A method for valuing architecture-based business transformation and measuring the value of solutions architecture
Slot, R.G.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
This appendix describes the activities conducted for the raw data analysis and transformation phase, used to analyse the value of solution architecture, which is described in the chapters 6 and 7.

3.1 Budget Success Variable

3.1.1 Format of the Budget Success Variable
For each project the original budget and the actual budget was compared and standardized into a percentage. See the figure below for an overview of the budget success variable of these projects.

![Histogram of project budget success variable](image-url)
All projects are normalized to a standard scale. The value of 100 means that the project was exactly on budget, while, for example, the value of 120 means that the project was 20% over budget, etc.

3.1.2 Removing Anomalies – Budget Success Variable

Analyzing this histogram, we find that it has some unusual characteristics. First, the number of projects with 30% underrun (in the column that is labeled 70) is higher than the number of projects with 20% underrun; 5 versus 3. We considered several possible reasons for this increase and interviewed the involved project managers. The most plausible reason that emerged from this investigation is that project managers, in some cases, have overstated the cost of the project. Because of performance-related conditions, project managers have an incentive to calculate projects using margins that create an underrun.

A second noteworthy characteristic of the histogram is the presence of two projects with a score of 330, which means that there are two projects with 230% overrun. This overrun is surprisingly high, considering the gap between 140 and 190 and then the jump to 330. One would not expect two projects with this high overrun in a small sample of 49 projects. When investigating this issue, we found out that the selection of projects was not completely random. High-overrun projects cost a lot of money and attract a lot of management attention; they become well known. It is possible that these high-overrun projects were chosen because of the fact that they had high overrun and, consequently, that they are not representative for the total population of projects.

Both effects are aggravated because the number of projects is relative small; for both tails of the curve, the difference between high and low is only a few projects. Results that differ from the general trend, because of specific reasons that are not related to the trend, are called outliers. Because of the arguments above, we decided to remove these outlier results from our sample. The resulting histogram looks like:
3.1.3 Data Transformation – Budget Success Variable

We analyzed this distribution with Minitab for Normal and for Lognormal distribution characteristics, using a 95% confidence interval.

The analysis compares the frequency of the actual figures and compares them to a theoretical distribution, while trying to minimize the error between the theoretical value and the actual frequency. A red dot describes the results of one individual project. The middle line is the "ideal" Normal distribution and the higher and the lower lines denote the 95% confidence intervals. We find that many points are located far outside of the confidence intervals and we will therefore reject the hypothesis that the actual distribution represents a Normal distribution. In this case,
Minitab also provides a p-value for the probability of a normal distribution. The value is 0.5%, which is lower than the threshold value of 5%. Therefore, we reject the null hypothesis of normal distribution.

![Probability Plot for Budget](image)

*Figure 3-4. Analysis of project budget success variable for Lognormal distribution*

Analyzing the figure for a lognormal distribution, we find that one project falls outside the 95% confidence interval. However, the match is much better compared to the normal distribution. We decided not to reject the working hypothesis that the information is lognormal distributed.

The related parameter values for the lognormal distribution are: $\mu = 3.74, \sigma = 0.44$ and $\lambda = 64.9$. The most likely outcome is 100 and the expected outcome is 111. Based on this analysis, we can transform the budget variable normal distribution by using the following transformation instruction:

$$b_t = \ln (b - \lambda) \quad (3-1)$$

where

- $b_t = transformed \ budget \ value$
- $b = original \ budget \ value$
- $\lambda = threshold \ value = 64.9$

See Figure 3-5 for the results of the transformation of budget overrun.
The p-value for this distribution is 15%, therefore we do not reject the null hypothesis that the transformed data is normal distributed.

### 3.2 Raw Data Analysis and Transformation – Time Success Variable

#### 3.2.1 Data Transformation – Time Success Variable

The time success variable is also normalized to a standard scale with a value of hundred indicates that the project was on time, a value of 150 indicates that the project was 50% over time, etc. Following the same steps that we used for the budget distribution, we were also able to find a lognormal transformation instruction for the time variable.
Analyzing the figure for a lognormal distribution, we find that no project falls outside the 95% confidence interval. We decided not to reject the working hypothesis that the information is lognormal distributed.

The related parameter values for the lognormal distribution are: $\mu = 3.0$, $\sigma = 1.3$ and $\lambda = 94.3$. From the distribution follows, that the most likely outcome is 98 and the expected outcome is 141. Using formula (3-1) with a $\lambda$ value of 94.3, we transformed the time variable to the normal distribution. See the histogram of Figure 3-7.
The p-value for this distribution is 10%, therefore we do not reject the null hypothesis that the transformed data is normal distributed.

3.3 Reliability of Transformation for Budget and Time
The normal distributions of the budget and time success variables falls within the confidence intervals for normal distribution, albeit sometimes it is close. The basic problem that we face with the identification of the distributions is the limited number of data points that are available. Use of several hundred data points gives a more accurate understanding of the distributions. Pyzdek states: “the approach works best if you have at least 200 data points, and the more the merrier” (2003). However, only a limited number of projects were available for analysis and we will have to do with the information that we have. To our advantage, the ANOVA test that we use is rather robust to deviations from normality (Lindman, 1974). This allows us to use the ANOVA test even when the actual distribution of the data does not fully comply with the normal distribution.

3.4 Raw Data Analysis – Customer Satisfaction Success Variable
Customer Satisfaction is measured using a standard procedure and a standard scale, across all projects. The scale goes from 1 (very bad) to 5 (very good). The procedure that is used to measure customer satisfaction is called OTACE (On Time above Customer Expectation) and it is a standard procedure were the customer gives a weighted score, and based upon a number of criteria. The criteria (and the priority of each criterion) are defined before the project starts and scored afterwards. For the measurement of the customer satisfaction, 44 projects were available.
We were not able to determine a standard distribution type for this distribution. Because the variable is numerical and continuous, we will use regression for the analysis of this variable.

### 3.5 Raw Data Analysis – Percentage Delivered Success Variable

The success variable *Percentage Delivered* is represented in a percentage, where 100% indicates that all the required functionality was delivered. In the actual results, we find that the minimum was 60% and the maximum was more than 120%. In this last case, we find that the project delivered more functionality than those planned. Most of the projects deliver 100% of the planned functionality.
As the variable is numerical and continuous, we will use regression for the analysis of this variable.

### 3.6 Raw Data Analysis – Functional and Technical Fit Success Variable

The success variables Functional Fit and Technical Fit are both represented as a multiple-choice variable with three possible choices. The three choices are:

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional fit of the project results</td>
<td>&gt;90% of functional requirements are realized, including all the 'must haves'</td>
<td>What is the percentage of defined functions realized at the end of the project?</td>
</tr>
<tr>
<td>Technical fit of the project results</td>
<td>The project met the main technical requirements</td>
<td>Balance in the non-functional requirements of project results, such as maintainability, resilience, scalability, security, availability, etc.</td>
</tr>
<tr>
<td></td>
<td>Between 80 and 90 percent of the functions are realized</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Less than 80% of the functions are realized</td>
<td></td>
</tr>
</tbody>
</table>

*Table 3-1. Answers for functional and technical fit (from questionnaire)*

See the figures below for an overview of the histogram of the answers.
For the analysis of these success variables, we use the Kruskal-Wallis test.