimate factors, since the data has very small sample sizes.

Trudelle (2017) uses the Nunn-McCurdy breach thresholds as a way to determine the severity of a program's cost growth. Trudelle's (2017) models used the 15 percent for significant and the 25 percent for critical. Trudelle's (2017) research was looking at most recent cost growth and thus used those thresholds. This research used the 30 percent and the 50 percent thresholds because it is looking at cost growth and cost variance from the earliest POE available.

**Other Government Costs**

The research was unable to find any policies or research that provided a Cost Estimating Relationship (CER) for Other Government Costs (OGCs). OGCs are costs that a program office incurs in-house that are not part of a contract. These costs may include administrative equipment, travel costs for program office personnel, Analysis of Alternatives costs, etc. The Air Force Life Cycle Management Center (LCMC) is interested in seeing if a relationship could be created by looking at the OGC to Contract costs as a ratio.

The Department of Defense Risk, Issue, and Opportunity Management Guide for Defense Acquisition Programs, the U. S. Air Force Cost Risk and Uncertainty Analysis Handbook (CRUH), the Joint Agency Cost Schedule Risk and Uncertainty Handbook (CSRUH), and the Cost Risk and Uncertainty Analysis Metrics Manual (CRUAMM) do not mention a policy for estimating OGCs. This research questioned LCMC analysts and found that typically OGCs are estimated using analogous programs or using historical actuals.

OGCs have a chapter in the Air Force Cost Analysis Handbook (AFCAH), Chapter 14, which breaks down into separate sub-parts of OGC, such as Training, Government Test Services, Live Fire Test and Evaluation, etc. Each of these sub points has a paragraph that mentions estimating techniques but, no CER is provided. The Estimating techniques are analogy, actuals, or build-up. In the interest of creating such a CER this research looks at the distribution of the ratios of OGC to Contract costs (AFCAH, 2007).

**Analogy**

Currently the analogy technique is used in some OGC estimates. When a program is very early in its life cycle and it does not have a granular breakdown or actuals to build its estimate it can use historical actuals from an analogous program for its estimate. The analogy method assumes that weapon systems with similar characteristics to historical weapon systems can be estimated by using complexity factors that account for new technology. Complexity factors are generally developed by subject matter experts44. If a new fighter aircraft was being estimated than the cost estimator would use the actual costs of a historical fighter aircraft, multiplied by a complexity factor to formulate the cost estimate for the new fighter aircraft.

Analogy estimates use historical data from other programs which allows them to be used early when data from the new weapons system is unavailable. This method is quick to develop and can be developed at little cost44. Having the estimate stem from a similar weapon system in history lends itself to easy comprehension. However, the analogy method has downsides as well. The complexity factors are subjective and no historical weapon system is a direct match to a new weapon system (GAO, 2009).

**Engineering build-up**

Currently Engineering build-up is used in some OGC estimates. When OGC's can be broken down to a very granular level, then using the engineering build-up technique would be logical. Engineering build-up relies on the detail of the weapon system's WBS. The more details available and the more accurate the estimate of each element or component than the better the engineering build-up method will estimate the cost. This method starts at the lowest WBS level, such as labor hours or materials, and adds all the estimates together. Then each WBS level is summed up and added to the next level until all components and levels are calculated.

Engineering build-up estimates forces the cost estimator to track each piece of the weapons system. Tracking all the pieces can illuminate cost drivers. Yet, engineering build-up estimates are very time consuming and very data reliant. If the data is not available at the granular level then engineering build-up estimates become very difficult (GAO, 2009).

**Parametric**

This research is attempting to create a parametric technique that can be used for OGCs. Parametric cost estimating uses statistical relationships between historical actual costs a new weapon system. These statistical relationships are also known as a CER. CERs are made to predict the future costs of a new weapon system based on the historical data relationships. Regression is a common method of developing a CER because it allows the estimator to make statistical inferences. R-squared (R2), statistical significance (p-value), F Statistic, and t Statistic are the most important regression statistics to consider when using parametric estimating (GAO, 2009).

**Summary**

This chapter started off with POEs and a brief description of cost estimating techniques. The chapter then delved into the cost growth literature from RAND, IDA and AFIT. After that the chapter covered the predictor variables, the estimate factors, and the thresholds, while delving into the literature that discusses them. Then the chapter covered the OGCs. Finally, the chapter discussed specific cost estimating techniques used in estimating OGCs, delving into the analogy method, the engineering build-up method, and the parametric method.

# III. Methodology

## Chapter Overview

This chapter starts off with the source of the data, and then goes into the independent variables that are used in the analysis. After that the chapter covers the estimate factors that are used in the analysis along with the thresholds for those estimate factors. Then the chapter gets into the details of the database and the outlier programs. After this the chapter goes through the methodology of contingency tables and Fisher's Exact Tests. Contingency tables are used to set up the relationship between the program characteristics and different cost growth or cost variance thresholds, while the Fisher's Exact Tests test for statistical significance. Finally the chapter covers the Other Government Costs (OGC) and creating a possible Cost Estimating Relationship (CER) by testing distribution fit tests.

## Data Source

The data for this research comes from Air Force Life Cycle Management (LCMC) Program Office Estimates (POE). POEs are not as scrutinized or as standardized as Selected Acquisition Reports (SAR) and this limits the characteristics or variables that can be seen in both data sets. The POEs did not have a standardized schedule or Earned Value Management (EVM) data. Multiple programs were missing data that other programs did have.

One of the benefits of this data source is that it differentiates between programs that are ACAT II and ACAT III. SARs tend to be Acquisition Category (ACAT) I programs and analysis on ACAT II and ACAT III programs is hard to find. Previous research has analyzed programs that are either ACAT I or not ACAT I--there is no breakdown of non-ACAT I programs into ACAT II or ACAT III. This data source has programs categorized as ACAT II and ACAT III.

**Independent Variables**

The program characteristics from this database that will be independent variables are concurrency, prototyping and ACAT level. Each of these variables are formulated as binary variables.

Drezner (1993) said that concurrency is designed to speed up the schedule so that there is less time for potential mishaps to occur but it also decreases the time available for early testing.

Birchler (2011) described concurrency as RDT&E appropriations being authorized within the same years that production appropriations are authorized. Jimenez (2016) and Trudelle (2017) used this definition. Drezner (1993) measured concurrency as the overlap in months between the completion of IO&E and MS-3a. In other words if testing overlaps with initial production then concurrency is occurring. This research does not provide which appropriations are being authorized for each year and the Milestone 3a or the IO&E dates are unavailable. The data does shows how much Work-to-Go there is for Engineering Manufacturing Development (EMD), Procurement and MILCON. The EMD phase is equivalent to the Research, Development, Testing, and Evaluation (RDT&E) phase in other papers. If EMD is not at 100% complete while Procurement is above 0% complete then concurrency is occurring; otherwise it is not. Thus concurrency is defined as a binary variable.

Jimenez (2016) and Trudelle (2017) categorized programs that created a prototype, or prototypes, before the beginning of production as programs with prototyping. These authors used SAR data which specifically mention prototypes. This research did not find the word prototypes in the POE documentation. Instead, programs were categorized as having prototyping if there were EMD unit quantities. If a program has unit quantities in the EMD phase then that program is considered to have prototyping. If a program does not have unit quantities in the EMD Phase, then the program does not have prototyping.

The ACAT levels were derived using two methods. The first method used the original designation of the program--28 of the 50 viable programs had a designated ACAT level found in the documentation. The second method determined whether or not a program's latest EMD or Procurement costs were at or above the DoD ACAT thresholds—the other 22 programs were categorized using this method. Specifically, programs that have an EMD cost of $480 Million and above or have a Procurement cost of $2,790 Million and above are ACAT I programs. Programs that have an EMD cost of $185 Million to $479 Million or a Procurement cost of $835 Million to $2,789 Million are ACAT II programs. Programs that do not meet these thresholds are ACAT III programs. The ACAT thresholds are in 2014 Base Year (BY) dollars.

The research used inflation tables from the Secretary of the Air Force Financial Management (SAF/FM) website to convert the costs of each program from the individual Current Year (CY) dollars to BY 2014 dollars. Equation 3.1 was used:

$$\frac{\text{Target CY dollar amount}}{\text{BY 2014 Inflation Index for that CY}} \tag{3.1}$$

The EMD numbers were divided by the RDT&E Inflation Index, while the Procurement numbers were divided by the Aircraft, Space, and Missile Procurement Inflation Index.

In the event a program's EMD and procurement costs led to two different ACAT categorizations, the higher ACAT level was chosen. For example, if a program's EMD costs met the ACAT II threshold but its procurement costs met the ACAT I threshold, then it was considered an ACAT I program.

The estimates are just estimates and the data that directly links to actuals are the percent work-to-go (PWTG) numbers that tell the analyst what percentage of funds, or percentage of dollars, they have in their budget left to spend. The research concludes this to mean that actuals are used in the POE but POE also has estimates that are added to the actuals to provide the overall POE. This research understands PWTG to mean that as PWTG decreases a program uses more actuals to generate the overall estimate. The smaller the PWTG, the further along the program is. PWTG was transformed into percent complete (% Complete) by taking the PWTG from 1 or: 1 - PWTG = % Complete.

**Estimate Variance Factor**

This research defines two metrics to measure the amount a program office's estimate changes over time. The first is the Estimate Variance Factor (EVF) and its calculations is equation 3.2:

$$\frac{\text{Absolute Value (Target Estimate – Initial Estimate)}}{\text{Initial Estimate}} \tag{3.2}$$

This is an absolute value because whether a program has too much money or not enough causes problems of re-budgeting and the possible loss of projects that did not receive funds in a previous budget cycle. There are unfunded programs and projects every year and if funds are not available when they are needed then those projects are either forgotten or passed over to be done at a later date. When a program is passed over to be done later, it can cost more than its original price due to inflation, change in available technology, or other factors. Having programs go unfunded can also impact the mission by military members not having the most up-to-date technology or capabilities at their disposal. When cost estimates are more stable, then funds may be allocated more efficiently. Programs with too much money and good estimating can identify excess funds and can shift money to programs that need funds earlier rather than later.

However, not all analysts or researchers think in absolute value. Some focus on cost growth as more impactful than cost variance.

**Estimate Growth Factor**

The second factor used in this research is the Estimate Growth Factor (EGF) and its calculation is equation 3.3:

$$\frac{\text{Target Estimate} - \text{Initial Estimate}}{\text{Initial Estimate}} \qquad (3.3)$$

This is considered because the Nunn-McCurdy thresholds, which are used in this research, were created with the idea of cost growth not cost variance. The thresholds are discussed in the next section.

## Thresholds for Estimate Factors

This research evaluates the EVF and EGF factors relative to three different thresholds. Two are based off the Nunn-McCurdy significant and critical breach thresholds and the other threshold is "any cost growth". The Nunn-McCurdy thresholds are applied to both the EVFs and the EGFs. If a program experiences a cost growth of more than 30% from its initial estimate then it has breached the Nunn-McCurdy significant threshold. This research also looks at cost variances (overrun or underrun) that are more than 30% away from the initial estimate. The "cost growth" threshold focuses on cost growth and is only applied to the EGFs.

Nunn-McCurdy legislation was enacted in 1983 to notify congress when a Major Defense Acquisition Program (MDAP) overran its cost estimate by a designated percentage. A MDAP is a program estimated to have EMD costs greater than $480M or procurement costs greater than $2.79B, in FY2014 constant dollars (Schwartz, 2016). The overrun is considered significant if the program has experienced a 30% increase above its original cost estimate. The overrun is considered critical if the program has experienced a 50% increase above its original cost estimate (10 U.S.C. §2433).

The inclusion of the any cost growth threshold is summarized well by Kozlak: "The purpose of identifying programs with positive cost growth is to focus the attention of the estimator on the 'troubled programs'. If an estimator can identify predictors of positive cost growth in the troubled programs, they may determine a method to cut down total cost growth" (Kozlak, 2016).

**Database**

The dataset was initially comprised of 778 POE which consisted of 140 programs. These programs did not all have the necessary data for an analysis. To properly analyze the estimate factor as a function of percent complete, programs needed more than one POE that provided percent complete. Only 64 programs had 2 or more POEs with different percent complete values. 60 programs met this requirement. Costs were already in Constant Year (CY) dollars and these values were divided by unit quantities to determine the Acquisition Procurement Unit Cost (APUC) and Program Acquisition Unit Cost (PAUC) values for that POE. There were 50 programs that had quantities measured in either the EMD phase or the Procurement phase. Note, APUC only uses procurement data, while typical PAUC includes Engineering Manufacturing Development (EMD), Procurement and MILCON data.

The 50 programs considered did not have any MILCON data so the PAUC metric in this research is only on EMD and Procurement data--22 of these 50 programs had PAUC data but not APUC data. At the same time, 1 of the 50 programs had APUC data but not PAUC data because the procurement percent complete had changes but the EMD percent complete did not have changes. This caused a difference in the number of programs and POEs for PAUC and APUC. At this point PAUC had 49 programs and 222 POEs, while APUC had 28 programs and 127 POEs.

This analysis is trying to identify a trend in estimate factors as a function of percent complete. Percent complete was binned into 21 bins of 5% intervals from 0% to 100% with 2.5% breaks. The 0% bin included data with percent complete from 0% to

2.5%, while the 5% bin included data from 2.5% to 7.5%, etc. Each bin would be showing an estimate factor for applicable programs.

The initial APUC and PAUC POEs for a given program are not part of the data set because they provide a baseline from which subsequent POE change is measured. This left 99 POEs for APUC and 173 POEs for PAUC. In the interest of maintaining independence within a bin, the POEs that did not show a percent complete change more than 2.5% were averaged together for their appropriate bin. This resulted in 80 APUC POEs from 28 programs and 150 PAUC POEs from 49 programs.

These programs and POEs are all using EVFs and thus have positive values. Due to the Nunn-McCurdy thresholds being linked to cost growth, or positive EGFs, it is important to note the amount of negative EGFs in the dataset. There are 31 APUC POEs that have a negative EGF and there are 62 PAUC POEs that have a negative EGF. This reduces the number of APUC POEs by 39 percent when the outliers are included and 41 percent when the outliers are not included. This reduces the number of PAUC POEs by 41 percent when the outliers are included and 43 percent when the outliers are not included.

In the interest of providing data that is useful to leadership this analysis also breaks out the data by percent complete that is binned into 5 bins. These bins follow the percent complete mean and median values found by Kozlak (2017) for the five different program reviews: Critical Design Review (CDR), First Flight (FF), Development Test & Evaluation End (DT&E), Initial Operating Capability (IOC), and Last SAR (LS).

The mean percent complete bins are classified as the group 1 bin structure and the median percent complete bins are classified as the group 2 bin structure. The number of programs reduced to 18 for APUC group 1 and group 2. The number of programs reduced to 34 for group 1 and 33 for group 2. In the interest of maintaining independence within a bin, the POEs that did not show a percent complete change more than the associated percentage were averaged together for their appropriate bin. This averaging of POE values changed the naming convention within the associated tables and graphs from POEs to observations. The bin ranges for the group 1 bin structure and the group 2 bin structure are provided in Table 3.1 and 3.2 respectively:

**Table 3.1: Group 1 Bin Structure**

| Bin Range | 0-20 | 21-38 | 39-50 | 51-75.5 | 76.5-100 |
|---|---|---|---|---|---|
| Last % in Bin (Bin ID + Half the Distance between Previous Bin ID) | 20 | 38 | 50 | 75.5 | 100 |
| First % in Bin (Previous Last % in Bin + 1) | 0 | 21 | 39 | 51 | 76.5 |
| Bin ID | 13 | 27 | 49 | 51 | 100 |
| Distance between Previous Bin ID | | 14 | 22 | 2 | 49 |
| Half the Distance between Previous Bin ID | | 7 | 11 | 1 | 24.5 |

**Table 3.2: Group 2 Bin Structure**

| Bin Range | 0-18.5 | 19.5-34.5 | 35.5-46 | 47-74 | 75-100 |
|---|---|---|---|---|---|
| Last % in Bin (Bin ID + Half the Distance between Previous Bin ID) | 18.5 | 34.5 | 46 | 74 | 100 |
| First % in Bin (Previous Last % in Bin + 1) | 0 | 19.5 | 35.5 | 47 | 75 |
| Bin ID | 12 | 25 | 44 | 48 | 100 |
| Distance between Previous Bin ID | | 13 | 19 | 4 | 52 |
| Half the Distance between Previous Bin ID | | 6.5 | 9.5 | 2 | 26 |

Table 3.1 and Table 3.2 were formulated using the highlighted rows, which were provided in Kozlak's (2017) research. The ranges of each bin were based off of half the distance from the previous bin's designator (Bin ID).

The initial APUC and PAUC POEs for a given program are not part of the data set because they provide a baseline from which subsequent observation change is measured. This left 34 observations for APUC group 1 and 35 observations for APUC group 2. It also left 71 observations for PAUC group 1 and 73 observations for PAUC group 2. Table 3.3 summarizes the exclusion criteria for the dataset. Table 3.4 shows final number of programs and POEs in the 21 bin structure analysis. Table 3.5 shows final number of programs and observations in the group 1 and group 2 bin structure analyses:

**Table 3.3: Program Exclusion Criteria**

| Data | Taken Out | Result |
|---|---|---|
| POEs | | 778 |
| Individual Programs | | 140 |
| Programs that had 1 or 0 Percent Complete Observations | 76 | 64 |
| Programs that did not show a change from the initial Percent Complete to the Latest | 4 | 60 |
| Programs that did not have a quantity recorded in either EMD or Procurement | 10 | 50 |
| Note 1: There is 1 program that appears in APUC but not in PAUC because the APUC Percentages are more than 2.5% apart but when they are aggregated with the PAUC percentages they become less than 2.5% apart | | |
| Note 2: There are 22 programs that appear in PAUC but not in APUC because the EMD Percent Complete has seen a change but the Procurement Percent Complete has not | | |
| **POEs within Viable Programs** | **APUC** | **PAUC** |
| POEs | 127 | 222 |
| POEs not including the initial APUC or PAUC amounts | 99 | 173 |
| POEs with a Change in Percent Complete above 2.5 (Observations with less than 2.5 percent change were averaged together for their appropriate percent bin) | 80 | 150 |
| POEs that had an Estimate Variance Factor within 3 standard deviation above the mean | 76 | 144 |

**Table 3.4: Number of Viable Programs and POEs**

| Resulting POEs and Viable Programs with Outliers | APUC | PAUC |
|---|---|---|
| Programs | 28 | 49 |
| POE | 80 | 150 |
| **Resulting POEs and Viable Programs without Outliers** | **APUC** | **PAUC** |
| Programs | 26 | 47 |
| POE | 76 | 144 |

**Table 3.5: Number of Viable Programs and Observations (Group 1 and Group 2 bin structure)**

| Resulting Observations and Viable Programs with Outliers | APUC Grp 1 | PAUC Grp 1 | APUC Grp 2 | PAUC Grp 2 |
|---|---|---|---|---|
| Programs | 18 | 34 | 18 | 33 |
| Observations | 34 | 71 | 35 | 73 |
| **Resulting Observations and Viable Programs without Outliers** | **APUC Grp 1** | **PAUC Grp 1** | **APUC Grp 2** | **PAUC Grp 2** |
| Programs | 17 | 33 | 17 | 32 |
| Observations | 33 | 69 | 34 | 71 |

## Assumptions/Limitations

This research is assuming that the data is valid and correct. The dataset is limited to POEs that can be adjusted for quantity. Estimate factors were calculated using APUC and PAUC to adjust for quantity. The dataset is very sparse. It started out with 778 POEs but only 222 had valid PAUC data and only 127 had APUC data. As for the OGC data there were 83 viable POEs for EMD and 82 for Procurement. The dataset is of POEs not SARs and POEs only have estimates not actuals. This research is assuming that when a POE provides a PWTG value it is using actuals and estimates to provide the current estimate.

## Outlier Programs

Chebychev's rule states that no more than $1/k^2$ of a probability distribution can be supported at values further than k true standard deviations from the true mean (McClave, 2012). This infers that for a given distribution, 89% of data must fall within 3 standard deviations above the mean. The research found two programs with POE values higher than 3 standard deviations from the mean. With the outliers removed, the number of programs becomes 26 for APUC and 48 for PAUC. The number of POEs becomes 76 for APUC and 144 for PAUC.

The group 1 and group 2 bin structures use the same outlier programs mentioned earlier. These outliers cause a shift in the number of programs and the number of observations within the group 1 and group 2 bin structures. The number of programs reduced to 17 for APUC group 1 and group 2. The number of programs reduced to 33 for group 1 and 32 for group 2. This left 33 observations for APUC group 1 and 34 observations for APUC group 2. It also left 69 observations for PAUC group 1 and 71 observations for PAUC group 2.

One of the programs with an extreme EVF POE value was an ACAT III program while the other was an ACAT II program. The ACAT II program POE is 9.844 standard deviations from the mean. The ACAT III program POE is 6.804 standard deviations away from the mean. Neither of these programs had concurrency or prototyping.

**Contingency Tables and Fisher's Exact Tests**

This research seeks to determine if there is a correlation between certain program characteristics and the amount of estimate change as defined by EVF and EGF. Furthermore, the program characteristics are categorical and the EVF and EGF metrics can be categorized based their relationship to one of the three thresholds previously discussed. Thus, frequency counts can be constructed and statistical significance tested using contingency tables. Contingency tables are used to test whether or not independent variables (i.e., concurrency, prototyping, and ACAT level) can predict dependent variables (e.g. EGF exceeding the 25% Nunn-McCurdy significant breach threshold). The typical statistical test that is used for contingency tables is the Pearson test. Yet the Pearson test fails to provide a viable assessment when the sample size is small. Fisher's Exact Tests are geared to account for small sample sizes within contingency tables.

The Fisher's Exact Test first assumes all observations are independent which is true when considering a single bin. It also operates under the assumption that the counts in the contingency table are fixed, or conditioned. The test is assuming that the values are the population values or that the results are not going to be predictive for a different dataset. Fisher's Exact Test is distinguished from other statistical tests, with unconditioned rows and columns, due to its second assumption (McDonald, 2009). "A benefit of using Fisher's Exact Test is the test does not estimate the probability of a value; rather the test calculates the exact probability of receiving the observed data" (Kozlak, 2016).

This test provides a right-tailed, a left-tailed and a two tailed-test each with an associated p-value. The p-value of the two-tailed test only determines if the characteristic has a significant effect on the EVF or EGF threshold. The p-value of the right or left-tail denotes whether or not the program characteristic has a significant effect on the EVF or EGF being above the Nunn-McCurdy significant or critical breach thresholds and the EGF "cost growth" threshold. This research does not count a one-tailed test with a significant result as valid if associated the two-tailed test has an insignificant result. The significant threshold p-value for this research is 0.1 since this thesis is more exploratory than explanatory. The null hypothesis is that the program characteristic does not have a significant effect on the possibility of an EVF or EGF significant breach, critical breach, or EGF "cost growth". The two-tail test alternative hypothesis states that the program characteristic does have a significant effect on possibility of an EVF or EGF significant breach, critical breach, or EGF "cost growth". The left-tail test alternative hypothesis states that the probability of an EVF or EGF significant or critical breach is greater when the program characteristic is not present. The right-tail test alternative hypothesis states that the probability of an EVF or EGF significant or critical breach is greater when the program characteristic is present:

Ho: (Prob[Y =1] if X = 1) = (Prob[Y =1] if X = 0)

Two-Tailed Test Ha: (Prob[Y=1] if X = 1) =/= (Prob[Y=1] if X = 0)

Left-Tail Ha: (Prob[Y =1] if X = 0) > (Prob[Y =1] if X = 1)

Right-Tail Ha: (Prob[Y =1] if X = 1) > (Prob[Y =1] if X = 0)

where X = 1 when the program possesses a given characteristic and Y = 1 when the EVF or EGF exceeds the Nunn-McCurdy significant or critical breach thresholds. Suppose the two-tailed test for prototyping has a significant result. This means that for this dataset prototyping has a significant effect on an EVF or EGF breaching or not breaching the threshold. Another example is: suppose the right-tail test for prototyping being related to a critical EVF breach had a p-value less than 0.1; then the conclusion is that when prototyping is present, it is more likely that a critical EVF breach will occur. Recall, these tests are conducted within a given bin and thus are applicable only for the associated level of program completion.

The Fisher's Exact Tests look at the EVF, EGF and "cost growth" thresholds in different bins by percentage. The percentage is measuring the program's percent complete. Percent complete from the beginning of the program life cycle till the end is broken out into three different bin structures. The first bin structure is the 21 bin structure that starts with 0% complete and ends with 100% complete, while each bin is separated by 5% complete. The next 2 bin structures were implemented to give more context to the results. The research expects that leadership will appreciate bins that line up with program reviews as opposed to arbitrary 5% bins.

Kozlak (2017) looked at Fisher's Exact Tests for 5 different reviews within the program acquisition cycle: Critical Design Review (CDR), First Flight (FF), Development Test & Evaluation End (DT&E), Initial Operating Capability (IOC), and Last SAR (LS). These 5 reviews had different mean and median percent complete values associated with them. Group 1 bin structure, is associated with the mean percent complete values: CDR is at 13 percent, FF is at 27 percent, DT&E is at 49 percent, IOC is at 51 percent, and LS is at 100%. Group 2 bin structure, is associated with the median percent complete values: CDR is at 12 percent, FF is at 25 percent, DT&E is at 44 percent, IOC is at 49 percent, and LS is at 100%. These program reviews should be more valuable to leadership than the arbitrary 5% bins.

It is important to note that Kozlak's bins were created with aircraft in mind yet every program has CDR, FF, DT&E, IOC, and LS. FF may seem like an aircraft specific milestone but every program has a similar milestones where prototypes are tested for the first time or simulations are run for the first time (AcqNotes 2019).

**Other Government Costs Distribution**

LCMC is interested in understanding the ratio of OGC-to-contract costs in order to support the development of future cost estimates. OGCs are costs that a program office incurs in-house that are not part of a contract. These costs may include administrative equipment, travel costs for program office personnel, Analysis of Alternatives costs, etc. This research worked with LCMC analysts and found OGCs are typically estimated using analogous programs or historical actuals. The research did not find any estimating policies or research that would provide cost estimators or contracting officers a CER. In the interest of creating such a CER, this research looks at the distribution of the ratios of OGC-to-contract costs.

The OGC data is not as sparse as the APUC and PAUC data since the ratios only need OGC and contract costs and are not concerned with quantity. This means that instead of being broken out into APUC and PAUC, the ratios are broken out by EMD and Procurement costs. In EMD there were 114 programs with OGC data, and in procurement there were 111 programs with OGC data. The data was reduced since not all of these programs had POEs with contract costs data. In EMD there were 83 programs with OGC and contract costs data and in Procurement there were 82 programs with OGC and contract costs data. The ratios were calculated based off the latest POEs and thus there is 1 ratio per program.

All of the OGC-to-procurement cost ratios are below 1. There are 3 OGC-to-EMD cost ratios above 1; one of the ratios is 13.57. This is 11.7 more than the next highest ratio (1.87). The program is an ACAT I program, but this does not explain such a large ratio that other ACAT I programs are not experiencing in this dataset. There are 18

ACAT I programs with EMD data, and their ratios are all below 1 except for this major outlier. It is possible this outlier occurred because of an in-house expense that is not obviously connected to any of the program characteristics. The outlier was removed.

To get an idea of possible distributions, histograms of the ratios were generated. Figure 3.1 and Figure 3.2 display these histograms with the outlier program taken out.

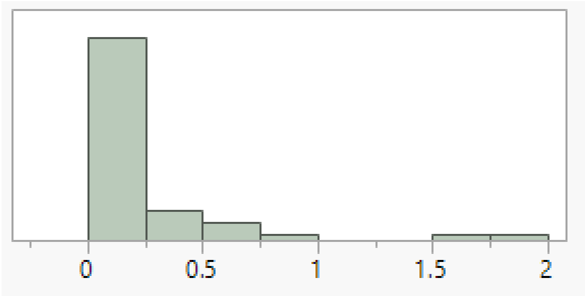



**Figure 3.1: EMD--OGC to Contract Ratio Distribution Without Outlier**

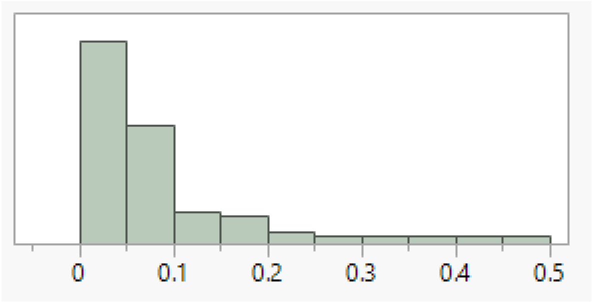



**Figure 3.2: Procurement--OGC to Contract Ratio Distribution Without Outlier**

These figures do not appear to fit a normal distribution. Based on the histograms, it appears as a distribution with a right-tail skew—such as an exponential distribution—may be an appropriate fit. This research will consider various distribution fits by conducting goodness of fit tests.

## Goodness of Fit Tests

When conducting a goodness of fit test for a distribution, there are multiple statistical tests and multiple distributions to test against. The tests and distributions used in this research were the first 8 distribution goodness of fit tests provided in JMP: normal, lognormal, Weibull, Weibull with threshold, extreme value, exponential, gamma, and beta. The p-value threshold is 0.05. JMP uses the Shapiro-Wilk W test for the normal distribution. JMP uses the Kolmogorov D test for the lognormal, exponential, and beta distributions. JMP uses the Cramer-von Mises W test for the Weibull, Weibull with threshold, extreme value, and gamma distributions. The null hypothesis of each test is that the distribution of the data fits the applicable distribution curve. The alternative hypothesis is that the distribution of the data is not a good fit for the applicable distribution curve. The summary statistics and a distribution of the OGC-to-contract costs are the building blocks for a CER.

## Summary

This chapter started off with the source of the data and then went into the independent variables that are used in the analysis. Next, the estimate factors used in the analysis were defined along with the thresholds for those estimate factors. Then the chapter discussed the details of database formulation and the outlier programs. After this the chapter went through the methodology of the analysis and the Fisher's Exact Tests. Finally the chapter covered the OGC and the related Kolmogorov's fit test.

# IV. Analysis and Results

## Chapter Overview

This chapter first reviews the Average Procurement Unit Cost (APUC) and Program Acquisition Unit Cost (PAUC) metrics with respect to the percent of program completion. Next, it reviews the results from the Fisher's Exact Tests and concludes with an analysis of Other Government Costs (OGC) to Contract ratios.

## PAUC and APUC Descriptive Statistics

Table 4.1 provides summary statistics for APUC and PAUC. It has sections for statistics that include the outliers and statistics the exclude the outliers. The table provides numbers to the graphs lines and shapes.

**Table 4.1: APUC and PAUC Summary Statistics**

| | APUC | PAUC | APUC Without Outliers | PAUC Without Outliers |
|---|---|---|---|---|
| Number of Programs | 28 | 49 | 26 | 48 |
| Number of Percent Complete Observations | 80 | 150 | 76 | 144 |
| Max Estimate Variance Factor | 9.495 | 6.555 | 1.117 | 1.796 |
| Mean Estimate Variance Factor | 0.454 | 0.331 | 0.217 | 0.272 |
| Median Estimate Variance Factor | 0.167 | 0.152 | 0.132 | 0.123 |
| Min Estimate Variance Factor | 0.0003 | 0.0001 | 0.0003 | 0.0001 |
| Standard Deviation | 1.3288 | 0.6322 | 0.2597 | 0.3262 |

Right away the maximum values for the APUC and PAUC jump out because they are so much larger than the mean or median values. The maximum values are also larger than 3 times the standard deviation. When the outliers are taken out the maximum values are much closer to the mean and median values. The minimum values are very small values compared to the other values in the table. Yet all the minimums are within 1 standard deviation from the mean. The standard deviations decrease without the outliers.

Tables 4.2 through 4.5 show the descriptive statistics of APUC Estimate Variance Factor (EVF) and APUC Estimate Growth Factor (EGF) broken out by ACAT level. Figures 4.1 through 4.4 show the ACAT level breakouts for APUC EVF and APUC EGF.

**Table 4.2: APUC EVF Descriptive Statistics by ACAT**

| ACAT | Max | Mean | Median | Min | Std Dev | Amount of POEs | Amount of Programs |
|---|---|---|---|---|---|---|---|
| ACAT I | 0.839 | 0.140 | 0.121 | 0.000 | 0.169 | 28 | 8 |
| ACAT II | 7.395 | 0.502 | 0.143 | 0.001 | 1.300 | 32 | 11 |
| ACAT III | 9.495 | 0.681 | 0.307 | 0.000 | 1.900 | 24 | 9 |

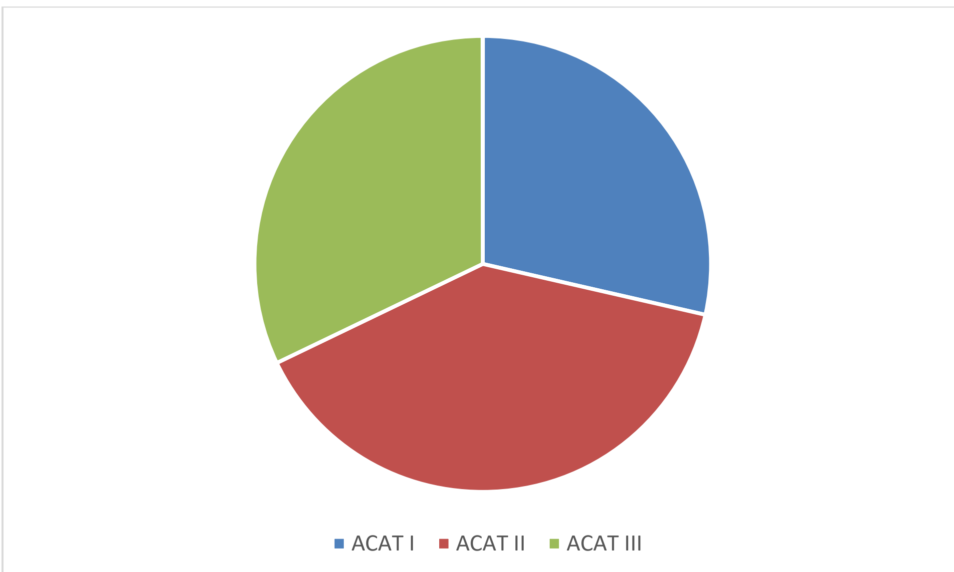



**Figure 4.1: ACAT Level Breakout for APUC EVF**

**Table 4.3: APUC EGF Descriptive Statistics by ACAT**

| ACAT | Max | Mean | Median | Min | Std Dev | Amount of POEs | Amount of Programs |
|---|---|---|---|---|---|---|---|
| ACAT I | 0.839 | -0.027 | -0.024 | -0.300 | 0.219 | 28 | 8 |
| ACAT II | 7.395 | 0.340 | 0.007 | -0.655 | 1.353 | 32 | 11 |
| ACAT III | 9.495 | 0.600 | 0.270 | -0.943 | 1.929 | 24 | 9 |

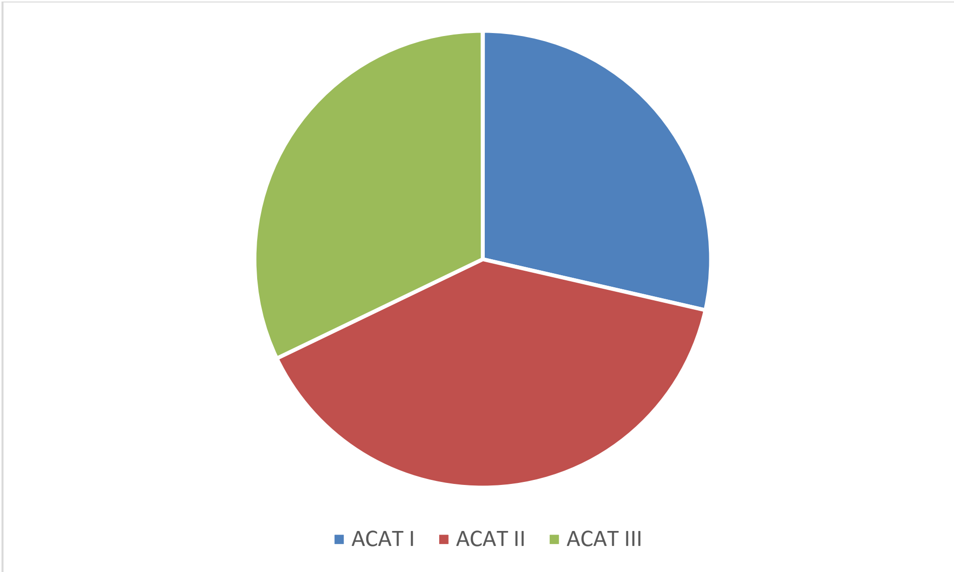



**Figure 4.2: ACAT Level Breakout for APUC EGF**

**Table 4.4: APUC EVF Descriptive Statistics by ACAT Without Outliers**

| ACAT | Max | Mean | Median | Min | Std Dev | Amount of POEs | Amount of Programs |
|---|---|---|---|---|---|---|---|
| ACAT I | 0.839 | 0.140 | 0.121 | 0.000 | 0.169 | 28 | 8 |
| ACAT II | 1.117 | 0.252 | 0.121 | 0.001 | 0.303 | 32 | 10 |
| ACAT III | 0.943 | 0.268 | 0.270 | 0.000 | 0.273 | 24 | 8 |

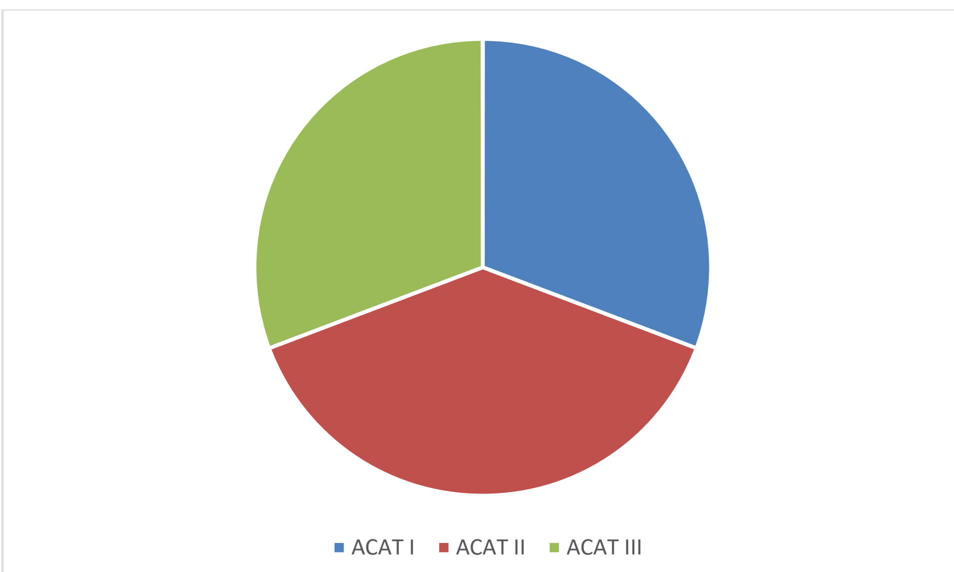



**Figure 4.3: ACAT Level Breakout for APUC EVF Without Outliers**

**Table 4.5: APUC EGF Descriptive Statistics by ACAT Without Outliers**

| ACAT | Max | Mean | Median | Min | Std Dev | Amount of POEs | Amount of Programs |
|---|---|---|---|---|---|---|---|
| ACAT I | 0.839 | -0.027 | -0.024 | -0.300 | 0.219 | 28 | 8 |
| ACAT II | 1.117 | 0.079 | 0.005 | -0.655 | 0.389 | 32 | 10 |
| ACAT III | 0.719 | 0.180 | 0.157 | -0.943 | 0.340 | 24 | 8 |

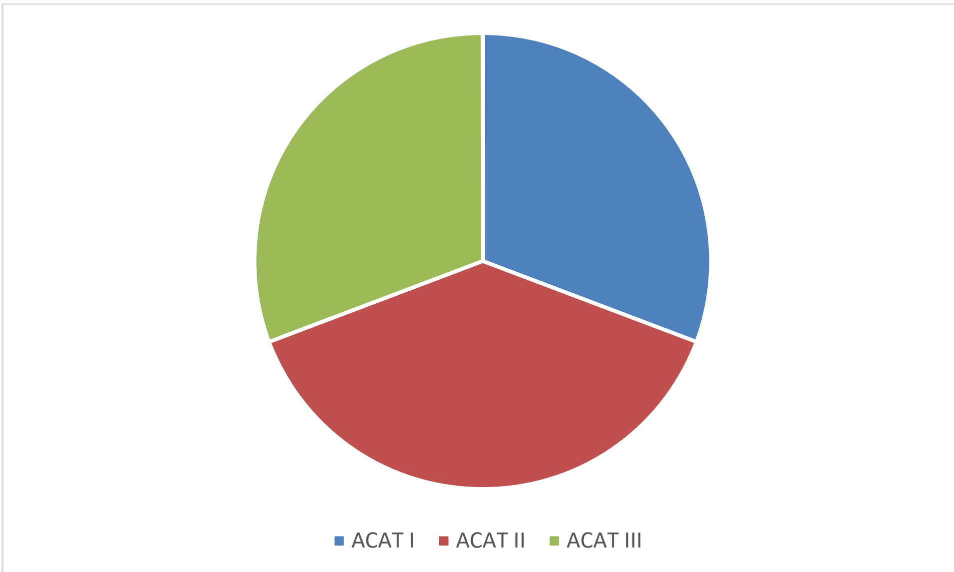



**Figure 4.4: ACAT Level Breakout for APUC EGF Without Outliers**

Tables 4.2 through 4.5 highlight that ACAT III programs have the highest median APUC EVF and APUC EGF. These tables also highlight ACAT II programs have the smallest impact on the cost. Tables 4.6 shows the descriptive statistics of PAUC EVF broken out by ACAT level. Figures 4.5 shows the ACAT level breakout for PAUC EVF. Figure 4.1 shows the number of programs per ACAT level to be pretty even (greatest difference in programs is 2), which is backed up by the data in Table 4.2. Figure 4.3 shows the number of programs per ACAT level to be pretty even (greatest difference in programs is 2), which is backed up by the data in Table 4.2. Figure 4.2 has the same values as Figure 4.1 while, Figure 4.4 has the same values as Figure 4.3.

**Table 4.6: PAUC EVF Descriptive Statistics by ACAT**

| ACAT | Max | Mean | Median | Min | Std Dev | Amount of POEs | Amount of Programs |
|---|---|---|---|---|---|---|---|
| ACAT I | 1.796 | 0.068 | -0.012 | -0.300 | 0.391 | 37 | 12 |
| ACAT II | 6.555 | 0.199 | 0.049 | -0.938 | 0.972 | 57 | 16 |
| ACAT III | 2.230 | 0.165 | 0.021 | -0.949 | 0.497 | 56 | 20 |

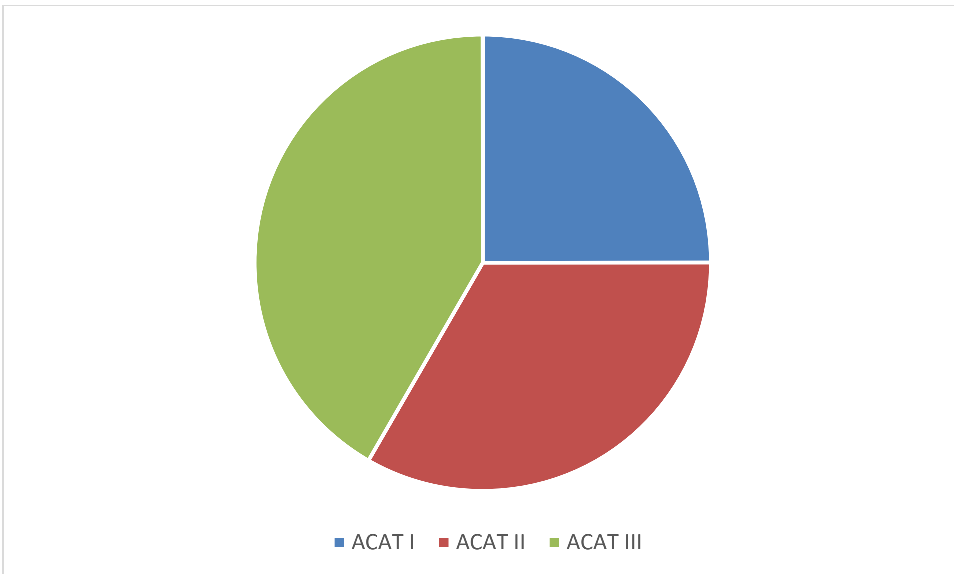



**Figure 4.5: ACAT Level Breakout for PAUC EVF**

Table 4.6 highlights that ACAT II programs have the highest median PAUC EVF. This table highlights ACAT III programs have the smallest impact on the cost. This would imply that EMD is affecting a change in impact to EVFs by ACAT, compared to the APUC EVFs. Figure 4.5 shows that ACAT IIIs occur most often, then ACAT IIs and ACAT Is show up the least, in this portion of the data. This is backed by the data in Table 4.6. Table 4.7 shows the descriptive statistics of PAUC EGF broken out by ACAT level. Figure 4.6 shows the ACAT level breakout for PAUC EGF.

**Table 4.7: PAUC EGF Descriptive Statistics by ACAT**

| ACAT | Max | Mean | Median | Min | Std Dev | Amount of POEs | Amount of Programs |
|---|---|---|---|---|---|---|---|
| ACAT I | 1.796 | 0.068 | -0.012 | -0.300 | 0.391 | 37 | 12 |
| ACAT II | 6.555 | 0.199 | 0.049 | -0.938 | 0.972 | 57 | 16 |
| ACAT III | 2.230 | 0.165 | 0.021 | -0.949 | 0.497 | 56 | 20 |

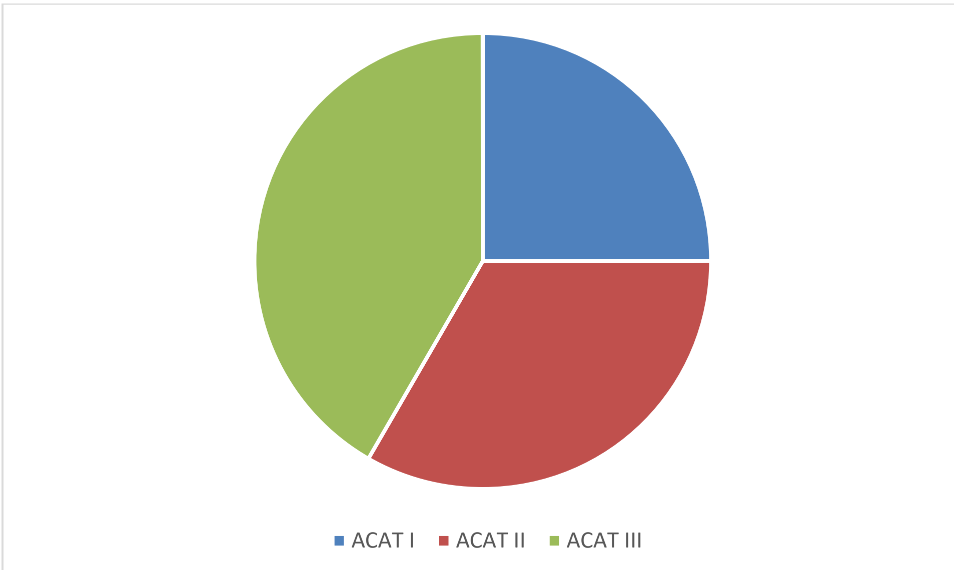



**Figure 4.6: ACAT Level Breakout for PAUC EGF**

Table 4.7 highlights that ACAT II programs have the highest median PAUC EGF. This table highlights ACAT I programs have the smallest impact on the cost. This would still imply that EMD is affecting a change in impact to EGFs by ACAT, compared to the APUC EGFs, even with the outlier programs removed. Figure 4.6 has the same values as Figure 4.5. This is backed by the data in Table 4.7. Table 4.8 shows the descriptive statistics of PAUC EVF broken out by ACAT level without the outlier programs. Figure 4.7 shows the ACAT level breakout for PAUC EVF without the outlier programs. Figure 4.6 has the same values as Figure 4.5.

**Table 4.8: PAUC EVF Descriptive Statistics by ACAT Without Outliers**

| ACAT | Max | Mean | Median | Min | Std Dev | Amount of POEs | Amount of Programs |
|---|---|---|---|---|---|---|---|
| ACAT I | 1.796 | 0.219 | 0.148 | 0.000 | 0.329 | 37 | 12 |
| ACAT II | 1.202 | 0.325 | 0.223 | 0.000 | 0.344 | 55 | 15 |
| ACAT III | 1.106 | 0.241 | 0.104 | 0.000 | 0.301 | 52 | 19 |

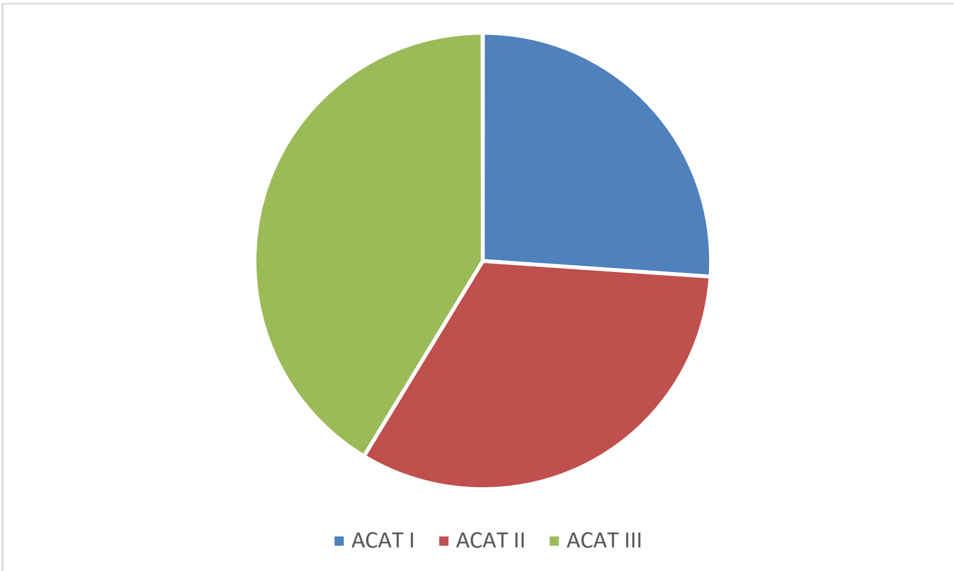



**Figure 4.7: ACAT Level Breakout for PAUC EVF Without Outliers**

Table 4.8 is similar to Table 4.6 in that it highlights that ACAT II programs have the highest median PAUC EVF. This table highlights ACAT III programs have the smallest impact on the cost. These results are very different from the other PAUC results but they do not match with the APUC results either. This would imply that EMD is affecting a change in impact to EVFs by ACAT, compared to the APUC EVFs. It also implies that removing the outlier programs impacted the results. Figure 4.7 shows that ACAT IIIs occur most often, then ACAT IIs and ACAT Is show up the least, in this portion of the data. This is backed by the data in Table 4.8. Table 4.9 shows the descriptive statistics of PAUC EGF broken out by ACAT level without the outlier programs.  Figure 4.8 shows the ACAT level breakout for PAUC EGF without the outlier programs.

**Table 4.9: PAUC EGF Descriptive Statistics by ACAT Without Outliers**

| ACAT | Max | Mean | Median | Min | Std Dev | Amount of POEs | Amount of Programs |
|---|---|---|---|---|---|---|---|
| ACAT I | 1.796 | 0.068 | -0.012 | -0.300 | 0.391 | 37 | 12 |
| ACAT II | 1.202 | 0.087 | 0.008 | -0.938 | 0.468 | 55 | 15 |
| ACAT III | 1.106 | 0.091 | 0.002 | -0.949 | 0.376 | 52 | 19 |

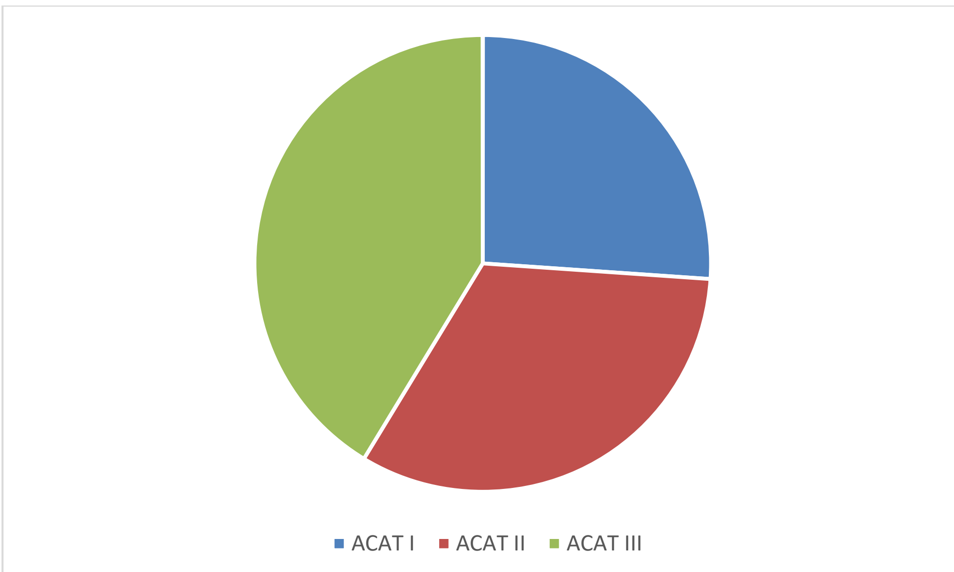



**Figure 4.8: ACAT Level Breakout for PAUC EGF Without Outliers**

Table 4.9 highlights that ACAT I programs have the highest median PAUC EGF. This table and figure highlights ACAT III programs have the smallest impact on the cost. This would imply that EMD is affecting a change in impact to EGFs by ACAT, compared to the APUC EGFs.  Figure 4.8 has the same values as Figure 4.7.  This is backed by the data in Table 4.9.

Figures 4.9 through 4.12 are the APUC and PAUC mean and median EVF graphs. The mean and median lines are in purple and are imposed over a histogram of the number of Program Office Estimates (POE) in each percent complete bin. The maximum (yellow) and minimum (orange) lines are also present to provide visual upper and lower bounds for the EVF values. All graphs show at least one point much higher than the others.

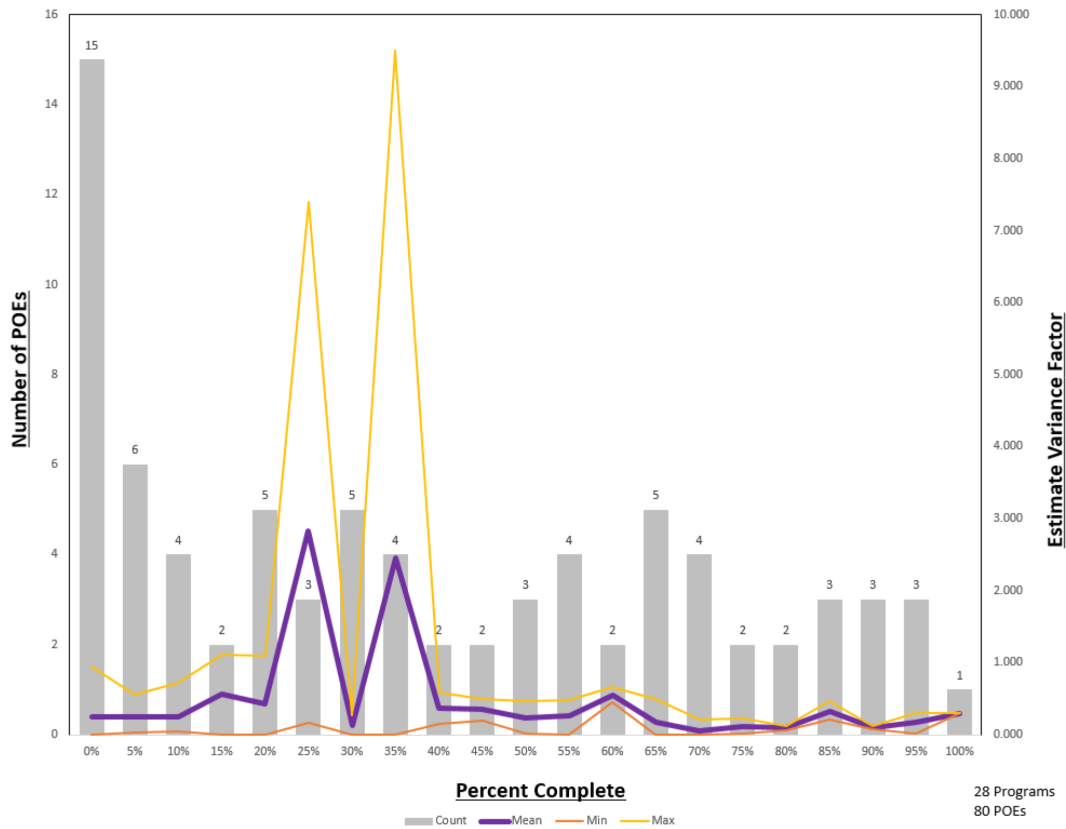



**Figure 4.9: APUC Mean Estimate Variance Factor**

Figure 4.9 has two major peaks in both the maximum line and the mean line. These really high peaks are in the 25% bin and the 35% bin. The minimum line is the most flat and has the least amount of peaks. The distance between the different lines starts off pretty small and then gets bigger where the peaks are and then the lines all converge in the 100% bin. The lines converge in the 100% bin because there is only one POE in the 100% bin.

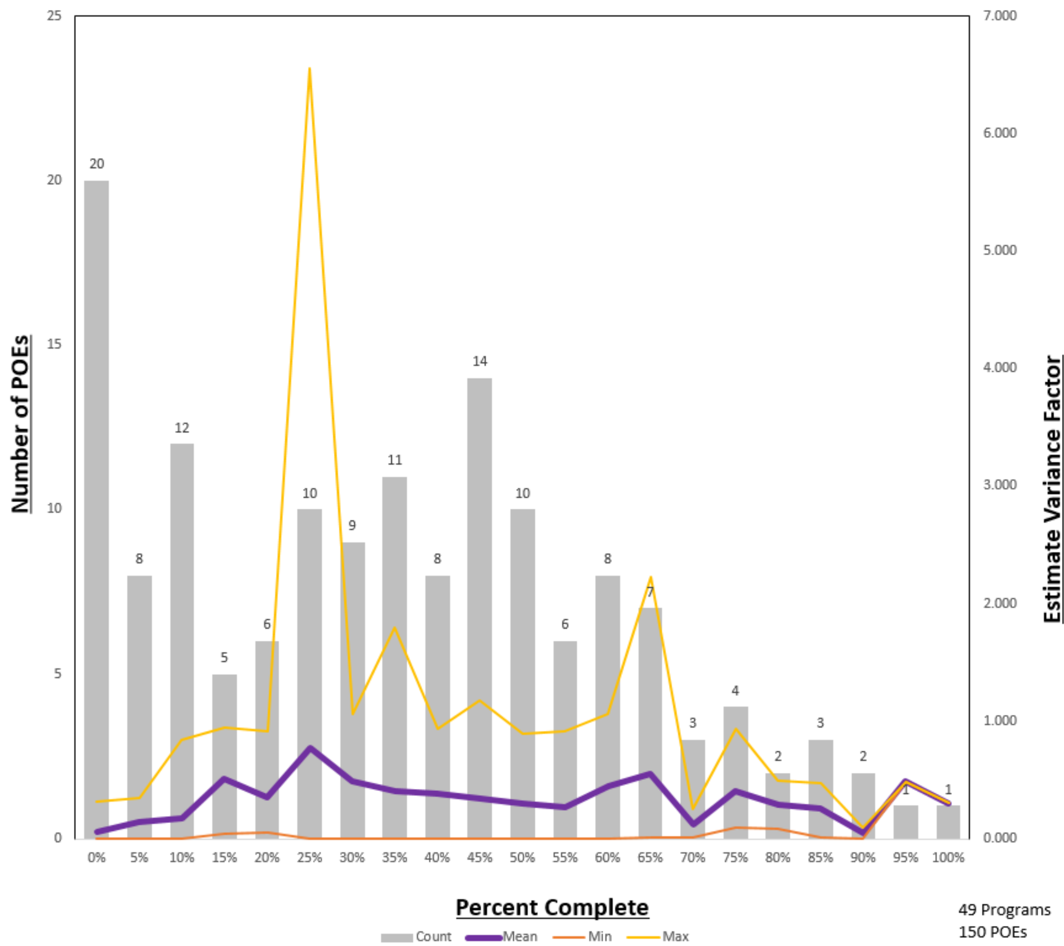



**Figure 4.10: PAUC Mean Estimate Variance Factor**

Figure 4.10 has one major peak in both the maximum line and the mean line. This peak is at the 25% bin. The mean line peaks are not as pronounced as the maximum line peaks. There is another peak in the 65% bin but it is not as clear a break from the mean line peak as the peak at the 25% bin. The minimum line is the most flat and has the least amount of peaks. The distance between the different lines starts off pretty small and then gets bigger where the peaks are and then the lines all converge in the 95% bin. The lines converge in the 95% bin because there is only one POE in the 95% and there is only one POE the 100% bin.

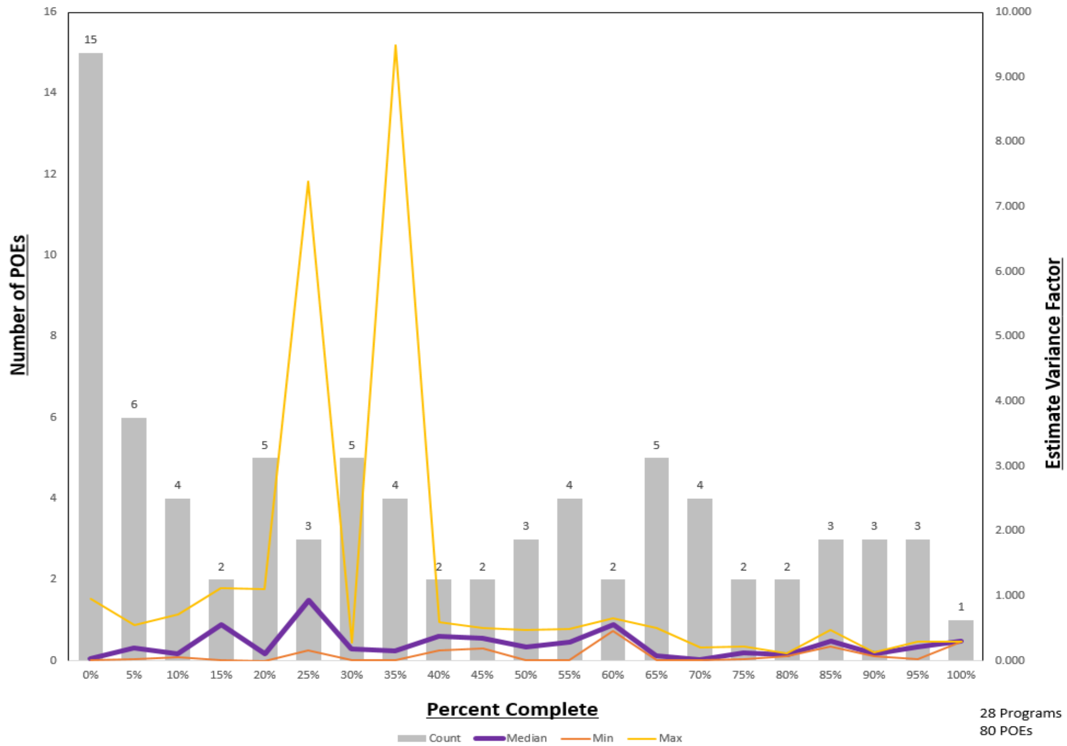



**Figure 4.11: APUC Median Estimate Variance Factor**

Figure 4.11 is very similar to Figure 4.9. Figure 4.11 has two major peaks in both the maximum line and the mean line. These really high peaks are in the 25% bin and the 35% bin. The minimum line is the most flat and has the least amount of peaks. The distance between the different lines starts off pretty small and then gets bigger where the peaks are and then the lines all converge in the 100% bin. The lines converge in the 100% bin because there is only one POE in the 100% bin. The big difference between Figure 4.11 and Figure 4.9 is that in Figure 4.11 the median line is much closer to the minimum line than the mean line in Figure 4.9.

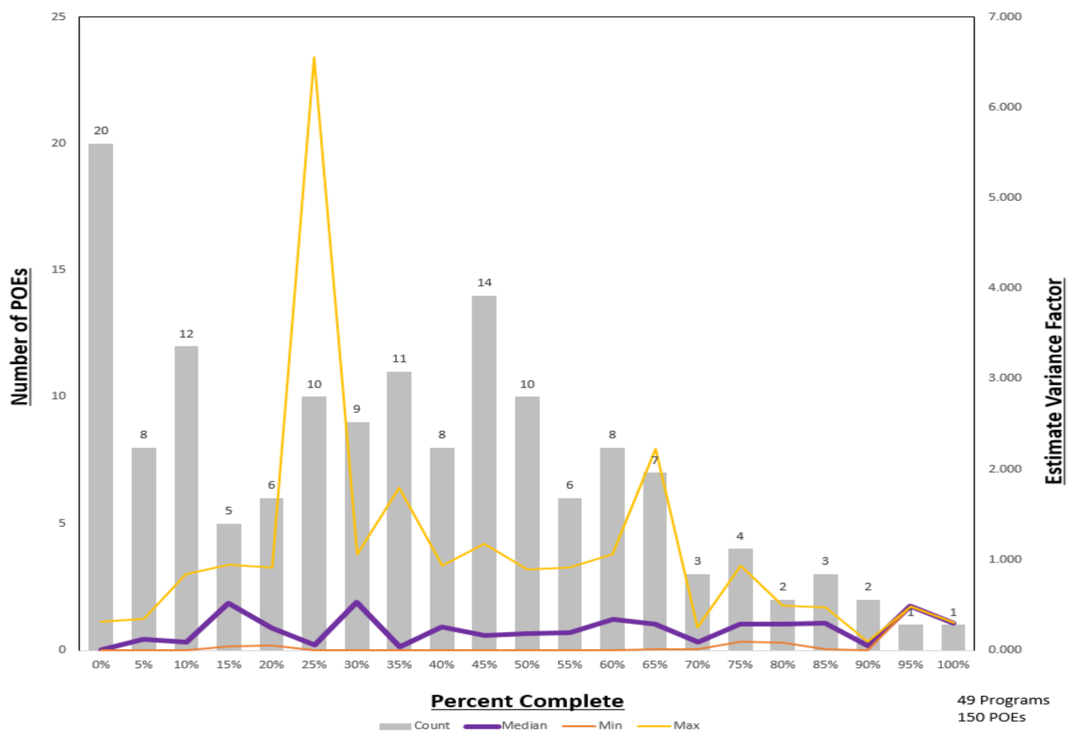



**Figure 4.12: PAUC Median Estimate Variance Factor**

Figure 4.12 is very similar to Figure 4.10. It has one major peak in both the maximum line and the median line. This peak is at the 25% bin. There is another peak in the 65% bin but it is not as clear a break from the mean line peak as the peak at the 25% bin. The median line has peaks in the 15%, 30%, and the 95% bins. The minimum line is the most flat and has the least amount of peaks. The distance between the different lines starts off pretty small and then gets bigger where the peaks are and then the lines all converge in the 95% bin. The lines converge in the 95% bin because there is only one POE in the 95% and there is only one POE the 100% bin. The big difference between Figure 4.12 and Figure 4.10 is that in Figure 4.12 the median line is much closer to the minimum line than the mean line in Figure 4.10.

Following Chebychev's rule that for a given distribution 89% of data must fall within 3 standard deviations of the mean. The research found two programs that were more than 3 standard deviations from the mean. Without the outliers the number of programs becomes 26 for APUC and 48 for PAUC. The number of POEs becomes 76 for APUC and 144 for PAUC. These points were taken out and the graphs were re-scaled.

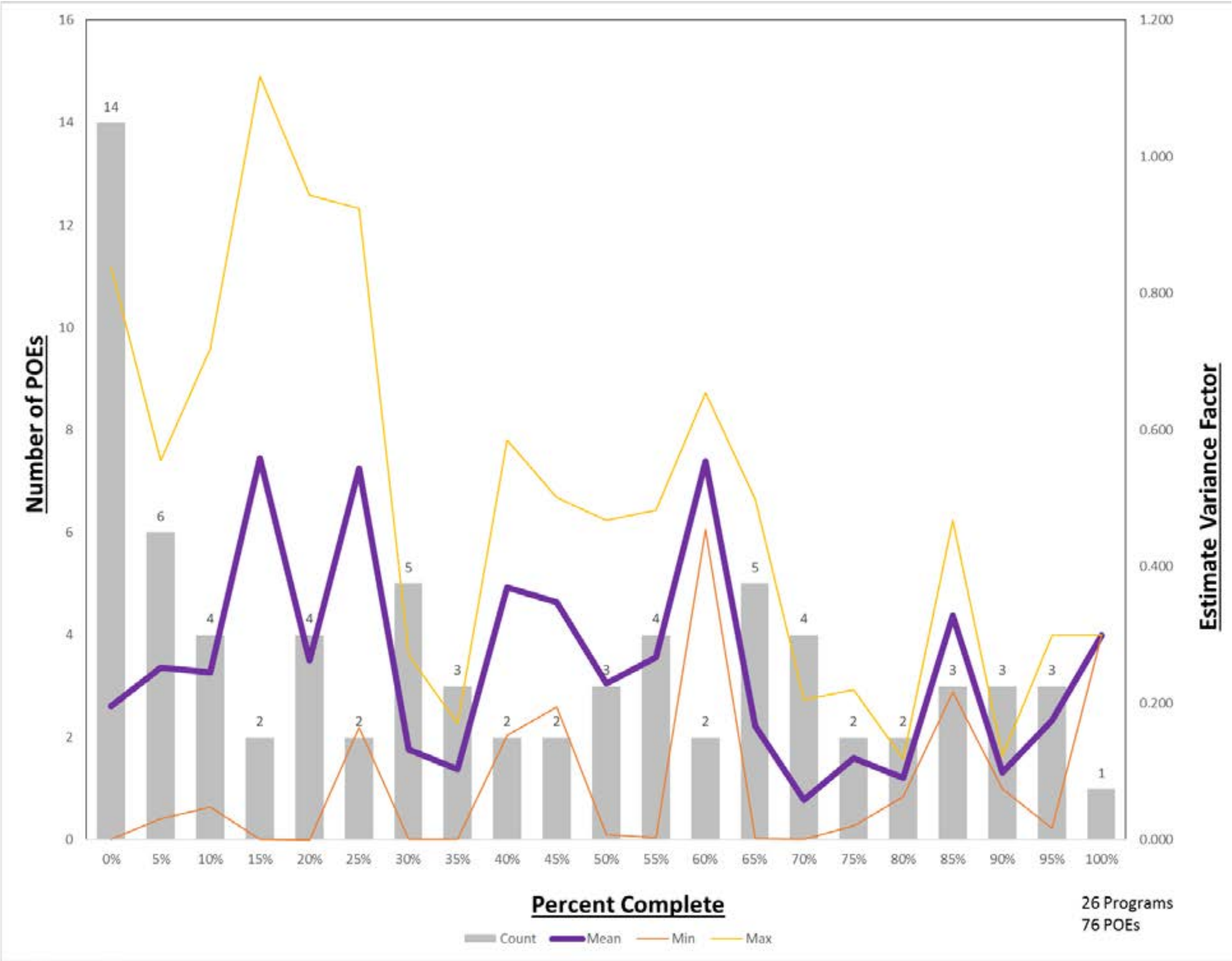



**Figure 4.13: APUC Mean Estimate Variance Factor (Without Outliers)**

Figure 4.13 has two major peaks in the mean line at the 15%, and the 25% bins. The maximum line does not have as many peaks but follows a similar basic shape to that of the mean line. The minimum line is also very similar in shape to the mean line but it does not match the peaks at the 15% or the 75% bins. The distance between the different lines starts off pretty broad and then gets smaller and smaller and then the lines all converge in the 100% bin. The lines converge in the 100% bin because there is only one POE in the 100% bin.

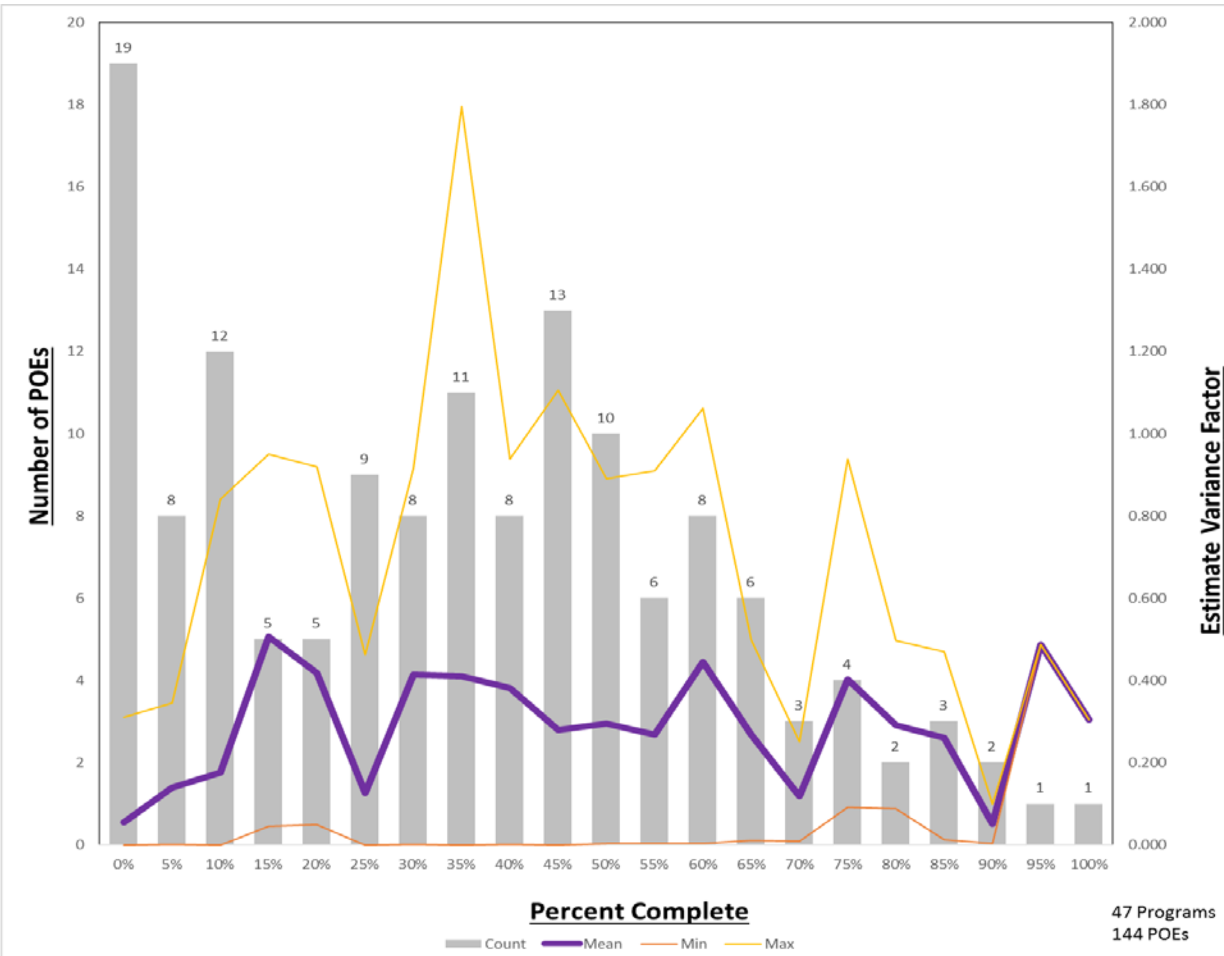



**Figure 4.14: PAUC Mean Estimate Variance Factor (Without Outliers)**

Figure 4.14 has one major peak in the maximum line at the 35% bin. The mean line does not have any major peaks. The minimum line is very flat with the least amount of peaks. The distance between the different lines starts off pretty small, gets bigger in the middle and then the lines all converge in the 95% bin.  The lines converge in the 95% bin because there is only one POE in the 95% and there is only one POE the 100% bin.

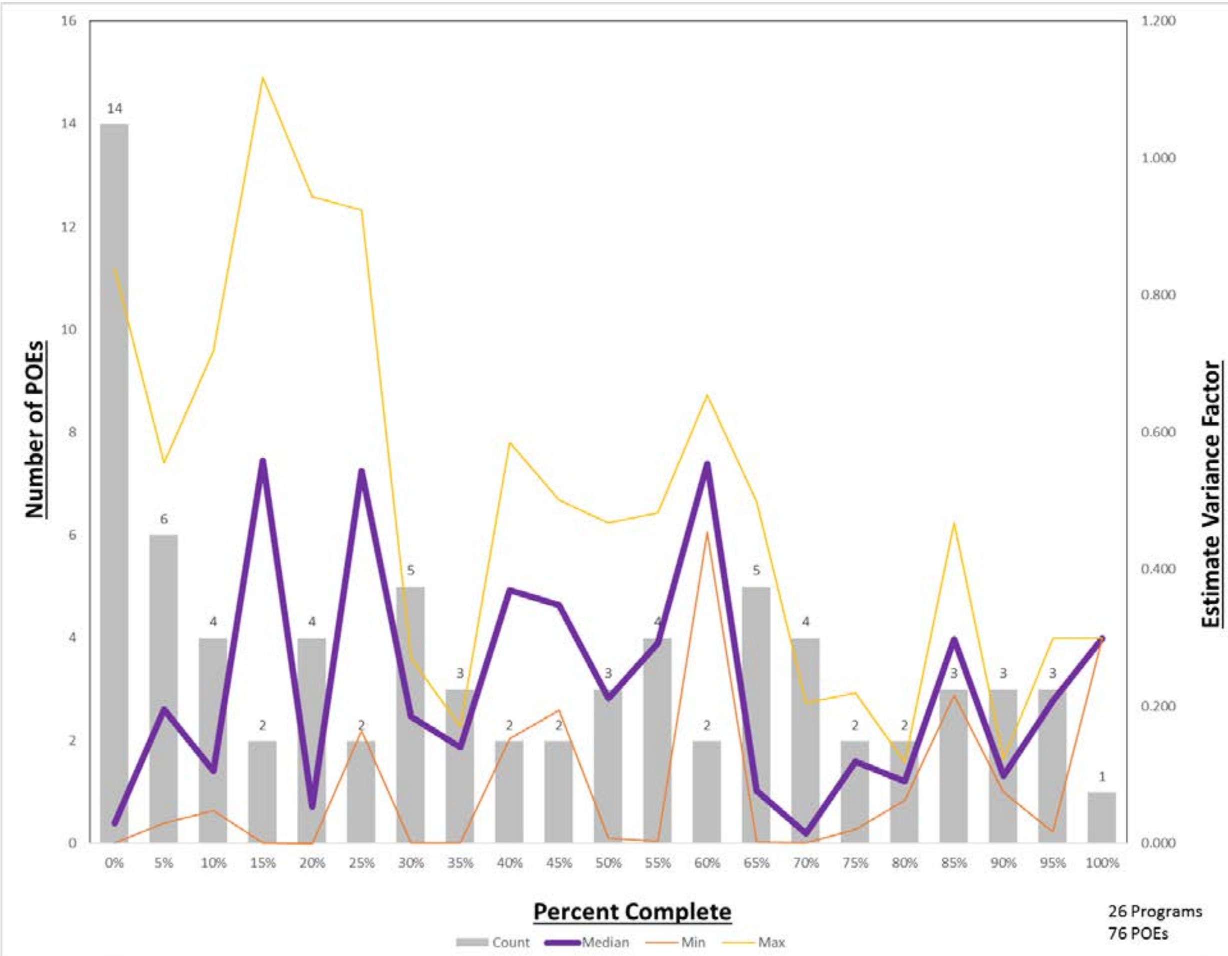



**Figure 4.15: APUC Median Estimate Variance Factor (Without Outliers)**

Figure 4.15 is very similar to Figure 4.13.  Figure 4.15 has three major peaks in the mean line at the 15%, and the 25% bins.  The maximum line does not have as many peaks but follows a similar basic shape to that of the mean line.  The minimum line is also very similar in shape to the mean line but it does not match the peaks at the 15% or the 75% bins.  The distance between the different lines starts off pretty broad and then gets smaller and smaller and then the lines all converge in the 100% bin.  Again this is due to their only being one data point in the 100% bin. There is one difference between Figure 4.13 and Figure 4.15. In Figure 4.15 the median line is closer to the minimum line than the mean line in Figure 4.13. The average distance from the median points to the minimum points is 0.1452 while, the average distance from the mean points to the minimum points is 0.1738.

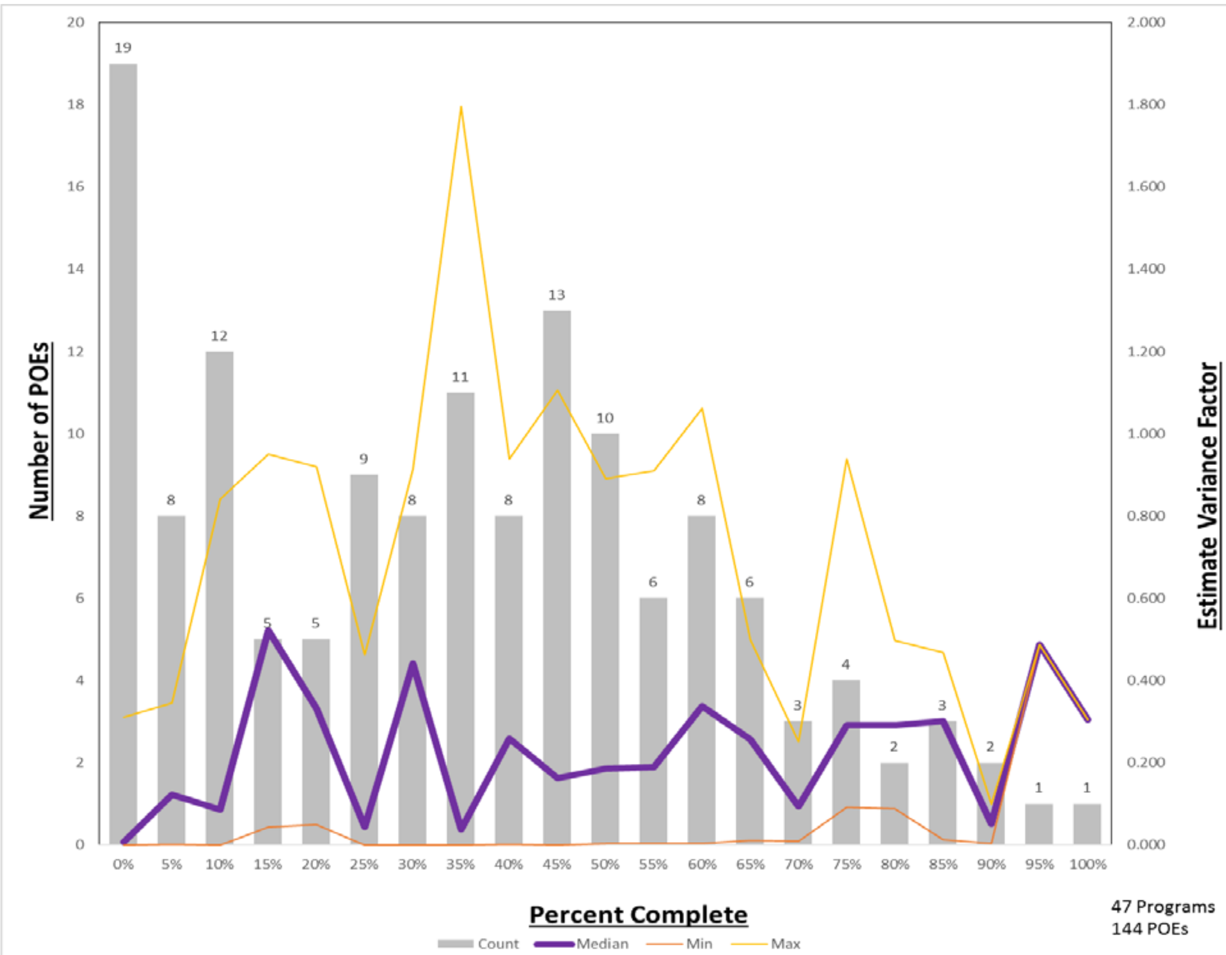



**Figure 4.16: PAUC Median Estimate Variance Factor (Without Outliers)**

Figure 4.16 has one major peak in the maximum line at the 35% bin. The mean line does not have any major peaks. The minimum line is very flat with the least amount of peaks. The distance between the different lines starts off pretty small, gets bigger in the middle and then the lines all converge in the 95% bin. Again this is due to their only being one data point in the 95% bin and one data point in the 100% bin. There are two differences between Figure 4.16 and Figure 4.14. In Figure 4.16 the median line is closer to the minimum line than the mean line in Figure 4.14. The average distance from the median points to the minimum points is 0.1755 while, the average distance from the mean points to the minimum points is 0.2375. The median line in Figure 4.16 also has more ups and downs than the mean line in Figure 4.14. These ups and downs are in the 30% through the 40% bins.

With the outliers being taken out Figures 4.13 through 4.16 more clearly show how varied the data is. Yet Figures 4.13 through 4.16 still have a bin that contains only one data point. The next group of figures breaks the data into 5 bins following Kozlak's (2017) mean and median percent completes for the phases he observed in his data: Critical Design Review (CDR), First Flight (FF), Development Test & Evaluation End (DT&E), Initial Operating Capability (IOC), and Last Selected Acquisition Report (LS).

**APUC and PAUC Percent Completes by Program Review**

Kozlak's (2017) mean percent complete for 5 major milestone, which is referred to as the group 1 bin structure in this research, are: CDR is at 13 percent, FF is at 27 percent, DT&E is at 49 percent, IOC is at 51 percent, while LS is at 100%;the median percent completes for the same 5 milestones, which is referred to as the group 2 bin structure in this research, are: CDR is at 12 percent, FF is at 25 percent, DT&E is at 44 percent, IOC is at 49 percent, while LS is at 100%.

Table 4.10 provides summary statistics for APUC and PAUC Kozlak bins with the outlier programs included.

**Table 4.10: APUC and PAUC Summary Statistics Kozlak Bins**

| | APUC Grp 1 | APUC Grp 2 | PAUC Grp 1 | PAUC Grp 2 |
|---|---|---|---|---|
| Number of Programs | 18 | 18 | 34 | 33 |
| Number of Percent Complete Observations | 34 | 35 | 71 | 73 |
| Max Estimate Variance Factor | 6.509 | 6.509 | 1.260 | 1.796 |
| Mean Estimate Variance Factor | 0.447 | 0.438 | 0.349 | 0.360 |
| Median Estimate Variance Factor | 0.220 | 0.206 | 0.261 | 0.262 |
| Min Estimate Variance Factor | 0.008 | 0.008 | 0.001 | 0.002 |
| Standard Deviation | 1.097 | 1.082 | 0.331 | 0.368 |

The outlier programs cause the maximum APUC EVF to jump out again because it is more than 3 times the standard deviation away from the mean. The value in question is actually 5.527 standard deviations away from the mean. The outlier programs in PAUC do not have an impact on the maximum EVF. In other words the maximum EVF does not change when the outlier programs are taken out. The outlier programs are based on a program to program analysis (Table 4.1) not on Table 4.10. Table 4.10 is the descriptive statistics of the Kozlak (2017) bin breakdown. Table 4.11 provides summary statistics for APUC and PAUC Kozlak bins without the outlier programs.

**Table 4.11: APUC and PAUC Summary Statistics Kozlak Bins (Without Outliers)**

| | APUC Grp 1 | APUC Grp 2 | PAUC Grp 1 | PAUC Grp 2 |
|---|---|---|---|---|
| Number of Programs | 17 | 17 | 33 | 32 |
| Number of Percent Complete Observations | 33 | 34 | 69 | 71 |
| Max Estimate Variance Factor | 0.942 | 0.942 | 1.260 | 1.796 |
| Mean Estimate Variance Factor | 0.263 | 0.259 | 0.343 | 0.354 |
| Median Estimate Variance Factor | 0.206 | 0.194 | 0.261 | 0.262 |
| Min Estimate Variance Factor | 0.008 | 0.008 | 0.001 | 0.002 |
| Standard Deviation | 0.240 | 0.237 | 0.323 | 0.362 |

With the outlier programs removed the maximum for APUC decreases by 5.567 in both group 1 and group 2 but as stated earlier there is no change to the maximum

PAUC values. The APUC mean EVF in group 1 decreases by 0.184 and by 0.179 in group 2. The PAUC mean EVFs for both group 1 and group 2 decrease by 0.006. The APUC median EVF in group 1 decreases by 0.013 and by 0.012 in group 2. The PAUC median EVFs for both group 1 and group 2 do not change. The minimum EVFs for APUC and PAUC do not change, regardless of group. The APUC standard deviation in group 1 decreases by 0.857 and by 0.845 in group 2. The PAUC standard deviation in group 1 decreases by 0.008 and by 0.006 in group 2. The values that experienced the most change were the maximum APUC values.

Figures 4.17 through 4.18 are the APUC and PAUC group 1 and group 2 EVF graphs. The mean line is in purple. The median line is in blue. The maximum (yellow) and minimum (orange) lines are also present to provide visual upper and lower bounds for the EVF values. These lines are imposed over a histogram of the number of POEs in each percent complete bin.

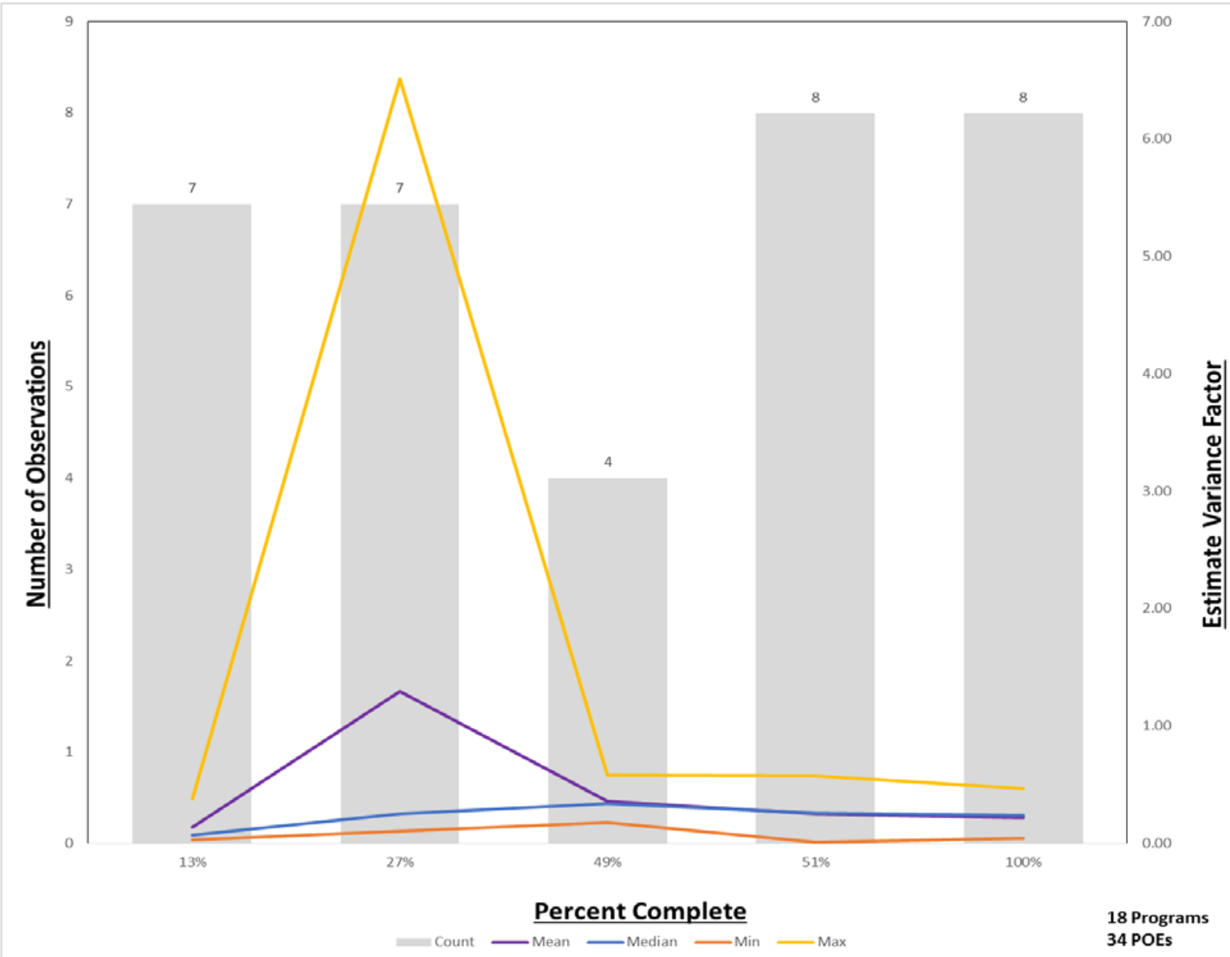



**Figure 4.17: APUC Grp 1 Estimate Variance Factor**

Figure 4.17 shows that in the 27 percent bin the maximum point is 5.22 from the mean point, which is 1.04 from the median point, which is 0.15 from the minimum point. After the spike in the 27 percent bin the lines stay pretty constant. The maximum and median lines have an average change of about 0.05, while the minimum and mean lines have an average change of about 0.07. The mean and median lines cross each other in the 51 percent bin and are only an average of 0.22 away from each other. The median line experiences an overall increase of 0.1677 from the 13 percent bin to the 100 percent bin. The mean line experiences an overall increase of 0.0776 from the 13 percent bin to the 100 percent bin.

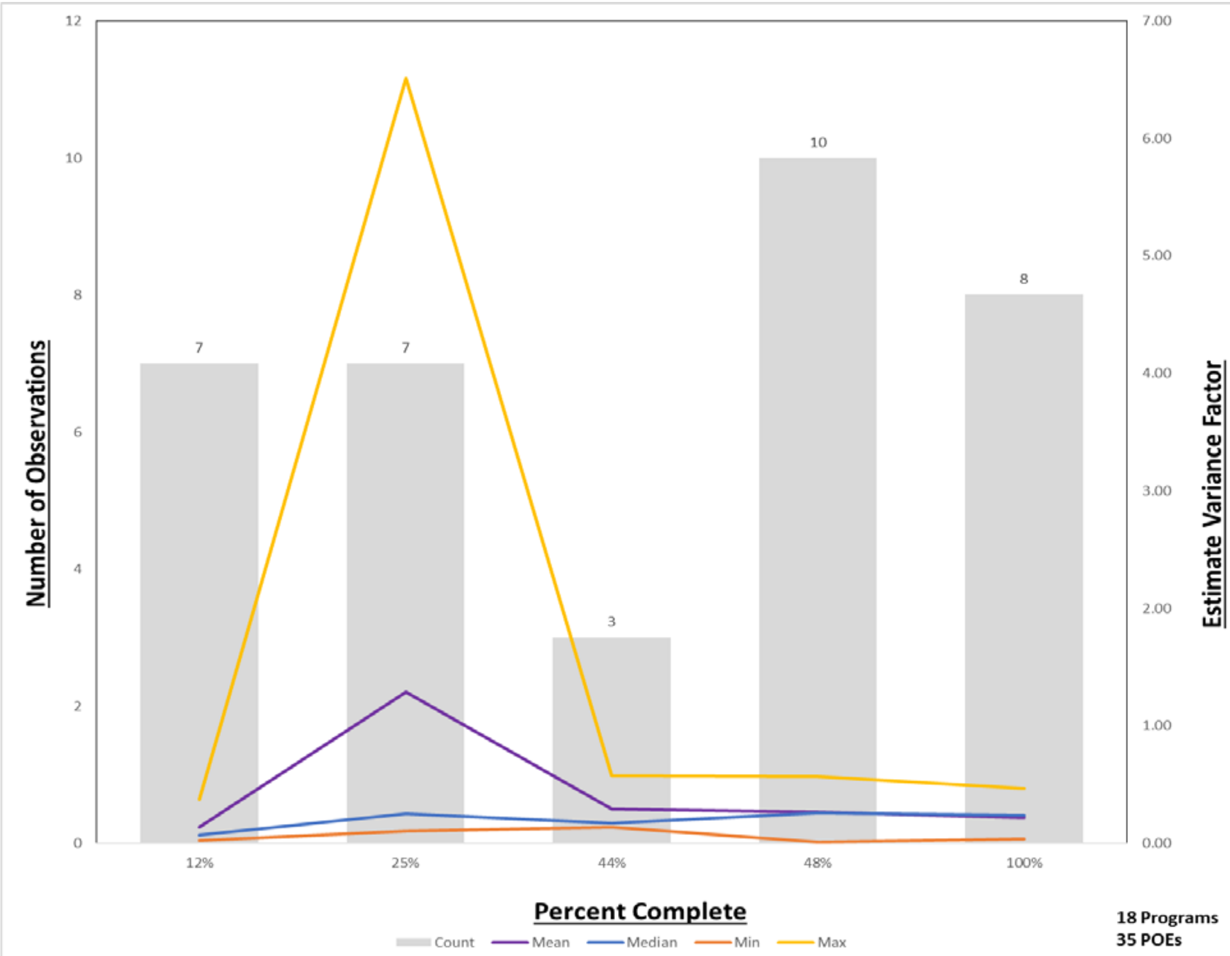



**Figure 4.18: APUC Grp 2 Estimate Variance Factor**

Figure 4.18 shows that in the 25 percent bin the maximum point is 5.22 from the mean point, which is 1.04 from the median point, which is 0.15 from the minimum point. After the spike in the 25 percent bin the lines stay pretty constant. The maximum and minimum lines have an average change of about 0.05. The median experiences an average change of about 0.03 while the mean experiences an average change of about 0.04. The mean and median lines cross each other after the 48 percent bin and are only an average of 0.24 away from each other. The median line experiences an overall increase of 0.1677 from the 12 percent bin to the 100 percent bin. The mean line experiences an overall increase of 0.0752 from the 12 percent bin to the 100 percent bin.

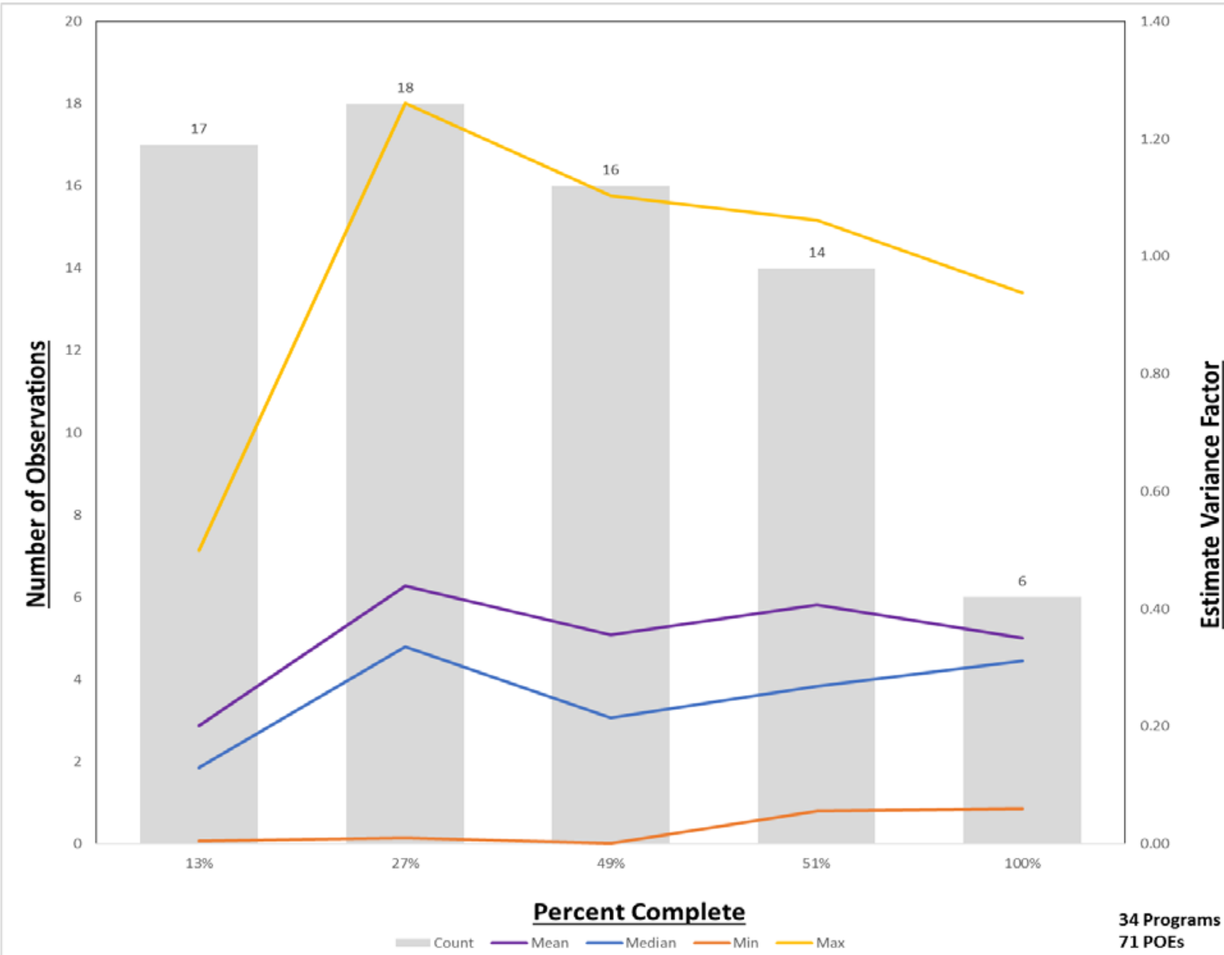



**Figure 4.19: PAUC Grp 1 Estimate Variance Factor**

Figure 4.19 shows that the maximum line starts with a spike, or increase of 0.76, in the 27 percent bin and then a steady drop off till the 100 percent bin. The minimum line experiences its slight spike in the 51 percent bin, with an increase of 0.055. Yet the minimum line only experiences an average change of 0.01 throughout all the bins. The median and mean lines have ups and downs but, do not vary too much. The median experiences an average change of about 0.05 while the mean experiences an average change of about 0.04. The mean and median lines are only an average 0.1 away from each other. The median line experiences an overall increase of 0.1825 from the 13 percent bin to the 100 percent bin. The mean line experiences an overall increase of 0.1498 from the 13 percent bin to the 100 percent bin.

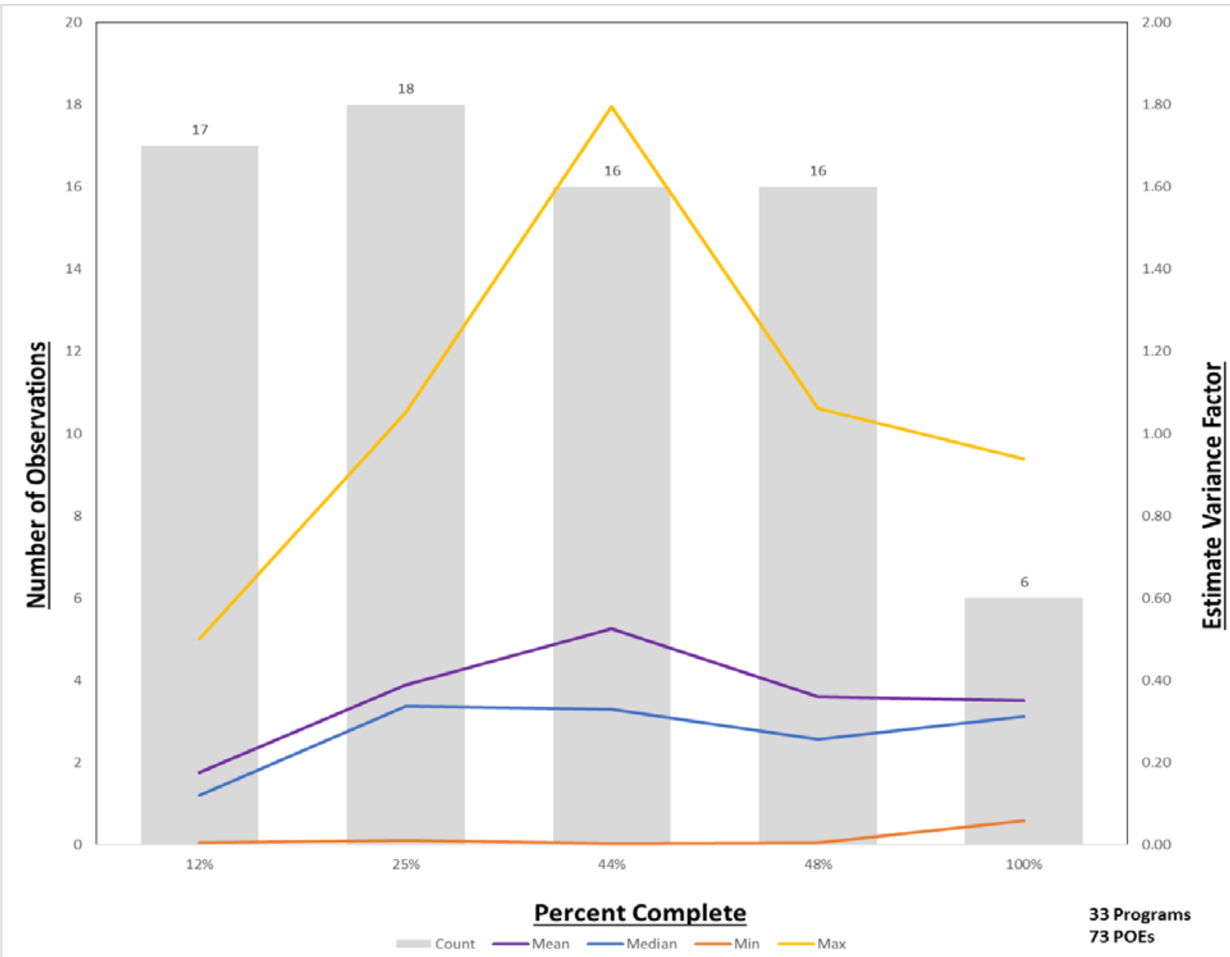



**Figure 4.20: PAUC Grp 2 Estimate Variance Factor**

Figure 4.20 shows that the maximum line starts with a climb from the 12 percent bin to the 25 percent bin, an increase of 0.55, and then peaks at the 44 percent bin, after an increase of 0.74. It decreases down, by 0.73, to the 48 percent bin and then even further to the 100 percent bin. The minimum line experiences its spike in the 100 percent bin, with an increase of 0.056. Yet the minimum line only experiences an average change of 0.01 throughout all the bins. The median and mean lines follow a similar path as the maximum line but, do not vary too much. The median experiences an average change of about 0.05 while the mean experiences an average change of about 0.04. The mean and median lines are only an average 0.1 away from each other. The median line experiences an overall increase of 0.1907 from the 12 percent bin to the 100 percent bin. The mean line experiences an overall increase of 0.1753 from the 12 percent bin to the 100 percent bin.

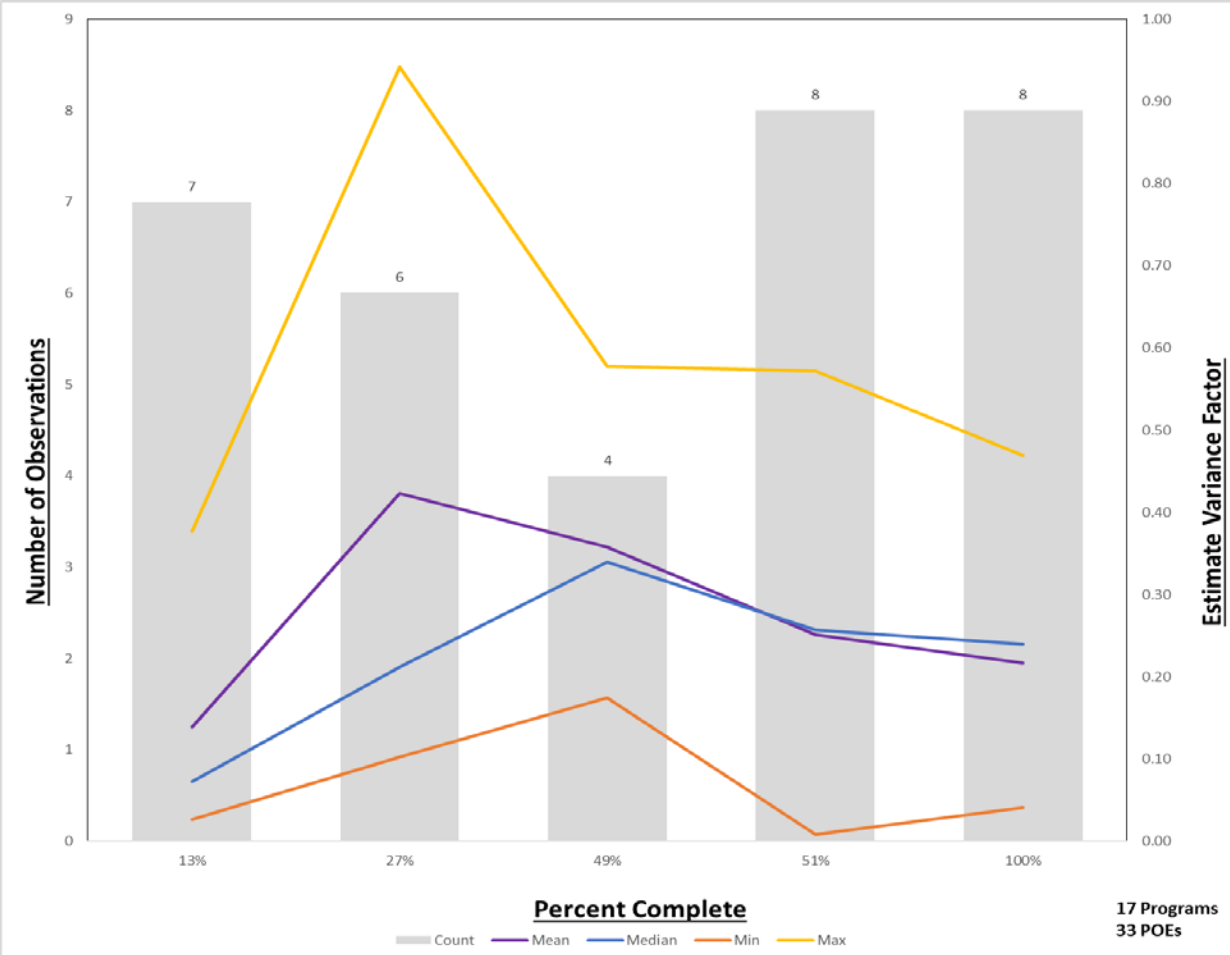



**Figure 4.21: APUC Grp 1 Estimate Variance Factor (Without Outliers)**

Figure 4.21 is basically a rescaled version of Figure 4.17 with a lower peak in the 27 percent bin. The peak went from 6.5 to 0.9. The 27 percent bin now shows the maximum point as 0.52 from the mean point, which is 0.21 from the median point, which is 0.11 from the minimum point. Now the mean and median lines are only an average of 0.05 away from each other; after the spike in the 27 percent bin. The rest of the values are the same as those in Figure 4.17.

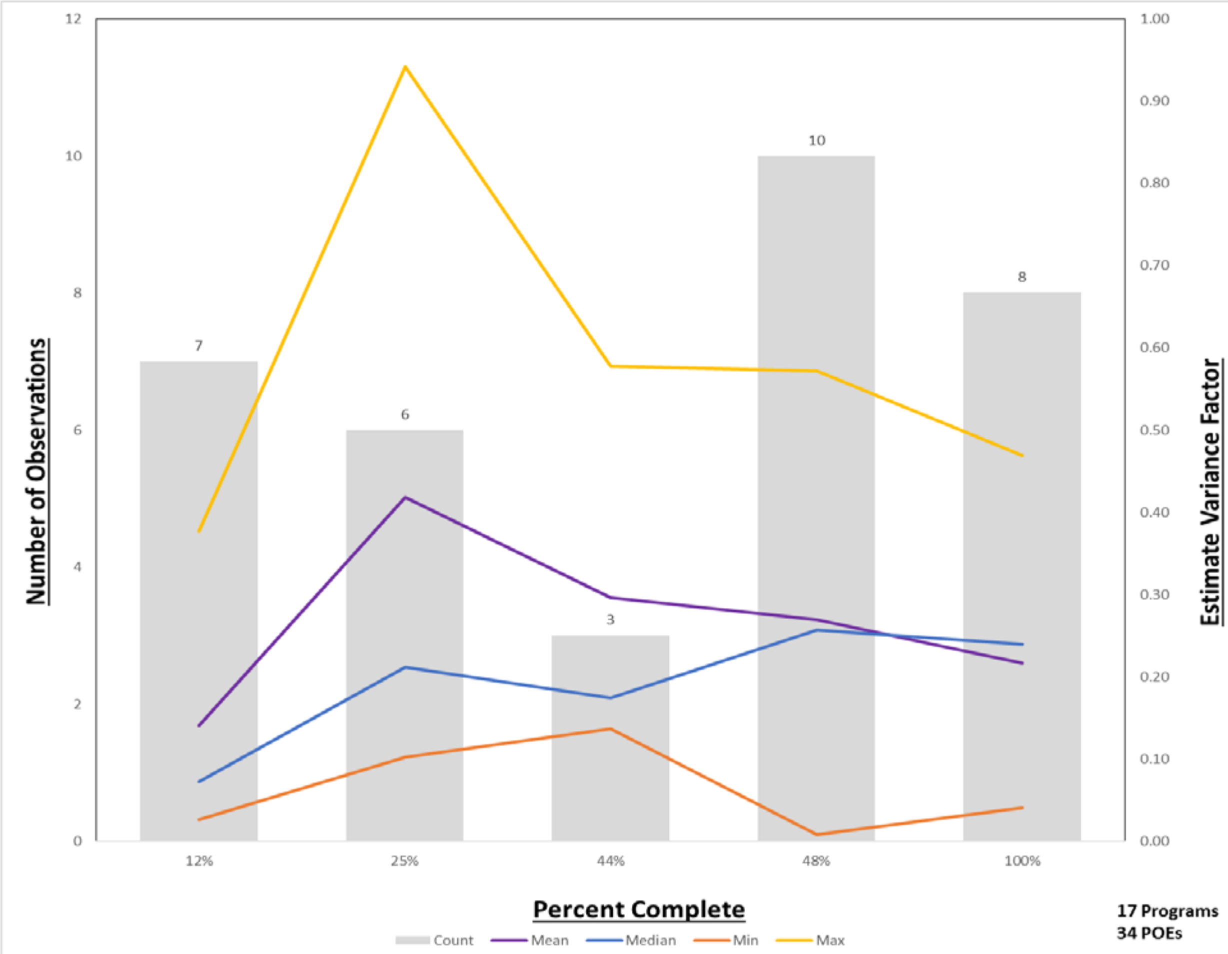



**Figure 4.22: APUC Grp 2 Estimate Variance Factor (Without Outliers)**

Figure 4.22 is a rescaled version of Figure 4.18 with the outlier program removed. Note, the maximum value in the 25 percent bin went from 6.5 to 0.9. The 25 percent bin now shows the maximum point as 0.52 from the mean point, which is 0.21 from the median point, which is 0.11 from the minimum point. Now the mean and median lines are only an average of 0.08 away from each other; after the spike in the 25 percent bin. The rest of the values are the same as those in Figure 4.18.

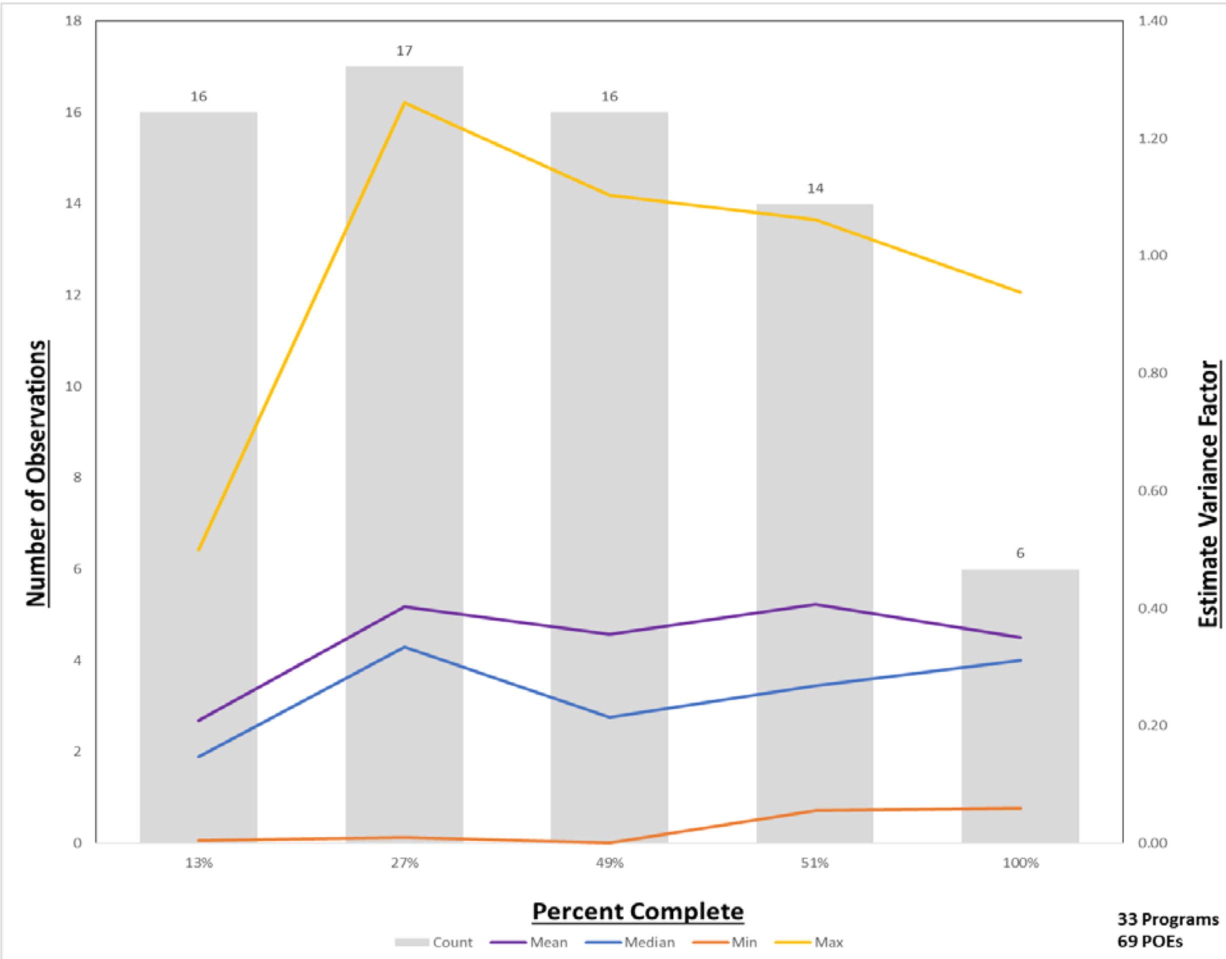



**Figure 4.23: PAUC Grp 1 Estimate Variance Factor (Without Outliers)**

Figure 4.23 is very similar to Figure 4.19, yet there are some differences. The median now experiences an average change of about 0.04 instead of the 0.05. The median line now experiences an overall increase of 0.1641, instead of the 0.1825, from the 13 percent bin to the 100 percent bin. The mean line now experiences an overall increase of 0.1412, instead of the 0.1498, from the 13 percent bin to the 100 percent bin. The rest of the values are the same as those in Figure 4.19.

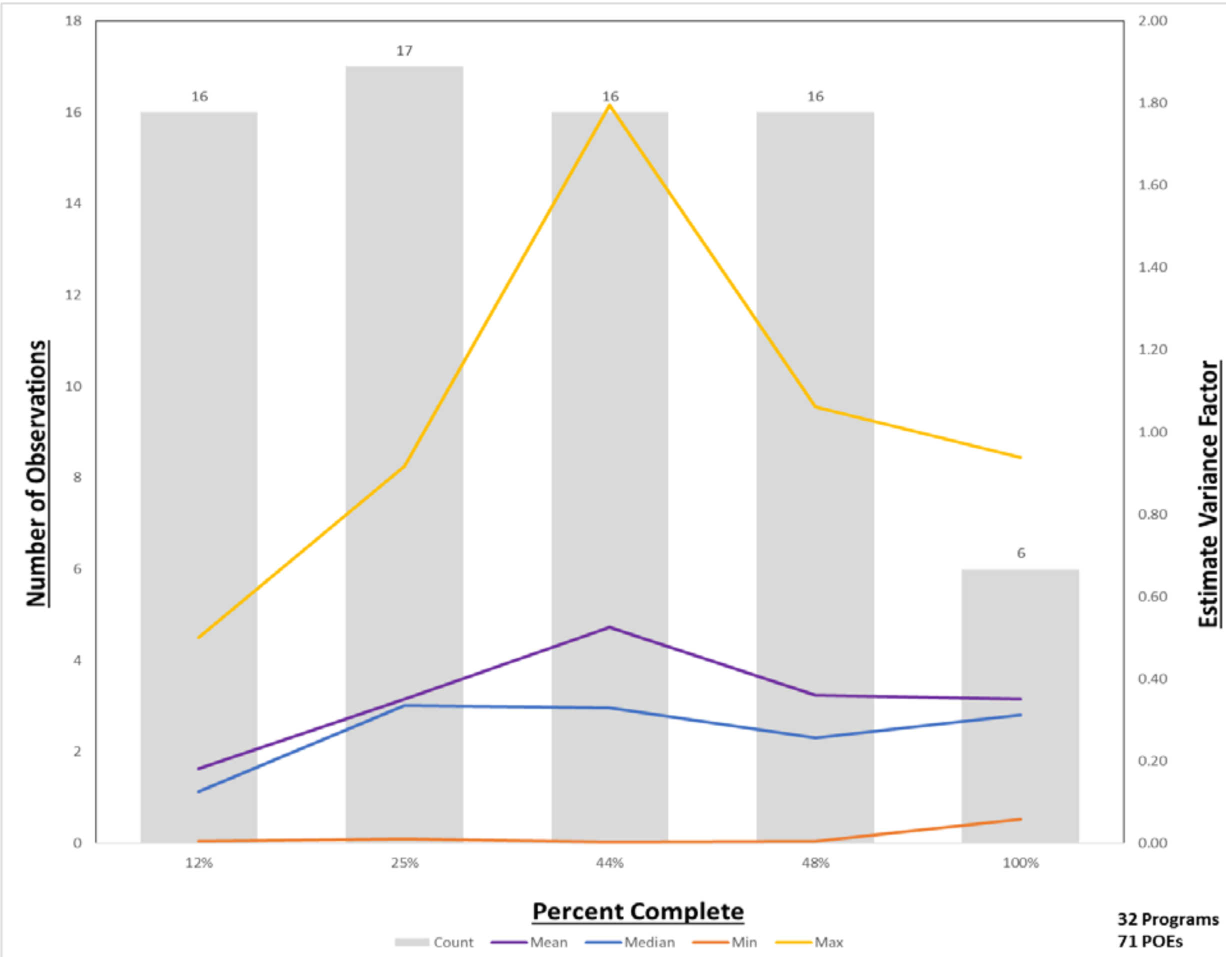



**Figure 4.24: PAUC Grp 2 Estimate Variance Factor (Without Outliers)**

Figure 4.24 is very similar to Figure 4.20, yet there are some differences. The climb from the 12 percent bin to the 25 percent bin is now an increase of 0.42. The peak at the 44 percent bin is an increase of 0.88. The median line now experiences an overall increase of 0.1866, instead of the 0.1907, from the 12 percent bin to the 100 percent bin. The mean line now experiences an overall increase of 0.1684, instead of the 0.1753, from the 12 percent bin to the 100 percent bin. The rest of the values are the same as those in Figure 4.20. Tables 4.12 through 4.14 have the difference between the 0% EVF and the 100% EVF. Note: WOO in the tables stands for Without Outlier programs.

**Table 4.12: Change from 0% to 100% 21 Bin Structure**

| | Median | Mean |
|---|---|---|
| APUC | 0.266 | 0.053 |
| APUC WOO | 0.269 | 0.104 |
| PAUC | 0.297 | 0.249 |
| PAUC WOO | 0.298 | 0.250 |
| **Average** | 0.283 | 0.164 |
| **Std Dev** | 0.018 | 0.101 |

**Table 4.13: Change from 0% to 100% Grp 1 Bin Structure**

| | Median | Mean |
|---|---|---|
| APUC | 0.168 | 0.075 |
| APUC WOO | 0.168 | 0.075 |
| PAUC | 0.191 | 0.175 |
| PAUC WOO | 0.187 | 0.168 |
| **Average** | 0.178 | 0.124 |
| **Std Dev** | 0.012 | 0.056 |

**Table 4.14: Change from 0% to 100% Grp 2 Bin Structure**

| | Median | Mean |
|---|---|---|
| APUC | 0.168 | 0.078 |
| APUC WOO | 0.168 | 0.078 |
| PAUC | 0.182 | 0.150 |
| PAUC WOO | 0.164 | 0.141 |
| **Average** | 0.171 | 0.112 |
| **Std Dev** | 0.008 | 0.039 |

For tables 4.12 through 4.14 the standard deviation for the mean values is higher than the standard deviation for the median values. These comparisons are for different breakouts of the dataset and represent different populations thus statistical comparisons are invalid. Yet the research believes it is good to see the numerical change in EVF for the different populations. In order to not have the research be overly influenced by outliers the research will focus on the median values. The research shows on average the APUC EVF experienced an increase of 0.27 within the 21 bin structure. It shows that on average the PAUC EVF experienced an increase of 0.30 within the 21 bin structure.

The research shows on average the APUC EVF experienced an increase of 0.17 within the group 1 bin structure. It shows that on average the PAUC EVF experienced an increase of 0.19 within the group 1 bin structure. The research shows on average the APUC EVF experienced an increase of 0.17 within the group 2 bin structure. It shows that on average the PAUC EVF experienced an increase of 0.18 within the group 2 bin structure, when the outlier programs were included. When the outlier programs were not included, the research shows on average the PAUC EVF experienced an increase of 0.16 within the group 2 bin structure.

**Fisher's Exact Test: Estimate Variance Factor**

Fisher's Exact Test is appropriate for small sample sizes when looking for a statistical association between two categorical variables. For this research the test is between a program characteristic and EVF values that exceed a defined threshold. The thresholds for this set of Fisher's Exact Tests are the Nunn-McCurdy breach percentage thresholds. The program characteristics are Acquisition Category (ACAT) I, ACAT II, ACAT III, concurrency and prototyping.

A Fisher's Exact Test was conducted at each percent complete bin to ensure that each test only had one observation of each program to maintain independence. The tests were also broken up by APUC and PAUC. Table 4.15 provides a summary of the EVF Fisher's Exact Test significant results. If a program characteristic does not have a significant effect on the EVF or EGF threshold breach then the p-value column and the columns after it are marked as "N/A". There are 4 instances of program characteristics having significant effects on the EVF.

**Table 4.15: Fisher's Exact Test EVF Results**

| Characteristic | Effect | One Tail p-value | Two Tail p-value | Percentage Bin | PAUC or APUC | Outliers Present | POEs |
|---|---|---|---|---|---|---|---|
| ACAT I | No effect on Estiamte Variance | N/A | N/A | N/A | N/A | N/A | N/A |
| ACAT II | No effect on Estiamte Variance | N/A | N/A | N/A | N/A | N/A | N/A |
| ACAT III | No effect on Estiamte Variance | N/A | N/A | N/A | N/A | N/A | N/A |
| Concurrency is not present | Estimate Variance Significant Breach | 0.0470 | 0.0889 | 0% | APUC | No | 14 |
| Prototyping is present | Estimate Variance Significant Breach | 0.0242 | 0.0242 | 35% | PAUC | N/A | 11 |
| Prototyping is present | Estimate Variance Critical Breach | 0.0061 | 0.0061 | 35% | PAUC | N/A | 11 |

These results indicate that it is more likely that a program will have a PAUC EVF significant breach at 35% complete if prototyping is present. It is more likely that a program will have a PAUC EVF above the critical breach threshold at 35% complete if prototyping is present. It is also more likely that a program will have an APUC Estimate Variance significant breach between 0% complete and 2.5% complete if concurrency is not present. The other characteristics did not have a significant effect on EVF significant or critical breaches.

The outlier programs mentioned earlier had POEs between the 0% complete and 2.5% complete point. When the outliers were included, the concurrency results became insignificant.

It makes sense that if prototyping was predictive of a critical breach at the 35% complete then it would also be predictive of a significant breach at the same time since the critical is a higher threshold than significant. None of the ACAT levels turn out to have a significant effect on EVF breaching the significant or critical thresholds. Table 4.16 has the results for the group 1 and group 2 bins. Same comment as Table 4.15 regarding N/As.

**Table 4.16: Fisher's Exact Test EVF Results for Kozlak Bins**

| Characteristic | Effect | One-Tail p-value | Two-Tail p-value | Group | Percentage Bin | PAUC or APUC | Outlier Present | POEs |
|---|---|---|---|---|---|---|---|---|
| ACAT I is not present | Estimate Variance Significant Breach | 0.0105 | 0.0114 | Grp 2 | 48 | PAUC | N/A | 16 |
| ACAT II is present | Estimate Variance Significant Breach | 0.0245 | 0.0350 | Grp 2 | 48 | PAUC | N/A | 16 |
| ACAT III | No Effect on Estimate Variance | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Concurrency is present | Estimate Variance Significant Breach | 0.0430 | 0.0498 | Grp1 | 13 | PAUC | Yes | 17 |
| Concurrency is not present | Estimate Variance Significant Breach | 0.0357 | 0.0357 | Grp1 | 100 | APUC | N/A | 8 |
| Concurrency is not present | Estimate Variance Significant Breach | 0.0357 | 0.0357 | Grp 2 | 100 | APUC | N/A | 8 |
| Prototyping is present | Estimate Variance Critical Breach | 0.0217 | 0.0217 | Grp1 | 27 | PAUC | Yes | 18 |
| Prototyping is present | Estimate Variance Critical Breach | 0.0099 | 0.0099 | Grp1 | 27 | PAUC | No | 17 |
| Prototyping is present | Estimate Variance Critical Breach | 0.0276 | 0.0276 | Grp 2 | 25 | PAUC | No | 17 |

These results indicate that it is more likely that a program will have a PAUC EVF significant breach at 48% complete if the program is not an ACAT I program. It is more likely that a program will have a PAUC EVF significant breach at 48% complete if the program is an ACAT II program. ACAT III was not significant.

It is more likely that a program will have a PAUC EVF significant breach at 13% complete if concurrency is present. It is more likely that a program will have an APUC EVF significant breach at 100% complete if concurrency is not present. This result is unaffected by which bin grouping is used.

It is more likely that a program will have a PAUC EVF significant breach at 27% complete if concurrency is not present. This result is unaffected by whether or not the outlier programs are present. It is more likely that a program will have a PAUC EVF significant breach at 25% complete if concurrency is not present and the outlier programs are not present.

### Fisher's Exact Test: Estimate Growth Factor

The above results were for EVF; recall EVF considers only the magnitude of the change in the POE (whether overestimated or underestimated). This section covers the results of Fisher's Exact Tests for EGF which categorizes POE increases above the thresholds differently than POE decreases above the threshold. This portion also better ties into the metrics of Nunn-McCurdy which is a breach in cost growth not in cost variance. A program does not have a Nunn-McCurdy breach if it is under budget.

The hypotheses are the same with EGF being swapped in for EVF. There are only 3 instances of program characteristics having significant effects on the EGF. Table 4.17 has a summary of the results:

**Table 4.17: Fisher's Exact Test EGF Nunn-McCurdy Breach Results**

| Characteristic | Effect | One Tail p-value | Two Tail p-value | Percentage Bin | PAUC or APUC | Outliers Present | POEs |
|---|---|---|---|---|---|---|---|
| ACAT I | No effect on Estiamte Growth | N/A | N/A | N/A | N/A | N/A | N/A |
| ACAT II | No effect on Estiamte Growth | N/A | N/A | N/A | N/A | N/A | N/A |
| ACAT III | No effect on Estiamte Growth | N/A | N/A | N/A | N/A | N/A | N/A |
| Concurrency | No effect on Estiamte Growth | N/A | N/A | N/A | N/A | N/A | N/A |
| Prototyping is present | Estimate Growth Significant Breach | 0.0242 | 0.0242 | 35% | PAUC | N/A | 11 |
| Prototyping is present | Estimate Growth Critical Breach | 0.0061 | 0.0061 | 35% | PAUC | N/A | 11 |

The ACAT levels again have no significant effect on the dependent variable, or in this case the EGF. Concurrency has no significant effect on the EGF. Prototyping being present in a program at the 35% complete bin is again correlated to both a critical and a significant breach in PAUC. Table 4.18 has the results for the group 1 and group 2 bin structures.

**Table 4.18: Fisher's Exact Test EGF Nunn-McCurdy Breach Results for Kozlak Bins**

| Characteristic | Effect | One-Tail p-value | Two-Tail p-value | Group | Percentage Bin | PAUC or APUC | Outlier Present | POEs |
|---|---|---|---|---|---|---|---|---|
| ACAT I is not present | Estimate Growth Significant Breach | 0.0262 | 0.0338 | Grp 2 | 48 | PAUC | N/A | 16 |
| ACAT II | No Effect on Estimate Growth | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| ACAT III | No Effect on Estimate Growth | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Concurrency is not present | Estimate Growth Significant Breach | 0.0410 | 0.0410 | Grp1 | 51 | PAUC | N/A | 14 |
| Concurrency is not present | Estimate Growth Significant Breach | 0.0357 | 0.0357 | Grp1 | 100 | APUC | N/A | 8 |
| Concurrency is not present | Estimate Growth Significant Breach | 0.0357 | 0.0357 | Grp 2 | 100 | APUC | N/A | 8 |
| Prototyping is present | Estimate Growth Critical Breach | 0.0441 | 0.0441 | Grp1 | 27 | PAUC | Yes | 18 |
| Prototyping is present | Estimate Growth Critical Breach | 0.0147 | 0.0147 | Grp1 | 27 | PAUC | No | 17 |
| Prototyping is present | Estimate Growth Critical Breach | 0.0294 | 0.0294 | Grp 2 | 25 | PAUC | No | 17 |

If the program is not an ACAT I program then it is more likely that a program will have a PAUC EGF significant breach in the 48% complete bin. ACAT II and ACAT III were not significant.

It is more likely that a program will have a PAUC EGF significant breach at 51% complete if concurrency is not present. It is more likely that a program will have an APUC EGF significant breach at 100% complete if concurrency is not present. This result is unaffected by which bin grouping is used.

It is more likely that a program will have a PAUC EGF significant breach at 27% complete if concurrency is not present. This result is unaffected by whether or not the outlier programs are present. It is more likely that a program will have a PAUC EGF significant breach at 25% complete if concurrency is not present and the outlier programs are not present.

**Fisher's Exact Test: Estimate Growth Factor Increasing or Decreasing**

The final set of Fisher's Exact Tests that were conducted were conducted with the EGF however, the metric changed from Nunn-McCurdy breaches to "cost growth". In other words did the program experience a change in its estimate that was positive or negative? The hypotheses are the same with any cost growth (i.e., EGF > 0) replacing the significant and critical breach thresholds. Again there are only 3 instances of program characteristics having significant effects on the EGF however, the differences here are more pronounced than the slight difference between the first two set of Fisher's Exact Tests. Table 4.19 has a summary of the results:

**Table 4.19: Fisher's Exact Test EGF "Cost Growth" Results**

| Characteristic | Effect | One Tail p-value | Two Tail p-value | Percentage Bin | PAUC or APUC | Outliers Present | POEs |
|---|---|---|---|---|---|---|---|
| ACAT I | No effect on Estiamte Growth | N/A | N/A | N/A | N/A | N/A | N/A |
| ACAT II | No effect on Estiamte Growth | N/A | N/A | N/A | N/A | N/A | N/A |
| ACAT III | No effect on Estiamte Growth | N/A | N/A | N/A | N/A | N/A | N/A |
| Concurrency is not present | Estimate Increase | 0.042 | 0.044 | 0% | APUC | Yes | 15 |
| Concurrency is not present | Estimate Increase | 0.0143 | 0.0286 | 60% | PAUC | N/A | 8 |
| Prototyping | No effect on Estiamte Growth | N/A | N/A | N/A | N/A | N/A | N/A |

In this set prototyping does not have a significant effect on a program EGF experiencing "cost growth". ACAT I, ACAT II, and ACAT III do not have a significant effect. If the outliers are present then a program that does not have concurrency is likely to experience an APUC EGF "cost growth" between the 0% and the 2.5% complete point. If the outliers are removed then the effect of concurrency becomes insignificant. When concurrency is not present the PAUC EGF is likely to have "cost growth" at the 60% complete bin. Table 4.20 has the results for the group 1 and group 2 bins.

**Table 4.20: Fisher's Exact Test EGF "Cost Growth" Results for Kozlak Bins**

| Characteristic | Effect | One-Tail p-value | Two-Tail p-value | Group | Percentage Bin | PAUC or APUC | Outlier Present | POEs |
|---|---|---|---|---|---|---|---|---|
| ACAT I | No Effect on Estimate Increase | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| ACAT II is present | Estiamte Increase | 0.0476 | 0.0476 | Grp 2 | 25 | APUC | Yes | 7 |
| ACAT III | No Effect on Estimate Increase | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Concurrency is not present | Estiamte Increase | 0.0286 | 0.0286 | Grp1 | 27 | APUC | Yes | 7 |
| Prototyping | No Effect on Estimate Increase | N/A | N/A | N/A | N/A | N/A | N/A | N/A |

These results indicate that it is more likely that a program will have an APUC "cost growth" at 25% complete if the program is an ACAT II program. ACAT I and ACAT III were not significant. It is more likely that a program will have an APUC "cost growth" in the 27% complete bin if concurrency is not present. If the outlier programs are taken out then the ACAT II and concurrency results become insignificant. Prototyping was insignificant.

### Other Government Costs-to-Contract Costs Ratio

Summary statistics were calculated for OGC-to-contract costs ratios; see Table 4.21:

**Table 4.21: OGC-to-Contract Costs Ratio Summary Statistics (With Outlier)**

| | EMD | Procurement |
|---|---|---|
| Programs | 140 | 140 |
| Programs With OGC Data | 114 | 111 |
| Programs with OGC and Contract Total Data | 83 | 82 |
| Max | 13.5652 | 0.4676 |
| Min | 0.0003 | 0.0008 |
| Mean | 0.3593 | 0.0812 |
| Median | 0.1062 | 0.0503 |
| Standard Deviation | 1.4973 | 0.0917 |

The maximum ratio for the EMD phase is 13.57; this is 11.7 more than the next highest ratio (1.87). The outlier was removed and the descriptive statistics were re-calculated. Table 4.22 provides the new summary statistics:

**Table 4.22: OGC to Contract Ratio Summary Statistics Without Outlier**

| | EMD | Procurement |
|---|---|---|
| Programs | 140 | 140 |
| Programs With OGC Data | 114 | 111 |
| Programs with OGC and Contract Total Data | 83 | 82 |
| Programs Without Outliers | 82 | 81 |
| Max | 1.8783 | 0.4676 |
| Min | 0.0003 | 0.0008 |
| Mean | 0.1983 | 0.0816 |
| Median | 0.1043 | 0.0517 |
| Standard Deviation | 0.3005 | 0.0922 |

The Procurement values are almost identical to the original values. The EMD values experienced more change. The mean decreased by 0.161 and the median decreased by 0.002. The standard deviation decreased by 1.197. The maximum decreased by 11.687. Table 4.21 and 4.22 show the descriptive statistics of the ratios for EMD and Procurement.

**Table 4.23: EMD Ratio Descriptive Statistics by ACAT**

| ACAT | Max | Mean | Median | Min | Standard Deviation | Number of Programs |
|---|---|---|---|---|---|---|
| ACAT I | 0.270 | 0.075 | 0.049 | 0.005 | 0.085 | 9 |
| ACAT II | 0.108 | 0.045 | 0.029 | 0.011 | 0.035 | 15 |
| ACAT III | 0.468 | 0.118 | 0.076 | 0.001 | 0.133 | 21 |

**Table 4.24: Procurement Ratio Descriptive Statistics by ACAT**

| ACAT | Max | Mean | Median | Min | Standard Deviation | Number of Programs |
|---|---|---|---|---|---|---|
| ACAT I | 0.212 | 0.103 | 0.056 | 0.006 | 0.085 | 9 |
| ACAT II | 1.878 | 0.239 | 0.102 | 0.016 | 0.469 | 15 |
| ACAT III | 0.478 | 0.138 | 0.116 | 0.010 | 0.109 | 20 |

The OGC-to-contract costs Ratio distributions were run against 8 different JMP distribution goodness of fit tests. Tables 4.23 and 4.24 are the summary of the results of those tests:

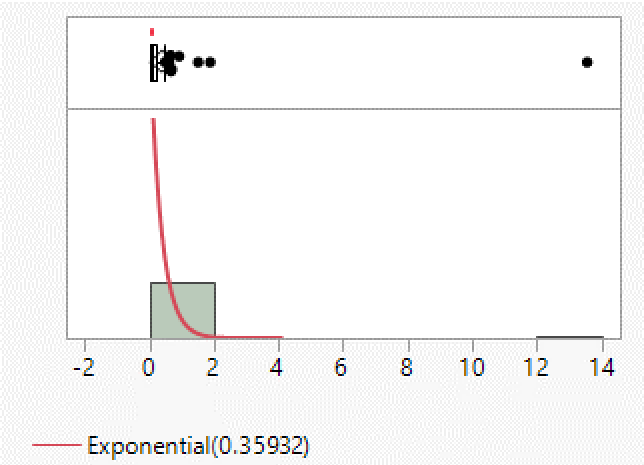
**Table 4.25: EMD Ratio Distribution Test Results**

| Fit | Test Result | Test Result Without Outliers |
|---|---|---|
| Normal | Fail | Fail |
| LogNormal | Pass | Pass |
| Weibull | Fail | Fail |
| Weibull With Threshold | Fail | Fail |
| Extreme Value | Fail | Fail |
| Exponential | Fail | Fail |
| Gamma | Undefined | Undefined |
| Beta | Undefined | Undefined |

**Table 4.26: Procurement Ratio Distribution Test Results**

| Fit | Test Result | Test Result Without Outliers |
|---|---|---|
| Normal | Fail | Fail |
| LogNormal | Pass | Pass |
| Weibull | Fail | Fail |
| Weibull With Threshold | Fail | Fail |
| Extreme Value | Fail | Fail |
| Exponential | Pass | Pass |
| Gamma | Undefined | Undefined |
| Beta | Undefined | Undefined |

Exclusion of the outlier program does not change the results of any of the tests. However, the results for the exponential fit test are different depending on which program phase is being analyzed. When the EMD and Procurement ratio distributions are fitted with an exponential curve, procurement passes and EMD fails. Figures 4.25 and 4.26 are the EMD distribution graphs with the fitted exponential curves.



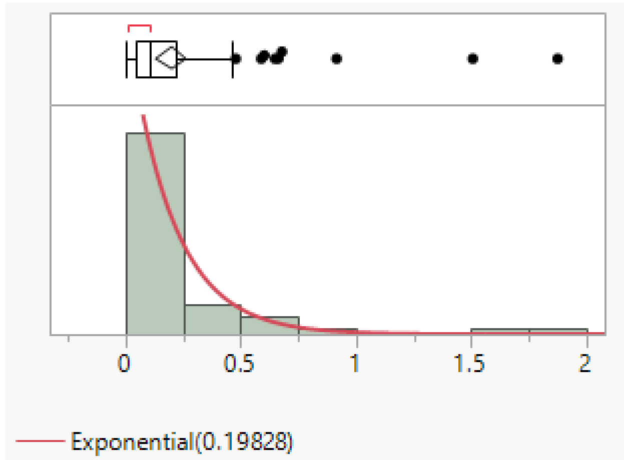




**Figure 4.25: EMD Ratio Exponential Distribution**

**Figure 4.26: EMD Ratio Exponential Distribution Without Outliers**

The data in Figure 4.26 has 8 bins between 0 and 2, while Figure 4.25 has only 1 bin between 0 and 2. Figure 4.26 has bin sizes of 0.25, while in Figure 4.25 has bin sizes of 2. However, the fit test results are the same. The p-value for both tests was 0.01, which fails the fit test. The EMD distribution failed the exponential fit test but the graph is very similar to an exponential downward sloping shape. There are 10 data points in Figure 4.25 and 9 data points in Figure 4.26 that are seen as outliers by the outlier box plots. It is possible these points could have influenced the distribution to fail the fit test. Figures 4.27 and 4.28 are the Procurement distribution graphs with the fitted exponential curves.

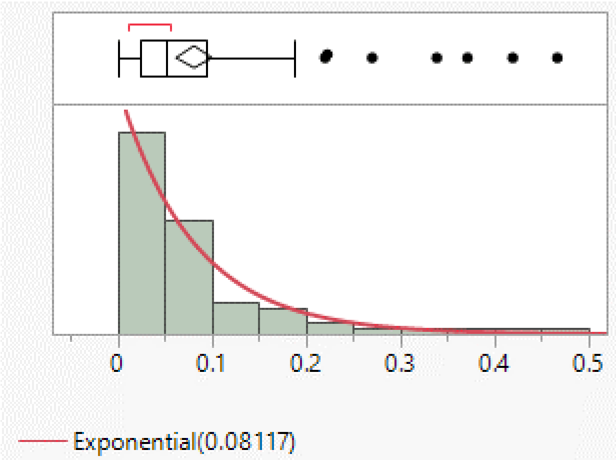



**Figure 4.27: Procurement Ratio Exponential Distribution**

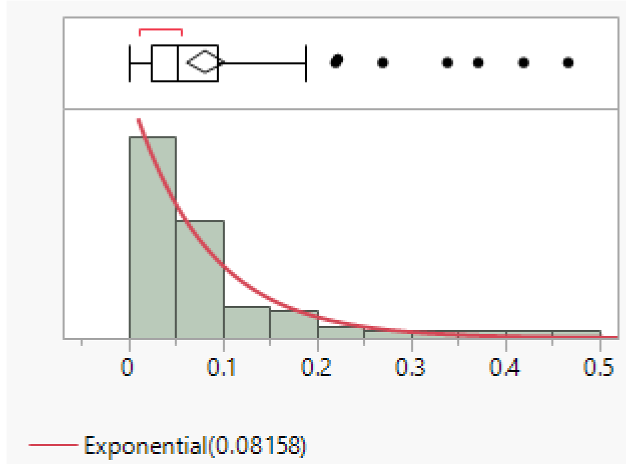



**Figure 4.28: Procurement Ratio Exponential Distribution Without Outliers**

Figure 4.27 is more or less visually identical to Figure 4.28, and the test results are the same. The p-value for both tests was 0.15, which passes the fit test. The Procurement distribution passed the exponential fit test and has a similar basic downward sloping shape as to that of the EMD distribution. There are 7 data points, in both figures, that are seen as outliers by the outlier box plot. Yet the distributions still pass the fit test.

The results for the lognormal tests were passed for both EMD and Procurement, regardless of whether or not the outlier program was included. Figures 4.29 and 4.30 are the EMD distribution graphs with the fitted lognormal curves.

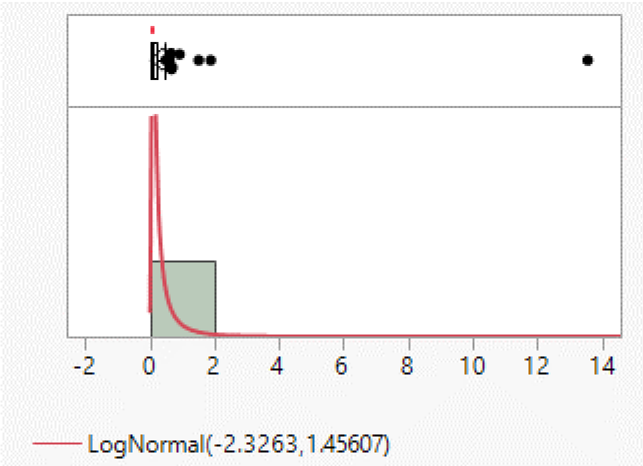



Figure 4.29: EMD Ratio LogNormal Distribution

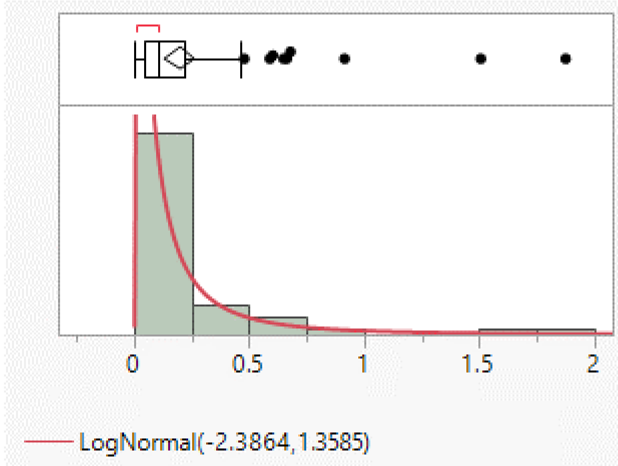



Figure 4.30: EMD Ratio LogNormal Distribution Without Outliers

The data in Figure 4.30 has 8 bins between 0 and 2, while Figure 4.29 has only 1 bin between 0 and 2. Figure 4.30 has bin sizes of 0.25, while in Figure 4.29 has bin sizes of 2. However, the fit test results are the same. The p-value for both tests was 0.15, which passes the fit test. There are 10 data points in Figure 4.29 and 9 data points in Figure 4.30 that are seen as outliers by the outlier box plots. These points did not cause the distribution to fail the fit test. Figures 4.31 and 4.32 are the Procurement distribution graphs with the fitted lognormal curves.

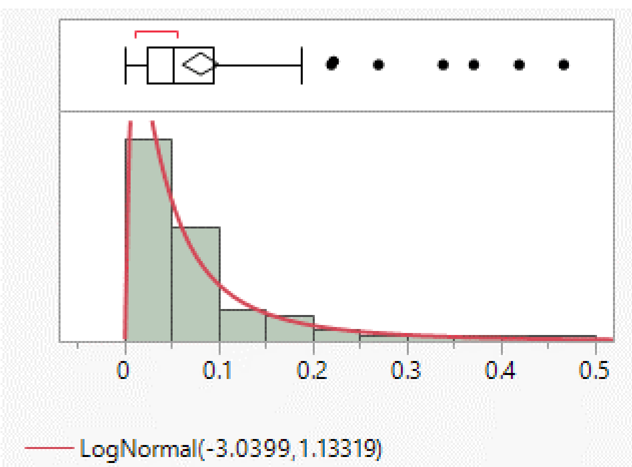



Figure 4.31: Procurement Ratio LogNormal Distribution

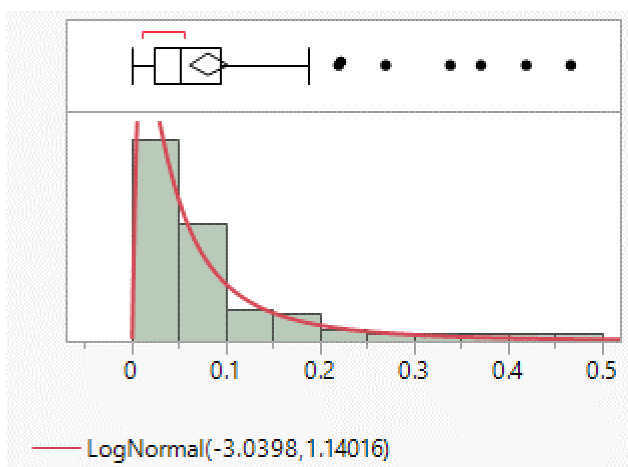



Figure 4.32: Procurement Ratio LogNormal Distribution Without Outliers

Figure 4.31 is more or less visually identical to Figure 4.32, and the test results are the same. The p-value for both tests was 0.15, which passes the fit test. The Procurement distribution passed the lognormal fit test and has a similar basic downward sloping shape as to that of the EMD distribution. There are 7 data points, in both figures, that are seen as outliers by the outlier box plot. Yet the distributions still pass the fit test. The results for the lognormal tests were passed for both EMD and Procurement, regardless of whether or not the outlier program was included.

**Summary**

The APUC and PAUC descriptive statistics for the dataset indicate that as time goes on PAUC can become less stable while APUC can become more stable. The Fisher's Exact Tests showed that ACAT III was not predictive in this dataset. ACAT II and ACAT I were predictive. Prototyping and concurrency were predictive 10 times each. The distribution of OGC-to-contract costs as a ratio for the procurement phase passed the exponential fit test. The distribution of OGC-to-contract costs as a ratio for EMD and procurement phases passed the lognormal fit test. Conversely, the distribution of the ratio for the EMD phase did not pass, but it did have a similarly shaped histogram. The next chapter will compare my results with those mentioned in my literature review and will also cover my recommendations.

# V. Conclusions and Recommendations

## Chapter Overview

This chapter will breakdown the responses to the research questions presented in Chapter I. It will be discussing confirmations or contradictions found in the current cost growth literature. Then it will discuss recommendations for future research and conclude the document. It needs to be reiterated that this research looked at estimate variance and estimate growth using initial program estimates and estimates throughout the program life cycle. The current literature that is being compared to this research looked at cost growth using estimates at MS B or MS II and completed program actuals throughout the program life cycle.

## Conclusions of Research

**1. How do Program Office Estimates (POE) change over the course of a program life cycle?**

From 0% complete to 100% complete Estimate Variance Factors (EVF) in the 21 bin structure experienced an increase of 27% in Acquisition Procurement Unit Cost (APUC) EVF and 30% in Program Acquisition Unit Cost (PAUC) EVF on average. The group 1 bin structure experienced an increase of 17% in APUC EVF and 19% in PAUC EVF on average. The group 2 bin structure experienced an increase of 17% in APUC EVF on average. The group 2 bin structure experienced an increase of 18% (on average) if the outlier programs were included. A 16% (on average) increase was experienced if the outlier programs were not included.

These values are defined differently from the Cost Growth Factors (CGF) of Drezner (1993) and Arena (2006), yet all look at changes in cost as a percentage. Assuming that all the Drezner (1993) found total CGFs for development, procurement and total program cost to be 25%, 18%, and 20% respectively (1993). Arena (2006) found total CGFs for development, procurement, and total program cost to be 58%, 44%, and 46% respectively.

The percentages of this research are closer to the percentages of Drezner (1993) than the percentages of Arena (2006). The research is comparative because an increase or decrease to a program's cost impacts the DoD budget. An increase may be looked at with more scrutiny by leadership but a decrease also causes funds to be reallocated and points to a deficiency in cost estimating. Leadership should be aware of any reallocation of funds and the reasoning behind that reallocation.

**2. What are predictive characteristics of POE cost variance or cost growth?**

**a. Prototyping**

Prototyping is tied, with concurrency, in the amount of instances it appears as a significant program characteristic. It appears significant in 10 different contingency tables. Prototyping is associated with PAUC EVF and PAUC Estimate Growth Factor (EGF) significant and critical breaches of the Nunn-McCurdy percentage thresholds in the 21 bin structure. Prototyping is associated with PAUC EVF and PAUC EGF critical breaches of the Nunn-McCurdy percentage thresholds in the group 1 and the group 2 bin structures.

Jimenez (2016) and Trudelle (2017) found prototyping to be insignificant. The results of this research contradict the findings of Jimenez (2016) and Trudelle (2017). This contradiction encourages future research to study prototyping in POEs.

#### b. Concurrency

Concurrency is tied, with prototyping, in the amount of instances it appears as a significant program characteristic. It appears significant in 10 different contingency tables. Concurrency is associated with significant breaches of the Nunn-McCurdy percentage thresholds in the 21 bin structure, the group 1 structure, and the group 2 bin structure. Concurrency is also a significant predictor when looking at "cost growth". Whether or not the EVF or EGF is APUC or PAUC depends on the contingency table.

Jimenez (2016), Trudelle (2017), and Drezner (1993) found concurrency to be insignificant. The results of this research contradict the findings of Jimenez (2016), Trudelle (2017), and Drezner (1993). This contradiction encourages future research to study concurrency in POEs.

#### c. ACAT I

Trudelle (2017) found that ACAT I programs experienced more cost growth than non-ACAT I programs. In this research there are 2 instances where if ACAT I is not present then it is more likely that the program will experience a PAUC EVF or EGF significant Nunn-McCurdy breach. These results show ACAT I to be significant and these results encourage future research to study ACAT I programs in POEs.

#### d. ACAT II

In this research there are 2 instances where if ACAT II is present then it is more likely that the program will experience a PAUC EVF significant Nunn-McCurdy breach, or an APUC “cost growth” in the group 2 bin structure. The impact of ACAT II programs is not confirmed or contradicted in the current literature. These results show ACAT II to be significant and these results encourage future research to study ACAT II programs in POEs.

#### e. ACAT III

ACAT III is not significant when the threshold is “cost growth” and the percent complete bins are at 5% increments or the Kozlak (2017) groupings. ACAT III is not significant when considering EGFs. ACAT III is also not significant when considering EVFs. The impact of ACAT III programs, or lack thereof, is not confirmed or contradicted in the current literature.

### 5. What is the distribution of the ratio Other Government Costs (OGC) to Contract costs?

The OGC-to-contract costs ratio distribution passed the lognormal fit test with both the EMD and the Procurement values. The OGC-to-contract costs ratio distribution also passed the exponential fit test with the Procurement values, but not the EMD values. Whether or not the OGC outlier program was included did not change the test results. Cost Analysts now have a Cost Estimating Relationship (CER) with summary statistics and distributions to possibly use with their estimates of OGCs. These CER estimates should strengthen their estimates beyond the mainstream analogous estimating of OGCs that takes place. There are no articles or research that can confirm or contradict these findings.

**Recommendations for Future Research**

The research suggests that more data collection and analysis would be beneficial for evaluating cost growth and cost variance in ACAT II and ACAT III programs. Increasing the size, standardization, and density of the dataset could better highlight possible trends within ACAT II and ACAT III programs. The CER created within the research should be tested on programs outside of the dataset for confirmation of application.

The results encourage future research to look at prototyping, concurrency, and the ACAT levels of a program when researching POEs. The database is sparse which invalidated the use of regression. The sparseness also limited the contingency table results to Fisher's Exact Tests which are geared toward small sample sizes but assume that the data being analyzed is the population or is representative of the population. This inference of the data being representative of all DoD programs or even just all Air Force programs is invalid considering the size and sparseness of the database. Yet it is the only POE database found by this research.

The sparseness of the database hinders the conclusions that can be drawn from a myriad of contingency tables with differing sample sizes and thus not truly comparable with each other. The results and the limitations of the database encourage further research into POEs.

Kozlak's bins were created with aircraft in mind but every program has Critical Design Review (CDR), First Flight (FF), Development Test & Evaluation End (DT&E), Initial Operating Capability (IOC), and Last SAR (LS). The only one of those milestones that may not be in other programs is FF, but every program has a similar milestone where prototypes are tested for the first time or simulations are run for the first time (AcqNotes 2019). Yet, the percent complete per milestone may vary and future research should look into different commodity types and their specific percent completes by milestone.

The research is not attempting to definitively evaluate or confirm the effects of program characteristics, but is rather trying to guide the bolstering of POE databases and POE research. This database and POE research should highlight cost growth and cost variance for ACAT II and ACAT III programs. Such programs are not highlighted in Selected Acquisition Reports (SAR) or the current cost growth literature.

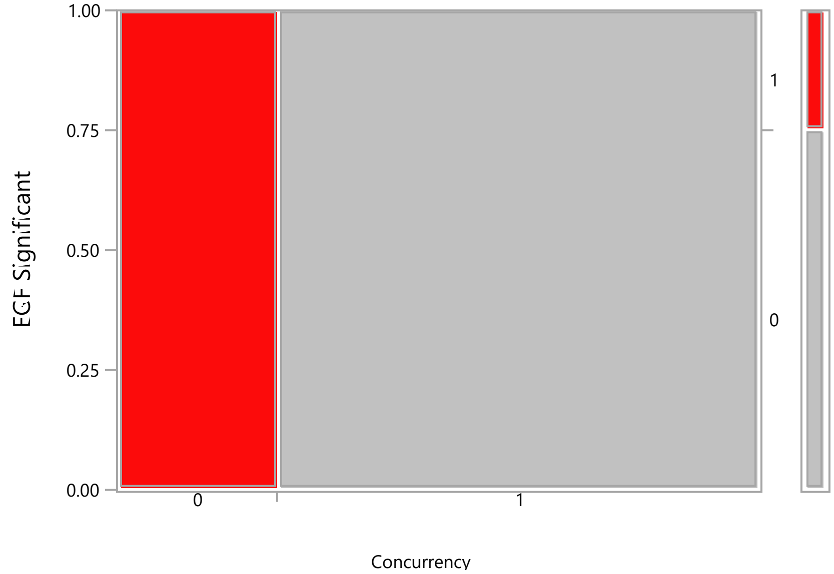
**Summary**

This chapter stated with the breakdown of the responses to the research questions presented in Chapter I. It discussed confirmations or contradictions found in the current cost growth literature to the results of this research. Then it discussed recommendations for future research. This research shows POE data and analyses to be limited and encourages future POE research to inform leadership of cost growth within ACAT II and ACAT III programs.

# Appendix

## Contingency Tables with Significant Results

### Contingency Analysis of EGF Significant By Concurrency
### Mosaic Plot (APUC Group 1, in the 100% bin)




### Contingency Table

Concurrency By EGF Significant

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 0<br>0.00<br>0.00<br>0.00 | 2<br>25.00<br>100.00<br>100.00 | 2<br>25.00 |
| 1 | 6<br>75.00<br>100.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 6<br>75.00 |
| Total | 6<br>75.00 | 2<br>25.00 | 8 |

### Tests

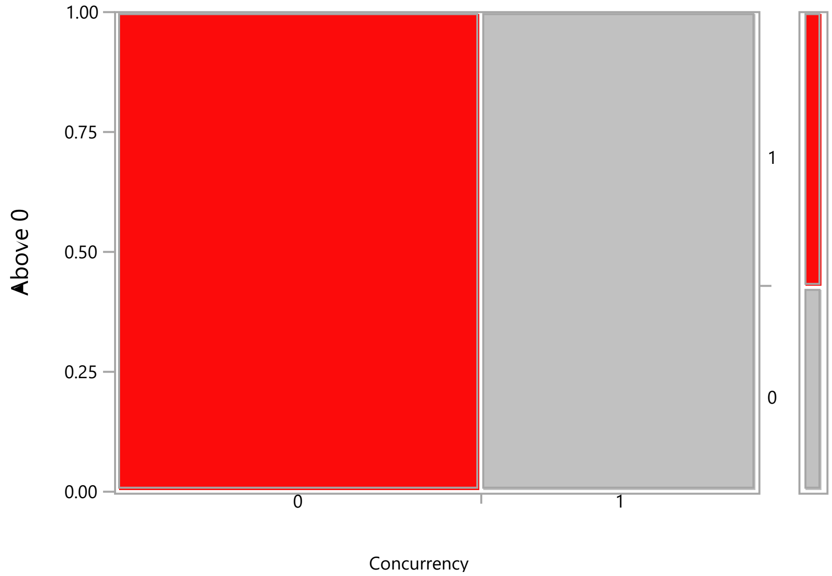
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 8 | 1 | 4.4986812 | 1.0000 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 8.997 | 0.0027* |
| Pearson | 8.000 | 0.0047* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.0357* | Prob(EGF Significant=1) is greater for Concurrency=0 than 1 |
| Right | 1.0000 | Prob(EGF Significant=1) is greater for Concurrency=1 than 0 |
| 2-Tail | 0.0357* | Prob(EGF Significant=1) is different across Concurrency |

## Contingency Analysis of Above 0 By Concurrency
## Mosaic Plot (APUC Group 1, in the 27% bin)




## Contingency Table

Concurrency By Above 0

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 0<br>0.00<br>0.00<br>0.00 | 4<br>57.14<br>100.00<br>100.00 | 4<br>57.14 |
| 1 | 3<br>42.86<br>100.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 3<br>42.86 |
| Total | 3<br>42.86 | 4<br>57.14 | 7 |

## Tests

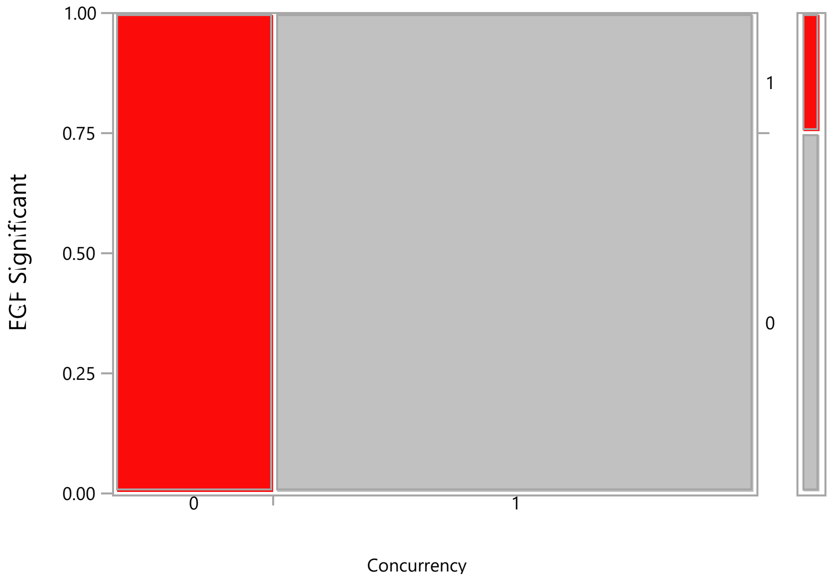
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 7 | 1 | 4.7803567 | 1.0000 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 9.561 | 0.0020* |
| Pearson | 7.000 | 0.0082* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.0286* | Prob(Above 0=1) is greater for Concurrency=0 than 1 |
| Right | 1.0000 | Prob(Above 0=1) is greater for Concurrency=1 than 0 |
| 2-Tail | 0.0286* | Prob(Above 0=1) is different across Concurrency |

## Contingency Analysis of EGF Significant By Concurrency
## Mosaic Plot (APUC Group 2, in the 100% bin)




## Contingency Table

Concurrency By EGF Significant

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 0<br>0.00<br>0.00<br>0.00 | 2<br>25.00<br>100.00<br>100.00 | 2<br>25.00 |
| 1 | 6<br>75.00<br>100.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 6<br>75.00 |
| Total | 6<br>75.00 | 2<br>25.00 | 8 |

## Tests

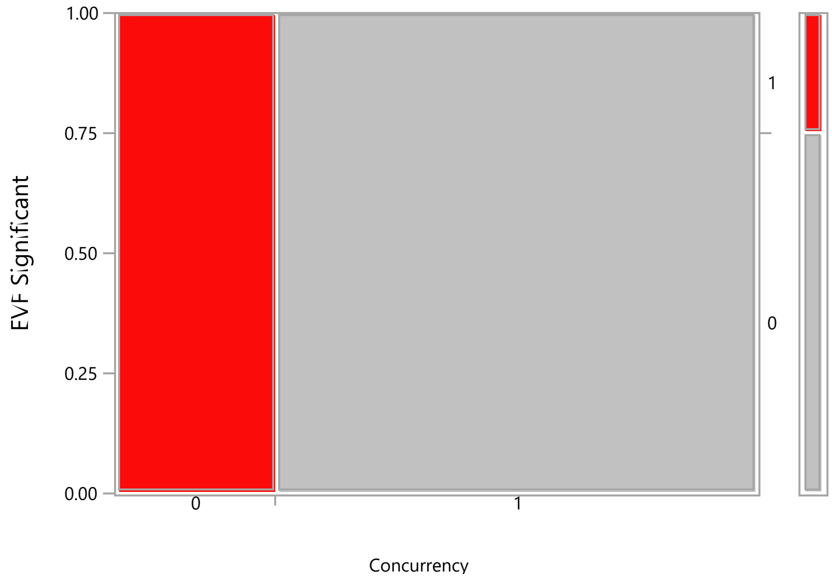
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 8 | 1 | 4.4986812 | 1.0000 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 8.997 | 0.0027* |
| Pearson | 8.000 | 0.0047* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.0357* | Prob(EGF Significant=1) is greater for Concurrency=0 than 1 |
| Right | 1.0000 | Prob(EGF Significant=1) is greater for Concurrency=1 than 0 |
| 2-Tail | 0.0357* | Prob(EGF Significant=1) is different across Concurrency |

## Contingency Analysis of EVF Significant By Concurrency
## Mosaic Plot (APUC Group 2, in the 100% bin)




## Contingency Table

Concurrency By EVF Significant

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 0<br>0.00<br>0.00<br>0.00 | 2<br>25.00<br>100.00<br>100.00 | 2<br>25.00 |
| 1 | 6<br>75.00<br>100.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 6<br>75.00 |
| Total | 6<br>75.00 | 2<br>25.00 | 8 |

## Tests

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 8 | 1 | 4.4986812 | 1.0000 |

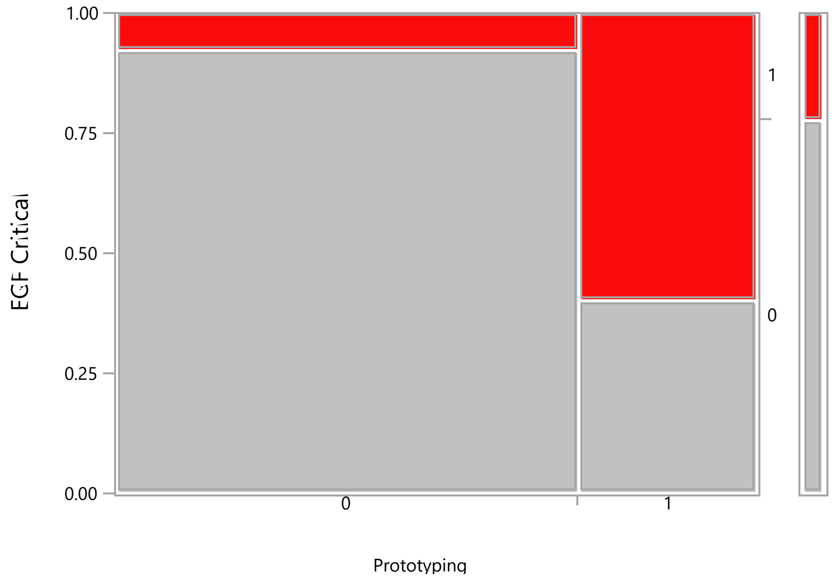
| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 8.997 | 0.0027* |
| Pearson | 8.000 | 0.0047* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.0357* | Prob(EVF Significant=1) is greater for Concurrency=0 than 1 |
| Right | 1.0000 | Prob(EVF Significant=1) is greater for Concurrency=1 than 0 |
| 2-Tail | 0.0357* | Prob(EVF Significant=1) is different across Concurrency |

## Contingency Analysis of EGF Critical By Prototyping

### Mosaic Plot (PAUC Group 1, in the 27% bin)




### Contingency Table

Prototyping By EGF Critical

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 12<br>66.67<br>85.71<br>92.31 | 1<br>5.56<br>25.00<br>7.69 | 13<br>72.22 |
| 1 | 2<br>11.11<br>14.29<br>40.00 | 3<br>16.67<br>75.00<br>60.00 | 5<br>27.78 |
| Total | 14<br>77.78 | 4<br>22.22 | 18 |

### Tests

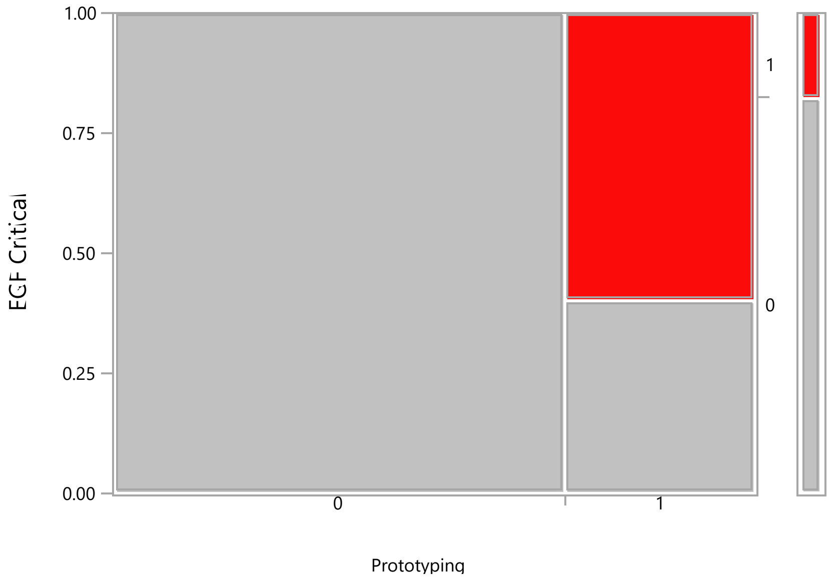
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 18 | 1 | 2.6441914 | 0.2773 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 5.288 | 0.0215* |
| Pearson | 5.716 | 0.0168* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.9984 | Prob(EGF Critical=1) is greater for Prototyping=0 than 1 |
| Right | 0.0441* | Prob(EGF Critical=1) is greater for Prototyping=1 than 0 |
| 2-Tail | 0.0441* | Prob(EGF Critical=1) is different across Prototyping |

## Contingency Analysis of EGF Critical By Prototyping
## Mosaic Plot (PAUC Group 1, in the 27% bin)




## Contingency Table

Prototyping By EGF Critical

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 12<br>70.59<br>85.71<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 12<br>70.59 |
| 1 | 2<br>11.76<br>14.29<br>40.00 | 3<br>17.65<br>100.00<br>60.00 | 5<br>29.41 |
| Total | 14<br>82.35 | 3<br>17.65 | 17 |

## Tests

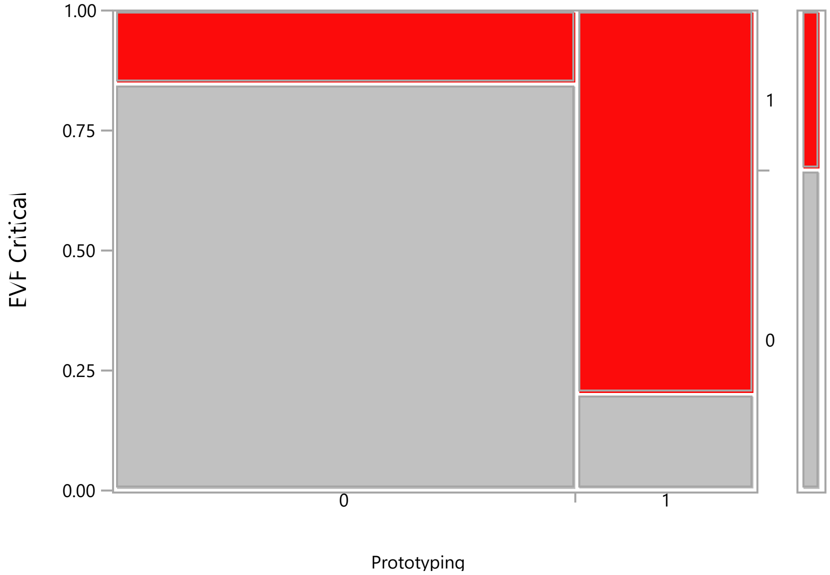
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 17 | 1 | 4.5569290 | 0.5752 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 9.114 | 0.0025* |
| Pearson | 8.743 | 0.0031* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 1.0000 | Prob(EGF Critical=1) is greater for Prototyping=0 than 1 |
| Right | 0.0147* | Prob(EGF Critical=1) is greater for Prototyping=1 than 0 |
| 2-Tail | 0.0147* | Prob(EGF Critical=1) is different across Prototyping |

**Contingency Analysis of EVF Critical By Prototyping**
**Mosaic Plot (PAUC Group 1, in the 27% bin)**




## Contingency Table

Prototyping By EVF Critical

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 11<br>61.11<br>91.67<br>84.62 | 2<br>11.11<br>33.33<br>15.38 | 13<br>72.22 |
| 1 | 1<br>5.56<br>8.33<br>20.00 | 4<br>22.22<br>66.67<br>80.00 | 5<br>27.78 |
| Total | 12<br>66.67 | 6<br>33.33 | 18 |

## Tests

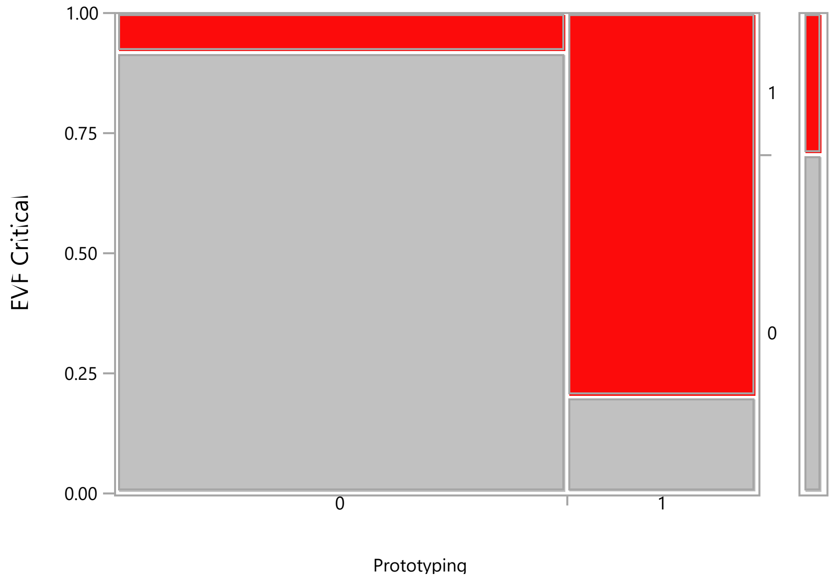
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 18 | 1 | 3.3740436 | 0.2945 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 6.748 | 0.0094* |
| Pearson | 6.785 | 0.0092* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.9993 | Prob(EVF Critical=1) is greater for Prototyping=0 than 1 |
| Right | 0.0217* | Prob(EVF Critical=1) is greater for Prototyping=1 than 0 |
| 2-Tail | 0.0217* | Prob(EVF Critical=1) is different across Prototyping |

## Contingency Analysis of EVF Critical By Prototyping
## Mosaic Plot (PAUC Group 1, in the 27% bin)




## Contingency Table
Prototyping By EVF Critical

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 11<br>64.71<br>91.67<br>91.67 | 1<br>5.88<br>20.00<br>8.33 | 12<br>70.59 |
| 1 | 1<br>5.88<br>8.33<br>20.00 | 4<br>23.53<br>80.00<br>80.00 | 5<br>29.41 |
| Total | 12<br>70.59 | 5<br>29.41 | 17 |

## Tests

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 17 | 1 | 4.3545136 | 0.4228 |

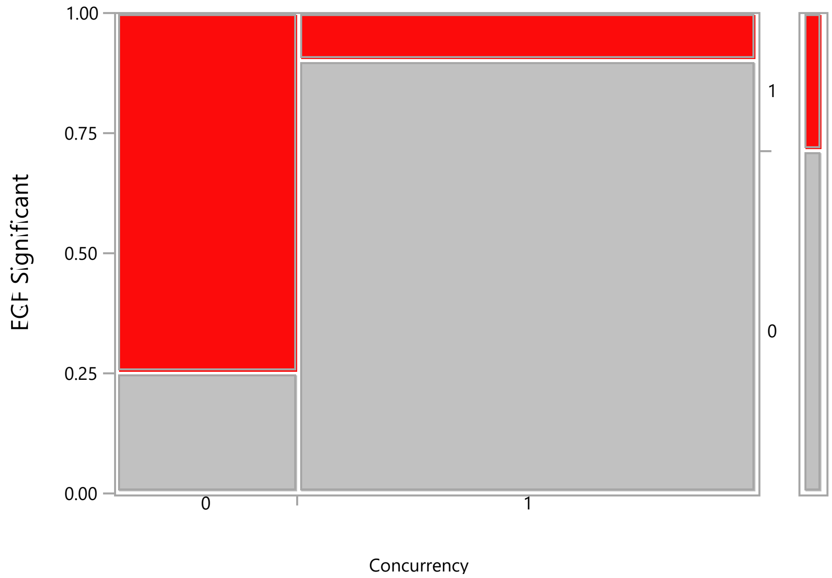
| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 8.709 | 0.0032* |
| Pearson | 8.731 | 0.0031* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.9998 | Prob(EVF Critical=1) is greater for Prototyping=0 than 1 |
| Right | 0.0099* | Prob(EVF Critical=1) is greater for Prototyping=1 than 0 |
| 2-Tail | 0.0099* | Prob(EVF Critical=1) is different across Prototyping |

## Contingency Analysis of EGF Significant By Concurrency

## Mosaic Plot (PAUC Group 1, in the 51% bin)




## Contingency Table

Concurrency By EGF Significant

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 1<br>7.14<br>10.00<br>25.00 | 3<br>21.43<br>75.00<br>75.00 | 4<br>28.57 |
| 1 | 9<br>64.29<br>90.00<br>90.00 | 1<br>7.14<br>25.00<br>10.00 | 10<br>71.43 |
| Total | 10<br>71.43 | 4<br>28.57 | 14 |

## Tests

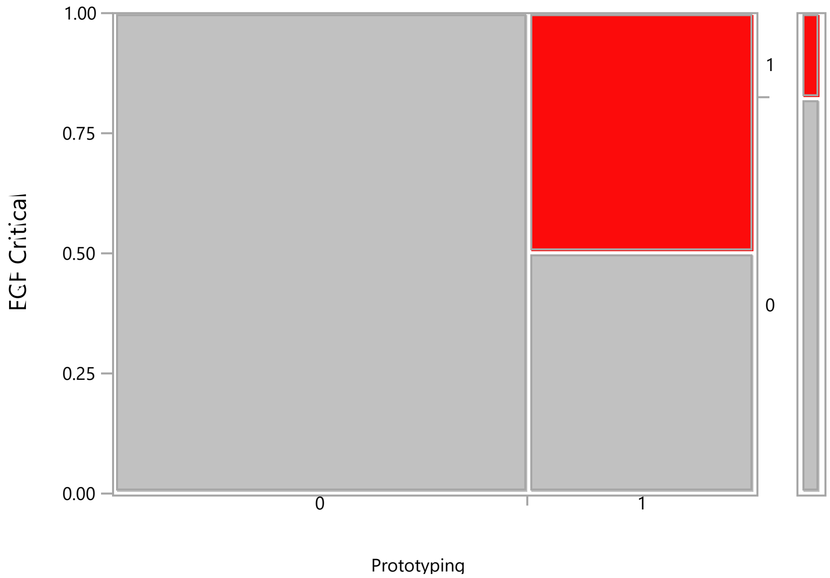
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 14 | 1 | 2.8756039 | 0.3433 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 5.751 | 0.0165* |
| Pearson | 5.915 | 0.0150* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.0410* | Prob(EGF Significant=1) is greater for Concurrency=0 than 1 |
| Right | 0.9990 | Prob(EGF Significant=1) is greater for Concurrency=1 than 0 |
| 2-Tail | 0.0410* | Prob(EGF Significant=1) is different across Concurrency |

## Contingency Analysis of EGF Critical By Prototyping
## Mosaic Plot (PAUC Group 2, in the 25% bin)




## Contingency Table

Prototyping By EGF Critical

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 11<br>64.71<br>78.57<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 11<br>64.71 |
| 1 | 3<br>17.65<br>21.43<br>50.00 | 3<br>17.65<br>100.00<br>50.00 | 6<br>35.29 |
| Total | 14<br>82.35 | 3<br>17.65 | 17 |

## Tests

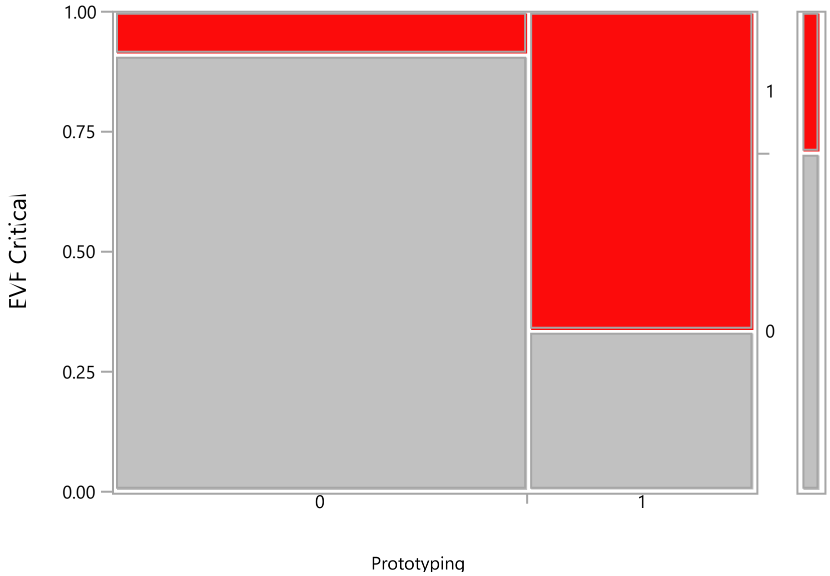
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 17 | 1 | 3.7631043 | 0.4750 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 7.526 | 0.0061* |
| Pearson | 6.679 | 0.0098* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 1.0000 | Prob(EGF Critical=1) is greater for Prototyping=0 than 1 |
| Right | 0.0294* | Prob(EGF Critical=1) is greater for Prototyping=1 than 0 |
| 2-Tail | 0.0294* | Prob(EGF Critical=1) is different across Prototyping |

**Contingency Analysis of EVF Critical By Prototyping**
**Mosaic Plot (PAUC Group 2, in the 25% bin)**




## Contingency Table

Prototyping By EVF Critical

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 10<br>58.82<br>83.33<br>90.91 | 1<br>5.88<br>20.00<br>9.09 | 11<br>64.71 |
| 1 | 2<br>11.76<br>16.67<br>33.33 | 4<br>23.53<br>80.00<br>66.67 | 6<br>35.29 |
| Total | 12<br>70.59 | 5<br>29.41 | 17 |

## Tests

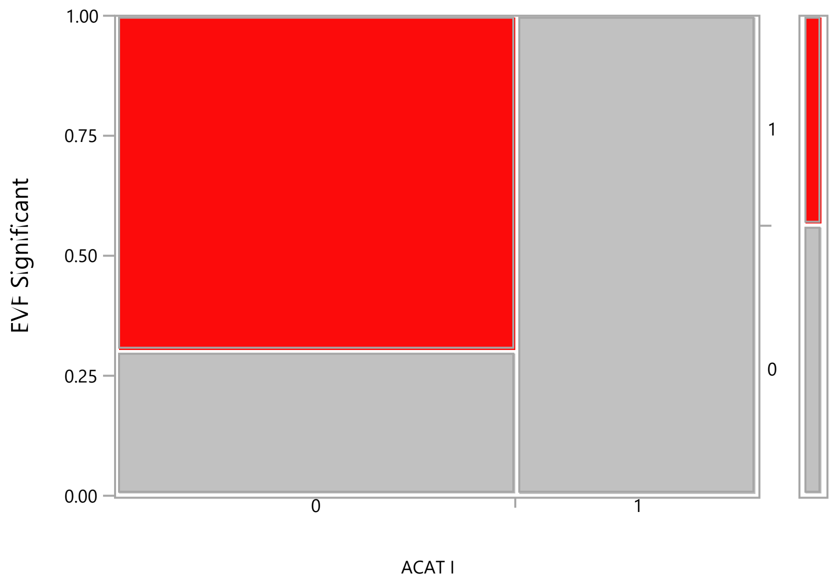
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 17 | 1 | 3.1284754 | 0.3038 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 6.257 | 0.0124* |
| Pearson | 6.199 | 0.0128* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.9990 | Prob(EVF Critical=1) is greater for Prototyping=0 than 1 |
| Right | 0.0276* | Prob(EVF Critical=1) is greater for Prototyping=1 than 0 |
| 2-Tail | 0.0276* | Prob(EVF Critical=1) is different across Prototyping |

## Contingency Analysis of EVF Significant By ACAT I
## Mosaic Plot (PAUC Group 2, in the 48% bin)




## Contingency Table

ACAT I By EVF Significant

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 3<br>18.75<br>33.33<br>30.00 | 7<br>43.75<br>100.00<br>70.00 | 10<br>62.50 |
| 1 | 6<br>37.50<br>66.67<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 6<br>37.50 |
| Total | 9<br>56.25 | 7<br>43.75 | 16 |

## Tests

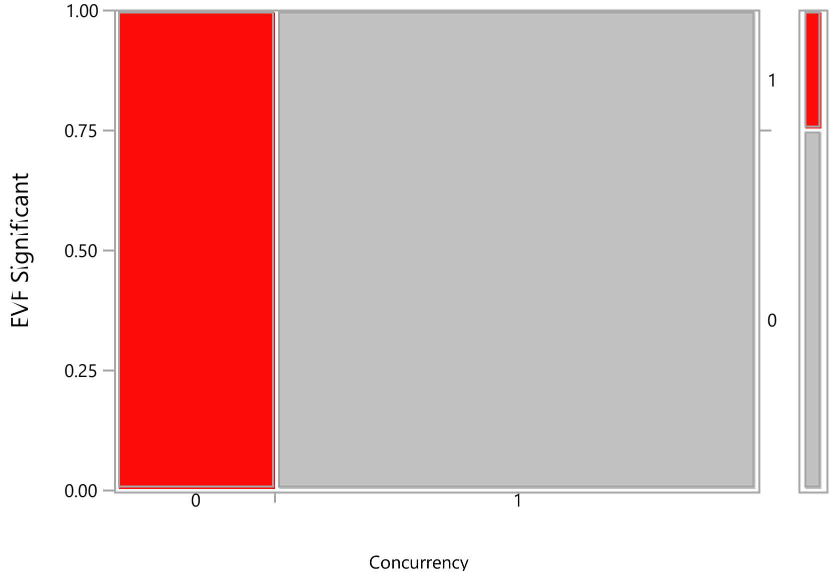
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 16 | 1 | 4.8563843 | 0.4429 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 9.713 | 0.0018* |
| Pearson | 7.467 | 0.0063* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.0105* | Prob(EVF Significant=1) is greater for ACAT I=0 than 1 |
| Right | 1.0000 | Prob(EVF Significant=1) is greater for ACAT I=1 than 0 |
| 2-Tail | 0.0114* | Prob(EVF Significant=1) is different across ACAT I |

### Contingency Analysis of EVF Significant By Concurrency
### Mosaic Plot (APUC Group 1, in the 100% bin)




### Contingency Table

Concurrency By EVF Significant

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 0<br>0.00<br>0.00<br>0.00 | 2<br>25.00<br>100.00<br>100.00 | 2<br>25.00 |
| 1 | 6<br>75.00<br>100.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 6<br>75.00 |
| Total | 6<br>75.00 | 2<br>25.00 | 8 |

### Tests

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 8 | 1 | 4.4986812 | 1.0000 |

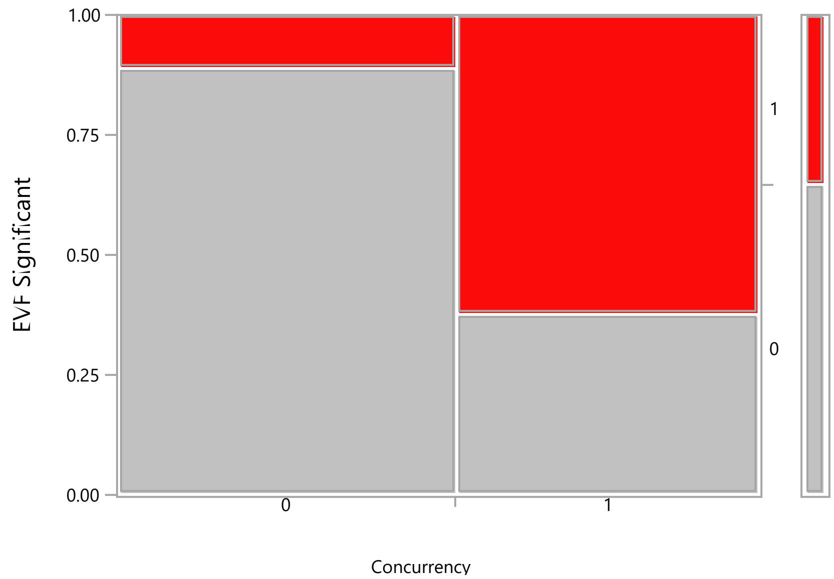
| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 8.997 | 0.0027* |
| Pearson | 8.000 | 0.0047* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.0357* | Prob(EVF Significant=1) is greater for Concurrency=0 than 1 |
| Right | 1.0000 | Prob(EVF Significant=1) is greater for Concurrency=1 than 0 |

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| 2-Tail | 0.0357* | Prob(EVF Significant=1) is different across Concurrency |

## Contingency Analysis of EVF Significant By Concurrency
## Mosaic Plot (PAUC Group 1, in the 13% bin)




### Contingency Table

Concurrency By EVF Significant

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 8<br>47.06<br>72.73<br>88.89 | 1<br>5.88<br>16.67<br>11.11 | 9<br>52.94 |
| 1 | 3<br>17.65<br>27.27<br>37.50 | 5<br>29.41<br>83.33<br>62.50 | 8<br>47.06 |
| Total | 11<br>64.71 | 6<br>35.29 | 17 |

### Tests

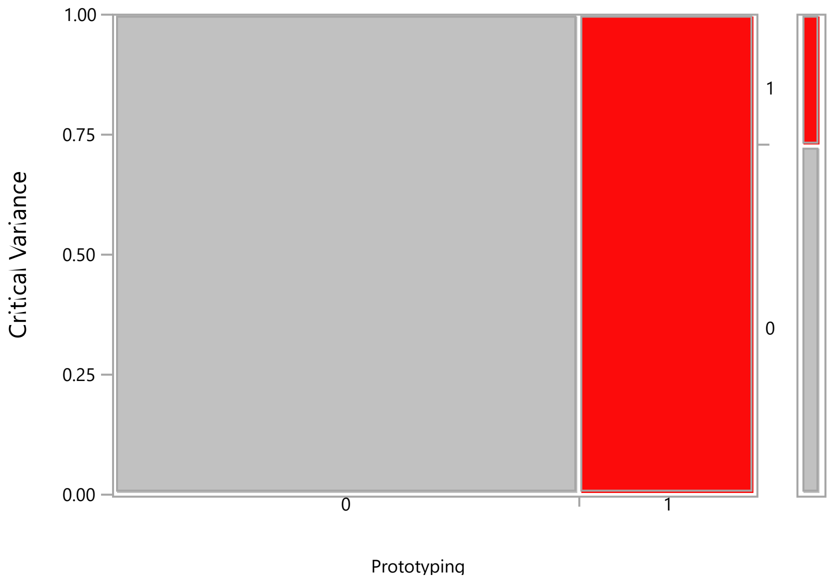
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 17 | 1 | 2.6052273 | 0.2360 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 5.210 | 0.0225* |
| Pearson | 4.898 | 0.0269* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.9977 | Prob(EVF Significant=1) is greater for Concurrency=0 than 1 |
| Right | 0.0430* | Prob(EVF Significant=1) is greater for Concurrency=1 than 0 |
| 2-Tail | 0.0498* | Prob(EVF Significant=1) is different across Concurrency |

## Contingency Analysis of Critical Variance By Prototyping
## Mosaic Plot (PAUC 21 bin, in the 35% bin)




## Contingency Table

Prototyping By Critical Variance

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 8<br>72.73<br>100.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 8<br>72.73 |
| 1 | 0<br>0.00<br>0.00<br>0.00 | 3<br>27.27<br>100.00<br>100.00 | 3<br>27.27 |
| Total | 8<br>72.73 | 3<br>27.27 | 11 |

## Tests

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 11 | 1 | 6.4454788 | 1.0000 |

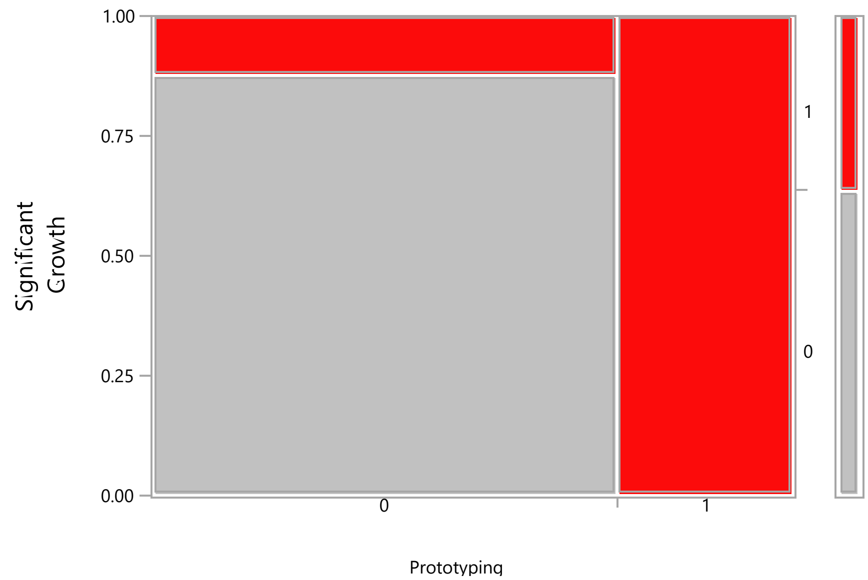
| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 12.891 | 0.0003* |
| Pearson | 11.000 | 0.0009* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 1.0000 | Prob(Critical Variance=1) is greater for Prototyping=0 than 1 |

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Right | 0.0061* | Prob(Critical Variance=1) is greater for Prototyping=1 than 0 |
| 2-Tail | 0.0061* | Prob(Critical Variance=1) is different across Prototyping |

## Contingency Analysis of Significant Growth By Prototyping
## Mosaic Plot (PAUC 21 bin, in the 35% bin)




## Contingency Table

Prototyping By Significant Growth

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 7<br>63.64<br>100.00<br>87.50 | 1<br>9.09<br>25.00<br>12.50 | 8<br>72.73 |
| 1 | 0<br>0.00<br>0.00<br>0.00 | 3<br>27.27<br>75.00<br>100.00 | 3<br>27.27 |
| Total | 7<br>63.64 | 4<br>36.36 | 11 |

## Tests

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 11 | 1 | 4.1961382 | 0.5820 |

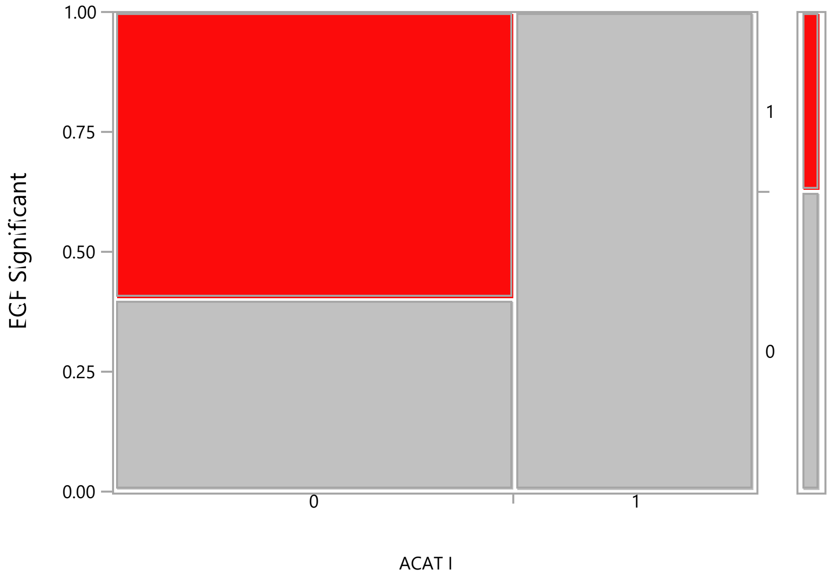
| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 8.392 | 0.0038* |
| Pearson | 7.219 | 0.0072* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 1.0000 | Prob(Significant Growth=1) is greater for Prototyping=0 than 1 |
| Right | 0.0242* | Prob(Significant Growth=1) is greater for Prototyping=1 than 0 |
| 2-Tail | 0.0242* | Prob(Significant Growth=1) is different across Prototyping |

| Fisher's Exact Test Table Probability (P) | Two-sided Prob ≤ P |
|---|---|
| 0.024242 | 0.0242* |

## Contingency Analysis of EGF Significant By ACAT I
## Mosaic Plot (PAUC Group 2, in the 48% bin)




## Contingency Table

ACAT I By EGF Significant

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 4<br>25.00<br>40.00<br>40.00 | 6<br>37.50<br>100.00<br>60.00 | 10<br>62.50 |
| 1 | 6<br>37.50<br>60.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 6<br>37.50 |
| Total | 10<br>62.50 | 6<br>37.50 | 16 |

## Tests

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 16 | 1 | 3.8548951 | 0.3642 |

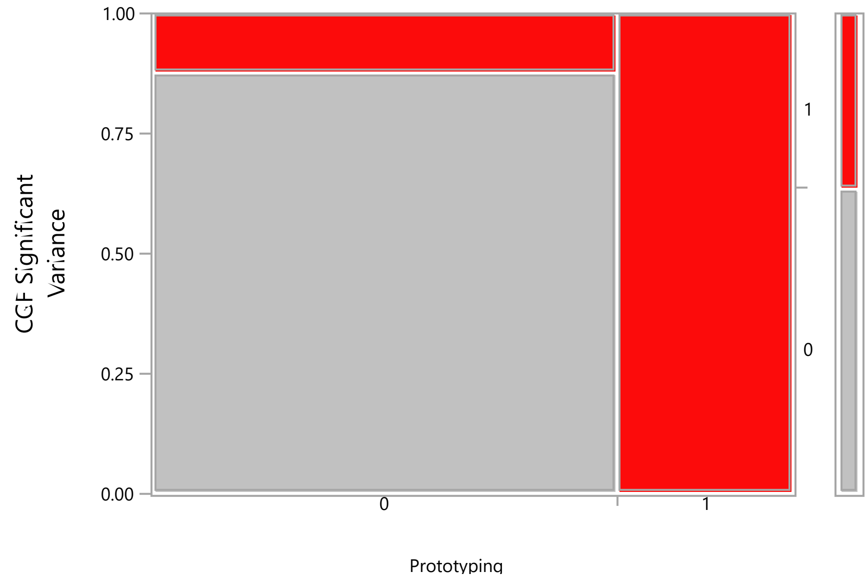
| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 7.710 | 0.0055* |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Pearson | 5.760 | 0.0164* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.0262* | Prob(EGF Significant=1) is greater for ACAT I=0 than 1 |
| Right | 1.0000 | Prob(EGF Significant=1) is greater for ACAT I=1 than 0 |
| 2-Tail | 0.0338* | Prob(EGF Significant=1) is different across ACAT I |

## Contingency Analysis of CGF Significant Variance By Prototyping
## Mosaic Plot (PAUC 21 bin, in the 35% bin)




## Contingency Table

Prototyping By CGF Significant Variance

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 7<br>63.64<br>100.00<br>87.50 | 1<br>9.09<br>25.00<br>12.50 | 8<br>72.73 |
| 1 | 0<br>0.00<br>0.00<br>0.00 | 3<br>27.27<br>75.00<br>100.00 | 3<br>27.27 |
| Total | 7<br>63.64 | 4<br>36.36 | 11 |

## Tests

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 11 | 1 | 4.1961382 | 0.5820 |

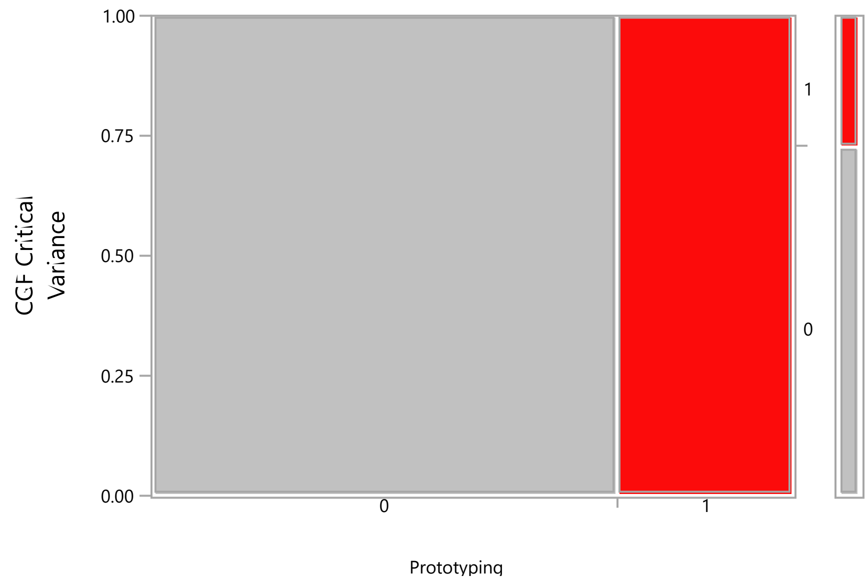
| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 8.392 | 0.0038* |
| Pearson | 7.219 | 0.0072* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 1.0000 | Prob(CGF Significant Variance=1) is greater for Prototyping=0 than 1 |
| Right | 0.0242* | Prob(CGF Significant Variance=1) is greater for Prototyping=1 than 0 |
| 2-Tail | 0.0242* | Prob(CGF Significant Variance=1) is different across Prototyping |

## Contingency Analysis of CGF Critical Variance By Prototyping

## Mosaic Plot (PAUC 21 bin, in the 35% bin)




## Contingency Table

Prototyping By CGF Critical Variance

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 8<br>72.73<br>100.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 8<br>72.73 |
| 1 | 0<br>0.00<br>0.00<br>0.00 | 3<br>27.27<br>100.00<br>100.00 | 3<br>27.27 |
| Total | 8<br>72.73 | 3<br>27.27 | 11 |

## Tests

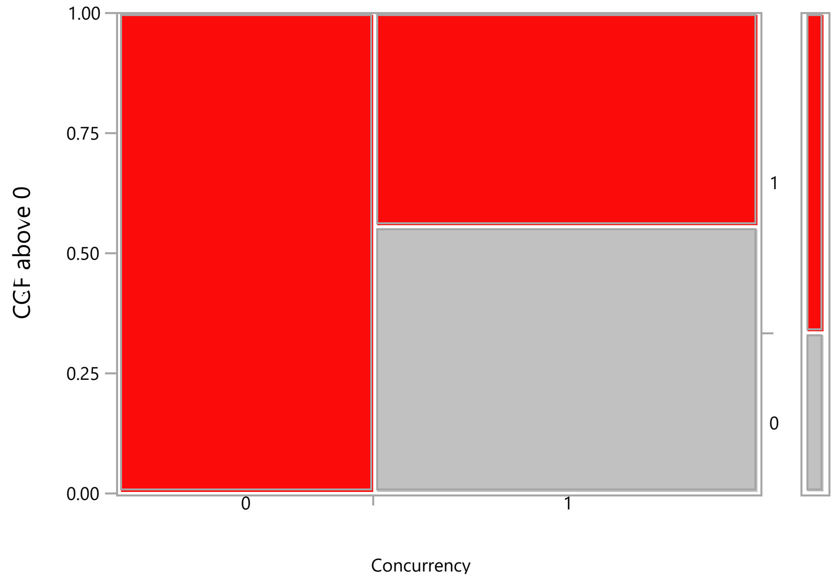
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 11 | 1 | 6.4454788 | 1.0000 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 12.891 | 0.0003* |
| Pearson | 11.000 | 0.0009* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 1.0000 | Prob(CGF Critical Variance=1) is greater for Prototyping=0 than 1 |
| Right | 0.0061* | Prob(CGF Critical Variance=1) is greater for Prototyping=1 than 0 |
| 2-Tail | 0.0061* | Prob(CGF Critical Variance=1) is different across Prototyping |

## Contingency Analysis of CGF above 0 By Concurrency
## Mosaic Plot (PAUC 21 bin, in the 0% bin)




### Contingency Table

Concurrency By CGF above 0

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 0<br>0.00<br>0.00<br>0.00 | 6<br>40.00<br>60.00<br>100.00 | 6<br>40.00 |
| 1 | 5<br>33.33<br>100.00<br>55.56 | 4<br>26.67<br>40.00<br>44.44 | 9<br>60.00 |
| Total | 5<br>33.33 | 10<br>66.67 | 15 |

### Tests

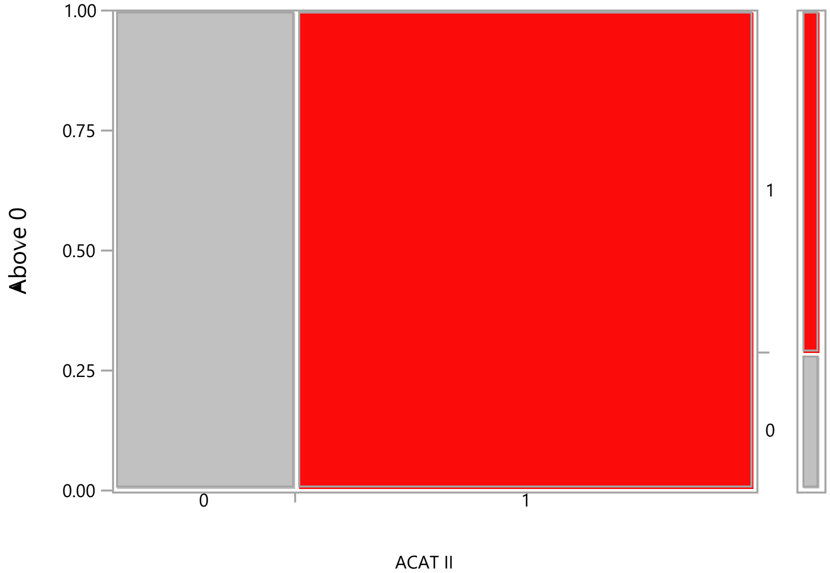
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 15 | 1 | 3.3650583 | 0.3524 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 6.730 | 0.0095* |
| Pearson | 5.000 | 0.0253* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.0420* | Prob(CGF above 0=1) is greater for Concurrency=0 than 1 |
| Right | 1.0000 | Prob(CGF above 0=1) is greater for Concurrency=1 than 0 |
| 2-Tail | 0.0440* | Prob(CGF above 0=1) is different across Concurrency |

## Contingency Analysis of Above 0 By ACAT II
## Mosaic Plot (APUC Group 2, in the 25% bin)




## Contingency Table

ACAT II By Above 0

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 2<br>28.57<br>100.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 2<br>28.57 |
| 1 | 0<br>0.00<br>0.00<br>0.00 | 5<br>71.43<br>100.00<br>100.00 | 5<br>71.43 |
| Total | 2<br>28.57 | 5<br>71.43 | 7 |

## Tests

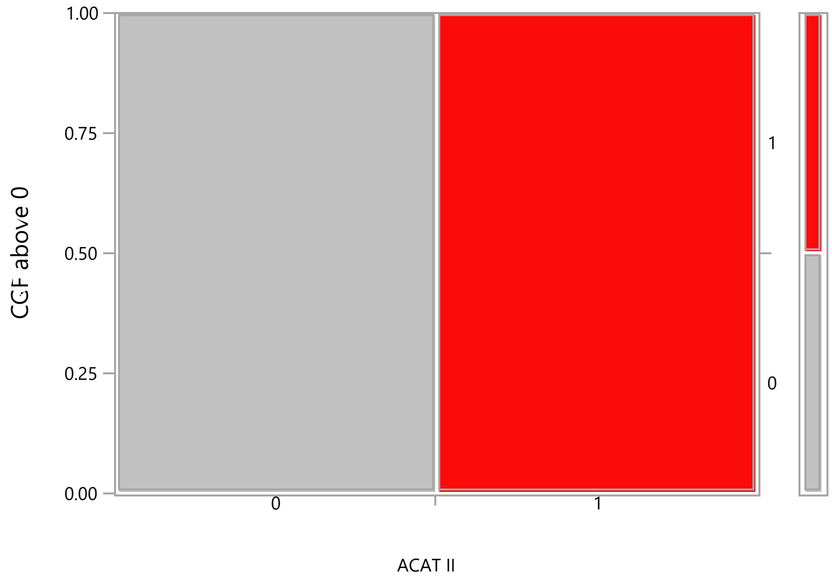
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 7 | 1 | 4.1878871 | 1.0000 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 8.376 | 0.0038* |
| Pearson | 7.000 | 0.0082* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 1.0000 | Prob(Above 0=1) is greater for ACAT II=0 than 1 |
| Right | 0.0476* | Prob(Above 0=1) is greater for ACAT II=1 than 0 |
| 2-Tail | 0.0476* | Prob(Above 0=1) is different across ACAT II |

## Contingency Analysis of CGF above 0 By ACAT II
## Mosaic Plot (PAUC 21 bin, in the 20% bin)




## Contingency Table

ACAT II By CGF above 0

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 3<br>50.00<br>100.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 3<br>50.00 |
| 1 | 0<br>0.00<br>0.00<br>0.00 | 3<br>50.00<br>100.00<br>100.00 | 3<br>50.00 |
| Total | 3<br>50.00 | 3<br>50.00 | 6 |

## Tests

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 6 | 1 | 4.1588831 | 1.0000 |

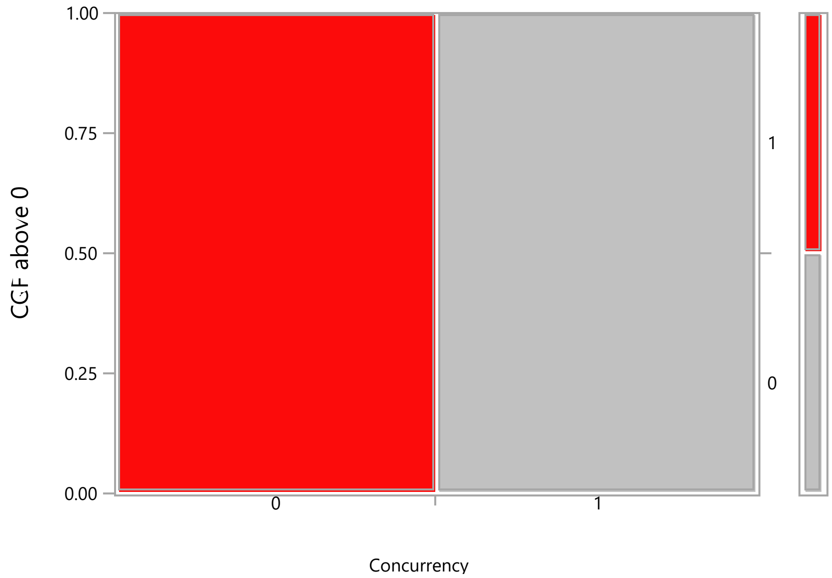
| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 8.318 | 0.0039* |
| Pearson | 6.000 | 0.0143* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 1.0000 | Prob(CGF above 0=1) is greater for ACAT II=0 than 1 |
| Right | 0.0500* | Prob(CGF above 0=1) is greater for ACAT II=1 than 0 |
| 2-Tail | 0.1000 | Prob(CGF above 0=1) is different across ACAT II |

## Contingency Analysis of CGF above 0 By Concurrency

## Mosaic Plot (PAUC 21 bin, in the 60% bin)




## Contingency Table

Concurrency By CGF above 0

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 0<br>0.00<br>0.00<br>0.00 | 4<br>50.00<br>100.00<br>100.00 | 4<br>50.00 |
| 1 | 4<br>50.00<br>100.00<br>100.00 | 0<br>0.00<br>0.00<br>0.00 | 4<br>50.00 |
| Total | 4<br>50.00 | 4<br>50.00 | 8 |

## Tests

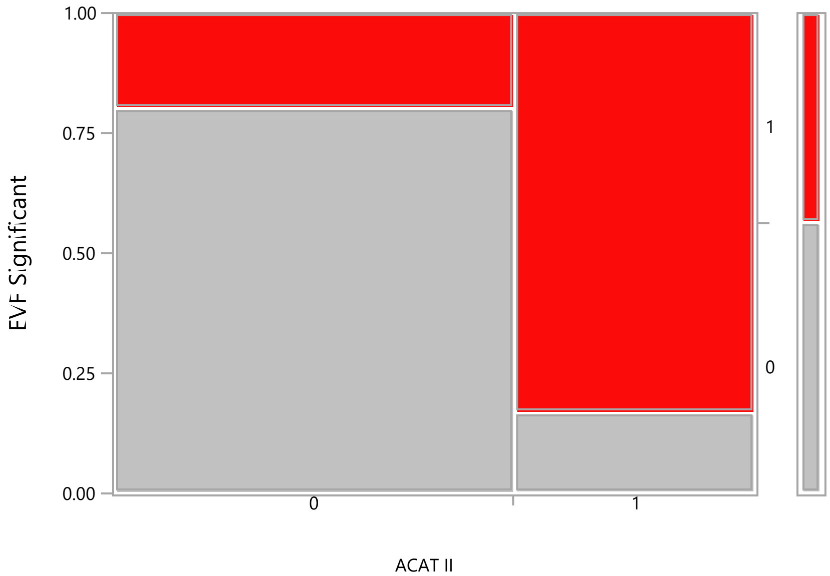
| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 8 | 1 | 5.5451774 | 1.0000 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 11.090 | 0.0009* |
| Pearson | 8.000 | 0.0047* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.0143* | Prob(CGF above 0=1) is greater for Concurrency=0 than 1 |
| Right | 1.0000 | Prob(CGF above 0=1) is greater for Concurrency=1 than 0 |
| 2-Tail | 0.0286* | Prob(CGF above 0=1) is different across Concurrency |

## Contingency Analysis of EVF Significant By ACAT II
## Mosaic Plot (PAUC Group 2, in the 48% bin)




## Contingency Table

ACAT II By EVF Significant

| Count<br>Total %<br>Col %<br>Row % | 0 | 1 | Total |
|---|---|---|---|
| 0 | 8<br>50.00<br>88.89<br>80.00 | 2<br>12.50<br>28.57<br>20.00 | 10<br>62.50 |
| 1 | 1<br>6.25<br>11.11<br>16.67 | 5<br>31.25<br>71.43<br>83.33 | 6<br>37.50 |
| Total | 9<br>56.25 | 7<br>43.75 | 16 |

## Tests

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 16 | 1 | 3.2576358 | 0.2971 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 6.515 | 0.0107* |
| Pearson | 6.112 | 0.0134* |

Warning: Average cell count less than 5, LR ChiSquare suspect.

| Fisher's Exact Test | Prob | Alternative Hypothesis |
|---|---|---|
| Left | 0.9991 | Prob(EVF Significant=1) is greater for ACAT II=0 than 1 |
| Right | 0.0245* | Prob(EVF Significant=1) is greater for ACAT II=1 than 0 |
| 2-Tail | 0.0350* | Prob(EVF Significant=1) is different across ACAT II |

## Bibliography

AcqNotes. *Acquisition Process: Exit & Entrance Criteria*, 2019. http://acqnotes.com/acqnote/acquisitions/exit-criteria

Arena, Mark A., Leonard, Robert S., Murray, Shelia E., Younossi, Obaid, *Historical Cost Growth of Completed Weapon System Programs*, Santa Monica, Calif,: RAND Corporation, 2006.

Asher, N. J., and T. F. Maggelet, *On Estimating the Cost Growth of Weapon Systems*, Institute for Defense Analyses, IDA Paper P-1494, September 1980.

Bielecki, J. V., E. D. White, *Estimating Cost Growth For Schedule Changes: A Regression Approach*, Cost Engineering, 47 (8), 28-34, 2005.

Birchler, D., Christle, G., & Groo, E. *Cost implications of design/build concurrency*. Defense Acquisition Research Journal, 18(3), 237–246, 2011.

Drezner, Jeffrey A., Jeanne M. Jarvaise, Ronald Wayne Hess, Paul G. Hough, and D. Norton, *An Analysis of Weapon System Cost Growth*, Santa Monica, Calif.: RAND Corporation, MR-291-AF, 1993.

Government Accountability Office. *Cost estimating and assessment guide* (Report No. GAO-09-3SP). Washington, DC: 2009.

Jimenez, C. A., White, E. D., Brown, G. E., Ritschel, J. D., Lucas, B. M., & Seibel, M. J. *Using pre-milestone B data to predict schedule duration for defense acquisition programs*. Journal of Cost Analysis and Parametrics, 9(2), 112–126, 2016. http://doi.org/10.1080/1941658X.2016.1201024

Kozlak, Scott J., White, Edward D., Ritschel, Jonathan D., Lucas, Brandon, & Seibel, Michael J. *Analyzing Cost Growth at Program Stages for DOD Aircraft*. Defense ARJ, Vol. 24 No. 3: 386-407, 2017. https://doi.org/10.22594/dau.16-763.24.03

Leonard, Robert, S., Akilah Wallace. *Air Force Major Acquisition Program Cost Growth Is Driven by Three Space Programs and the F-35A*. Santa Monica, CA: RAND Corporation, 2013.

McClave, J. T., Benson, P. G., & Sincich, T. *Statistics for business and economics* (10th ed.). Upper Saddle River, NJ: Pearson, 2008.

McDonald, J.H. Handbook of Biological Statistics (2nd ed.) Sparky House Publishing, Baltimore, Maryland, 2009.

McNichols, Gerald R. and Bruce J. McKinney. “Analysis of DoD Weapon System Cost Growth Using Selected Acquisition Reports,” *Sixteenth Annual Department of Defense Cost Analysis Symposium*. Arlington VA. 4-7 October 1981.

MCR Federal, LLC, & Air Force Cost Analysis Agency, *Chapter 14: Other Government Costs*, Air Force Cost Analysis Handbook, September 2007.

Moore, G. W., E.D. White., *A Regression Approach for Estimating Procurement Cost*, Journal of Public Procurement, 5 (2), 187-209, 2005.

Porter, Gene, Brian Gladstone, C. Vance Gordon, Nicholas Karvonides, R. Royce Kneece, Jr., Jay Mandelbaum, William D. O’Neil, *The Major Causes of Cost Growth in Defense Acquisitions*. Institute For Defense Analyses, Alexandria, VA, December 2009.

Schwartz, M., & O’Connor, C. V. *The Nunn-McCurdy act: Background, analysis, and issues for Congress* (Report No. R41293). Washington, DC.: Congressional Research Service, 2016.
Retrieved from https://fas.org/sgp/crs/natsec/R41293.pdf

Trudelle, Ryan, White, Edward D., Ritschel, Dan, Koschnick, Clay, & Lucas, Brandon, *Modeling median will-cost estimates for defense acquisition programs*, Journal of Defense Analytics and Logistics, Vol. 1 Issue: 1, pp.19-33, 2017.
https://doi.org/10.1108/JDAL-04-2017-0004

White, E.D., V.P. Sipple, M.A. Greiner, *Surveying Cost Growth*, Defense Acquisition Journal, 11 (1), 78-91, 2004

| REPORT DOCUMENTATION PAGE | *Form Approved* *OMB No. 074-0188* |
|---|---|



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From – To)* |
|---|---|---|
| 22-03-2019 | Master's Thesis | March 2017 – March 2019 |

**TITLE AND SUBTITLE**

AN ANALYSIS OF ESTIMATE VARIANCE IN PROGRAM OFFICE ESTIMATES

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**

Bonenfant, Benjamin J., Captain, USAF

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMINIMUMG ORGANIZATION NAMES(S) AND ADDRESS(S)**

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/ENV)
2950 Hobson Way, Building 640
WPAFB OH 45433-7765

**8. PERFORMINIMUMG ORGANIZATION REPORT NUMBER**

AFIT-ENV-MS-19-M-163

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Air Force Life Cycle Management Center
Cost Staff
Shawn Valentine, Civ, USAF
1865 Fourth Street, Building 14, Area B
Wright-Patterson AFB, OH 45433-7765
937-656-5504
AFLCMC.FZC.ResearchDataRequest@us.af.mil

**10. SPONSOR/MONITOR'S ACRONYM(S)**

AFLCMC/FM-FZ

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION/AVAILABILITY STATEMENT**

**DISTRUBTION STATEMENT A.** APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**13. SUPPLEMENTARY NOTES**

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

**14. ABSTRACT**

Past research has shown that predicting the cost growth within DoD systems is an important topic. Total program cost growth and predictors of program cost growth have been studied. Kozlak (2017) studied cost growth at four major reviews: Critical Design Review, First Flight, Development Test and Evaluation End, and Initial Operating Capability. This research attempts to assess cost growth and cost variance at similar points in a program life cycle. In the past the majority of studies have been done identifying programs as either: Acquisition Category (ACAT) I and non-ACAT I programs, or Major Defense Acquisition Programs (MDAP) and non-MDAP programs. This research has data that is able to highlight ACAT II and ACAT III programs. This research also attempts to create a CER for the relationship between Other Government Costs (OGC)-to-contract costs. The research is not attempting to definitively evaluate or confirm the effects of program characteristics, but is rather trying to guide the bolstering of POE databases and POE research. This database and POE research should highlight cost growth and cost variance for ACAT II and ACAT III programs. Such programs are not highlighted in Selected Acquisition Reports (SAR) or the current cost growth literature.

**15. SUBJECT TERMS**

Cost Growth, Program Office Estimates

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | | Lt Col Clay Koschnick, AFIT/ENV |
| U | U | U | UU | 111 | **19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-6565, ext 4638 clay.koschnick@afit.edu |

**Standard Form 298 (Rev. 8-98)**
Prescribed by ANSI Std. Z39-18