****

**Agile and Scrum - An Empirical Study of the Feasibility of Story Points**

Master Thesis

Name of the Study Programme

Project Management and Data Science

**Faculty 3**

from

Damian Spencer

Date:

Berlin, 16.03.2021

1st Supervisor: Professor Dr. Tilo Wendler

2nd Supervisor: Stefanie Supper

**Index**

[1 Abstract and Introduction 1](#_Toc66881609)

[1.1 Abstract 1](#_Toc66881610)

[1.2 Introduction 1](#_Toc66881611)

[2 Motivation 4](#_Toc66881612)

[3 Theory 7](#_Toc66881613)

[3.1 “Agile” 7](#_Toc66881614)

[3.2 What is Scrum 8](#_Toc66881615)

[3.2.1 Original intent – The Agile Manifesto 9](#_Toc66881616)

[3.3 Agile and Scrum 10](#_Toc66881617)

[3.3.1 Affiliation 10](#_Toc66881618)

[3.3.2 Comparison to Traditional Methods 12](#_Toc66881619)

[3.3.3 Industry and Evolution 13](#_Toc66881620)

[3.3.4 Scrum Roles 15](#_Toc66881621)

[3.3.5 Scrum Artifacts 16](#_Toc66881622)

[3.3.6 User Stories 16](#_Toc66881623)

[3.3.7 Story Points 18](#_Toc66881624)

[3.3.8 Story Point Estimation (Planning Poker) 20](#_Toc66881625)

[3.4 Agile and Scrum Critique 23](#_Toc66881626)

[3.5 Summary of Theoretical Findings 25](#_Toc66881627)

[3.6 Professional Experience and Practical Application 26](#_Toc66881628)

[4 Empiricism 28](#_Toc66881629)

[4.1 Outline and Procedure 28](#_Toc66881630)

[4.2 Case Study and Findings 30](#_Toc66881631)

[4.2.1 Case Study 30](#_Toc66881632)

[4.3 Summary of Empirical Findings 48](#_Toc66881633)

[5 Summary and Contribution 50](#_Toc66881634)

[6 Discussion 55](#_Toc66881635)

[List of Literature 61](#_Toc66881636)

[Statutory Declaration 68](#_Toc66881637)

**List of figures**

[Figure 1. Survey results of different Agile methodologies used globally 11](#_Toc66881672)

[Figure 2. Waterfall Model excerpt from Clinical Engineering 2nd Edition – Chapter 916 12](#_Toc66881673)

[Figure 3. Project Management Institute Fees and Costs 14](#_Toc66881674)

[Figure 4. MS DevOps hierarchy of work items37 18](#_Toc66881675)

[Figure 5. Example of a Burn down chart 19](#_Toc66881676)

[Figure 6. Demonstration of Case Study procedure 28](#_Toc66881677)

[Figure 7. Example of Product Backlog in Microsoft Azure DevOps 31](#_Toc66881678)

[Figure 8. Example of time logged Actual time spent 31](#_Toc66881679)

[Figure 9. Script showing summary of Case Study data 33](#_Toc66881680)

[Figure 10. a & b) Showing histograms of Estimate and Actual Hours 33](#_Toc66881681)

[Figure 11. a & b) Showing Q-Q Plots of Estimate and Actual Hours 33](#_Toc66881682)

[Figure 12. Script extract showing results of Shapiro-Wilk normality test 34](#_Toc66881683)

[Figure 13. Script extract showing result of Kruskal-Wallis test 35](#_Toc66881684)

[Figure 14. Script showing results of pairwise Wilcox Rank Sum (two-sided) Test 36](#_Toc66881685)

[Figure 15. Story Point Categorical Box Plot 36](#_Toc66881686)

[Figure 16. Script showing results of pairwise Wilcox Rank Sum (greater than) Test 37](#_Toc66881687)

[Figure 17. Script extract showing results of Cramer's V Test of Association 38](#_Toc66881688)

[Figure 18. Script extract showing results of Spearman's Rho Test of Association 39](#_Toc66881689)

[Figure 19. Extract demonstrating sample of created dummy variables 40](#_Toc66881690)

[Figure 20. Correlation Matrix of Dependent Variables 42](#_Toc66881691)

[Figure 21. Variable Inflation Factor Formula 43](#_Toc66881692)

[Figure 22. Variable Inflation Factor Formula (continued) 43](#_Toc66881693)

[Figure 23. Variable Inflation Factor Results 44](#_Toc66881694)

[Figure 24. Script extract showing results of Linear Regression Model 45](#_Toc66881695)

[Figure 25. Graph showing Linear Model Residuals (Homoscedasticity) 45](#_Toc66881696)

[Figure 26. Histogram of Linear Model Residuals 46](#_Toc66881697)

[Figure 27. Script extract showing test and results of Kolmogorov-Smirnov Test 46](#_Toc66881698)

[Figure 28. Script showing results of Test of Independence (Pearson’s Correlation Test) 47](#_Toc66881699)

[Figure 29. Poll about the Elements of Story Points 51](#_Toc66881700)

[Figure 30. Script of Stepwise function results 57](#_Toc66881701)

[Figure 31. Stepwise Selection Lineal Model Residual Plot 57](#_Toc66881702)

[Figure 32. Histogram of Stepwise Selection Linear Model Residuals 58](#_Toc66881703)

[Figure 33. Script of Kolmogorov-Smirnov Test for Residuals of Stepwise Selection Model 58](#_Toc66881704)

[Figure 34. Script showing test of Independence of Residuals of Stepwise Selection Model 59](#_Toc66881705)

List of tables

[Table 1. Direct comparison of Waterfall to Agile 13](#_Toc66881706)

[Table 2. Suggested method for using Story Points 52](#_Toc66881707)

# Abstract and Introduction

## Abstract

This paper examines the project management “Agile” framework through the lens of a real-world example. The paper asserts that “Agile” and derivative methodologies such as Scrum, as they are currently implemented and promoted today, have veered from the original intent of the philosophies from which they have been derived. Furthermore, the paper leans toward the view that certain techniques under Scrum are “non-agile”. One of these techniques – estimation of work units and velocity using “Story Points” – will be observed using a case study based on empirical data amassed from a product backlog of a live project. The findings will be discussed and recommendations for an improved estimation framework will be given. While this paper’s empirical focus is on whether the way Story Points are current and most popularly used would be best serving as an estimation measure for work units, it may be more broadly useful for those who seek understand of what being “Agile” was intended to look like by those who founded the framework.

## Introduction

The practice of agile frameworks for project management has evolved significantly over time. As part of this evolution, while many techniques and processes have been developed, some have been feasibly useful while others may be seen as resulting from groupthink and a bandwagon effect whereby the new attractive unknown object grabs the attention of professionals. This paper Stems from a desire to investigate the opinion that Story Points measurement is one of those shiny objects and the use of Story Points as a tool used for estimation of User Stories (as well as for measuring burn down) is ineffectual. This study will empirically test the hypothesis that **as Story Points are currently and most popularly used (Planning Poker) they are arbitrary and not feasible as an estimation measure specifically in an environment where deadlines are imperative**.

In this paper a theoretical review of Agile frameworks and Scrum methodologies (to which Story Points has been appended) is conducted. Whilst demonstrating that Agile Frameworks have brought significant benefit to businesses in many industries worldwide, the paper points to advocacy about Agile and Scrum, from the very founders of these frameworks and techniques, that these methods of project management stem away from their original intent. The review highlights the concerns of the founders that their work has been transformed and made to be marketable and, as such, industrialised. The writer, references and applies this same argument to this argument with respect to Story Points.

There are many protagonist and antagonist arguments that support the use of and advocate against the use of Story Points for User Story measurement estimation. These arguments are similar to the general lines of arguments one may observe about the ongoing debate about relative versus absolute estimation techniques – from the ease and quickness of relative estimates to the ability to relate absolute estimates to actual results in the same unit of measurement regarding absolute estimates. While there are many of the aforementioned arguments, the writer has not found any (resembling) study which uses the data available in this instance and this provides a unique situation. More specifically what makes the study potentially rare is not just the capture of data using both estimation techniques but the recording of actual time spent on units of work in addition.

Given the collection of this empirical data, the writer capitalises on this opportunity to move beyond a theoretical or subjective assessment, to that of an objective assessment which, although specific to the case study upon which the assassinated based, may be transferable as a use case and applied to other projects/states. The goal therefore, is to perform various statistical analyses which to test initial hypothesis and provide the formula for understanding the relationship between the two estimation techniques, overall resulting in the development a methodology which can be used by other project teams where there may be contention between project teams about using either estimation technique or where estimation using either technique resulted in project delays.

The case study is derived from a live company project where a revolutionary product is being launched. The parent company, referred to as “Company A”, hired a contractor, referred to as “Company B”, to assist with project management and product development. An Agile/Scrum methodology was implemented in managing the project. Given a significant delay in the launch of the product, Company A’s management, which already had reservations about estimating the project timeline as well as Sprints using Story Points, required a change in estimation from Story Point to time estimation. This business scenario provides the primary basis of this paper as well as the empirical data from the different estimates of time and Story Points, and the recorded actual time.

Recommendations will be made for how to assess either of the two estimation methods under discussion going forward. Finally, there will be a there will be an acknowledgement of the circumstances specific to this case and the empirical data associated, and how these considered/applied to other projects.

# Motivation

This thesis is inspired from personal/professional experience and an opinion of Agile and Scrum both as a student and working student.

While working as an Accounting Advisor, the author was a member of a team that was tasked with testing a new software product which would potentially use for reconciliations. Despite learning of the term at undergraduate degree level, this was a first introduction to project management in a professional setting. Not happy reporting to an auxiliary manager outside of daily and monthly tasks, there was a quick realisation of the earnestness and support project managers had from senior management; this was most noteworthy.

A couple years later, on a second occasion after being promoted to Senior Accounting Advisor, another more directly involved project abounded focused on the implementation of a different software application. This project, was noticeably more structured and the role played referred to on the development team was as a tester. Again, the backing that this auxiliary project team received from top leadership within the company was remarkable.

Later as a professional, after an assessment, a career change needed to be made from accounting to something that aligned itself more closely with coding and technical expertise. There was a need to develop new skills in order to remain relevant in the job market many years in the future. After pondering on many options for pursing a more technical field of study, including software programming and web development, it was consultation with a colleague who studied predictive data analytics (at a top university) that held the most appeal. Doing further research, the decision was made that data science would be a viable field of study to pursue.

The challenge then became more academic. Having a non-technical (non-computer science related) undergraduate degree, made being accepted into a graduate program for a computer programming related course more difficult. Fortunately found the Master of Project Management and Data Science at the Hochschule für Technik und Wirtschaft – Berlin catered for prospective students seeking a career change regardless of undergraduate background. What was most captivating about the opportunity was the project management portion of the degree program. Still not fully convinced that project management was most viable for employment, recollection of the importance it was given in previous employment experiences helped in trusting its value. The application was completed and fortunately there was acceptance.

During graduate student coursework in project management, although the terms were somewhat known to him, the author was introduced to and became more familiar with ‘Agile’ and Scrum (methodologies) yet still, the topic did not hold much favour and scepticism about the practices remained. As is customary with any course, there was a heavy amount of theory being taught initially but not until in the second semester when first practical experience was attained. In a practicum with a company, while on a team with fellow classmates, the team was involved in a real Scrum project where the company in question was the client. As part of the assignment, each member of the team was tasked, at least during one Sprint, with assuming each the Scrum roles. Seeing the theory that was learned in the classroom being applied in the classroom, especially the iterative approach that is the nature of ‘Agile’ and Scrum, the author began become more convinced and changed views about project management in general to a more favourable one.

Fortunately, another practical exposure to the methodologies was presented; this time in the classroom. For another team exercise, students were required to build a project beginning through end, the final result being a generated financial valuation report. This was the first introduction to Story Points. Although, the theory that was taught in the classroom leaned more heavily on the side of the pros of using Story Points (in comparison to the cons) the author was remained an antagonist when undergoing the practical exercise. Agreement simply could not be found with the arguments made that were learned during the lectures when building User Stories and assigning Story Point values, making the exercise quite difficult.

Upon assuming a role as a Project Manager at a most recent internship at the time, there had been a considerable delay in the launch of the product that was being pursued. The delay was attributed to a few factors but relative estimation was deemed as detrimental to foreseeing the delay as a result of the use of Story Points. An argument was made by the head manager that Story Points are arbitrary and ineffectual for measuring the scope/degree of work and given the delay, the more absolute alternative – hours should be used and the project proceeded as such. Upon seeing someone with such seniority and with a great deal of experience sharing the same perspective, confidence was gained in the viewpoint of Story Points not being the most feasible measure of estimation.

A further protagonist argument for using the absolute method is the consideration of timesheet records. Given the need for employees (specifically consultants) to both estimate and report to clients in hours or days of work versus an alternative relative metric, employees become experienced at estimating the time they would require to execute tasks. Introducing the relative measure takes away from that knowledge, experience and learned gut feeling that had been learned and honed over time. This argument resounded and coming from top management, this spurred further desire to look to examine the hypothesis of one being more reasonably practical than the other.

Moreover, given the title of the Master’s degree program being Project Management and Data Science, exploring this topic of Story Point estimation using the empirical data that was available, provided a suitable blend of diving in to the practical elements of project management using statistical analysis to empirically examine the feasibility for using the absolute versus the non-practical method of estimation.

The topic of relative versus absolute measures of estimation can be quite long-winded but the rather unique situation of this project is that both sets of information were being recorded for each User Story and sub-tasks. Enquiring from previous employers and from colleagues who are Project Management professionals to various other sources, if both data sets are recorded in this manner was to no avail. Therefore, an opportunity is presented to assess this element of estimation empirically and it may be of benefit those who read this paper.

# Theory

## “Agile”

The term “Agile” stems from what has become known as the “Agile Manifesto” which was written by a small group of experts in their respective fields. **[[1]](#footnote-1)** The manifesto was a result of an effort to pioneer, revolutionise and improve software development processes – the standard of which initially followed a document driven and heavyweight approach but has now evolved to being more lightweight. But what does this mean for businesses?

Atlassian, the producer of the most widely used project management software “Jira”, [[2]](#footnote-2) defines “Agile” as “an iterative approach to project management and software development that helps teams deliver value to their customers faster and with fewer headaches”. [[3]](#footnote-3) Forbes, a reputable American business magazine, in its published article, “What is Agile?” states that, “Agile enables organisations to master continuous change”. [[4]](#footnote-4) The Agile Alliance, an organisation initiated by some of the members of the original concept of agility for software development[[5]](#footnote-5),defines Agile as, “a way of dealing with, and ultimately succeeding in, an uncertain and turbulent environment[[6]](#footnote-6).”

What practical benefit has evolved as a result of the development of “Agile” (and its methodologies) in business today? In a recent global survey of individuals in the software development community, 95% of respondents across diverse industries indicated that their companies have adopted Agile methodologies with 70% of respondents reporting that an ability to manage changing priorities is the primary benefit of Agile (Digital.ai, 2021, p. 10).[[7]](#footnote-7) A fair alignment of these definitions and reporting – Agile is an iterative approach tosuccessfullydelivering value to customers in a turbulent environment with continuously of changing priorities.

## What is Scrum

From the sport of Rugby, professors Hirotaka Takeuchi and Ikujiro Nonaka were guided to a concept which was co-authored in their Harvard Business Review paper in 1986 entitled, “The New Product Development Game” (Takeuchi, H. and Nonaka, I., 1986). In this paper, the authors advocate for a novel and more holistic approach to product development, making an analogy to rugby whereby “a team tries to go the distance as a unit, passing the ball back and forth”. They argue that the method of a team progressing forward holistically would provide more speed and flexibility which companies were quickly realising as necessary for keeping up with competition. The old, sequential relay-race method of specialists passing the baton to the next group was no longer adequate.

World Rugby, the global governing body for the sport of rugby defines “Scrum” as a means of restarting play after a stoppage which has been caused by a minor infringement. [[8]](#footnote-8) As a metaphor for rugby, Jeff Sutherland, inventor and co-creator of what is known today as “Scrum” (the project management process) borrowed this term. The cofounders modelled the Scrum methodology akin to rugby whereby (Sutherland, J., 2002, p. 2):

* The context is set by playing field (environment) and rugby rules (controls).
* The primary cycle is moving the ball forward.
* Rugby evolved from breaking soccer rules - adapting to the environment.
* The game does not end until environment dictates (business need, competition, functionality, timetable). Agile and Scrum – Industry and Evolution

Expanding on the third bullet, from a historical context, Webb Ellis, a seventeen-year-old rugby student, initiated rugby (which then eventually became codified in 1839) by breaking the rules of a soccer game – While playing soccer for an old college in East Warwickshire, seeing time running out while his time is behind, Ellis defied the rules, picked up the ball and ran with it to the goal. Scrum, therefore, has an air of rule breaking.

As a result, characteristics of a Scrum project defy traditionally conventional method of product/software development and project management (Schwaber K., 1995, p. 16). These include:

* Flexible deliverable - the content of the deliverable is dictated by the environment.
* Flexible schedule - the deliverable may be required sooner or later than initially planned.
* Small teams - each team has no more than 6 members. There may be multiple teams within a project.
* Frequent reviews - team progress is reviewed as frequently as environmental complexity and risk dictates (usually 1 to 4-week cycles). A functional executable must be prepared by each team for each review.
* Collaboration - intra and inter-collaboration is expected during the project.
* Object Oriented - each team will address a set of related objects, with clear interface and behavior.

### Original intent – The Agile Manifesto

The “Agile Manifesto”, as it is commonly referenced today, is actually the “Manifesto for Agile Software Development”. This manifesto originated in the year 2001 by seventeen professionals who gathered – with representation from the fields of Extreme Programming, Scrum and other aspects of software development and project management. The purpose of that gathering was to develop an alternative to document-driven fixed processes with the goal of delivering good products to customers genuinely girded around people being the most important asset. The result was a manifesto which promoted the following principles:

* Deliver working software frequently with a preference to the shorter timescale.
* Business personnel and developers must work together daily throughout the project.
* Build projects around supported, and trusted motivated individuals.
* Face-to-face conversation is most efficient and effective for conveying information.
* Working software is the primary measure of progress.
* Promote sustainability. Stakeholders should maintain indefinite constant pace.
* Continuous attention to technical excellence and good design enhances agility.
* Simplicity – the art of maximizing the amount of work not done – is essential.
* The best architectures, requirements, and designs emerge from self-organizing teams.
* Regularly reflect on becoming more effective, tune and adjust behavior.

Regarding an approach to achieving agility, perhaps Dave Thomas, one of the originally signatories of the ‘Manifesto’, in his address at the GOTO conference[[9]](#footnote-9) – “created for developers by developers; a series of software development conferences focused on gathering the best minds in the software community to bring the most interesting topics to light” – in 2015 proposed a noteworthy approach. Paraphrasing his comments, agility is, “three steps plus a loop. Understand where you are, take a small step to where you want to be and at the end of that step, evaluate what happened. And then repeat.”

From these principles, many practical applications have been developed or categorised under what is referred to today as “Agile”; these include “Scrum”, “Kanban” and “Lean Development” to name a few examples. This paper’s focus will remain on Scrum. A customised version of Scrum was used for the case study (which will be addressed later in this paper).

## Agile and Scrum

### Affiliation

Throughout the project management industry, the terms “Agile” and “Scrum” are used interchangeably. Writer and Search Engine Optimisation specialist for Northeastern University, Scott W. O’Connor, wrote, “On the surface, it is easy to see why Agile and Scrum can often be confused, as they both rely on an iterative process, frequent client interaction and collaborative decision making.” [[10]](#footnote-10) As stated by the Scrum Alliance, a non-profit certifying body who nurture the agile movement by offering education, advocacy, research, community and connection, “The difference between Agile and Scrum is that agile refers to a set of principles and values shared by several methodologies, process and practices; Scrum is one of several agile frameworks—and is the most popular[[11]](#footnote-11).” Here is where the distinction lies.

Scrum being the most popular framework of Agile is bolstered by multiple surveys. In the 14th “State of Agile” survey referenced above (the largest, longest running and most quoted Agile survey that exists) Scrum (and related variants) is reported in the year 2020 as the most commonly used Agile methodology at a minimum of 75% compared to various other methodologies[[12]](#footnote-12) – refer to Figure 1.[[13]](#footnote-13) This survey had a global audience of over one thousand respondents.

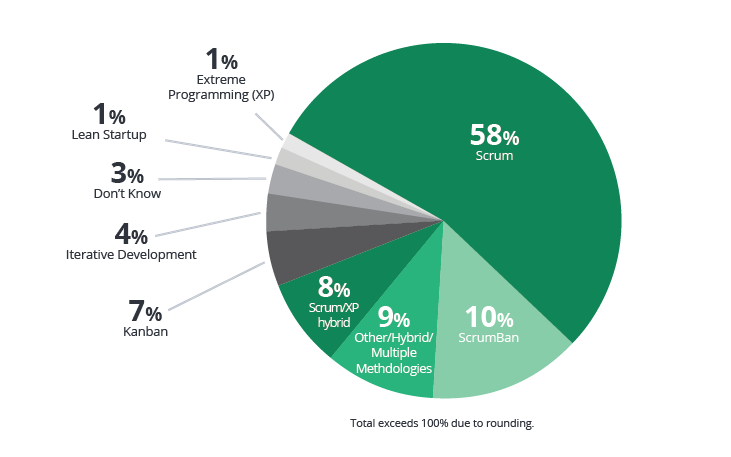


Figure 1. Survey results of different Agile methodologies used globally

The Scrum Alliance reported that approximately 72% respondents stated (some variant of) Scrum is used as its most common Agile methodology in its 2017-2018 “State of Agile” global survey (Digital.ai, 2018, p. 9).

According to the Agile Alliance the Scrum Master is a role within the team that is tasked with the responsibility of ensuring that the team “lives agile values and principles and follows the processes and practices that the team agreed they would use. [[14]](#footnote-14)” The Scrum Master role, while developed under the structure of Scrum, and highly respect as one of which the person filling that role is recognised as an expert in Scrum who can mentor to others, does not hold and true authority in the Scrum process. The Scrum Master role is aligned more closer with that of servant leadership – guiding and facilitating the process.

The term Scrum Master, has even become the term which refers to that role designated as the expert of a team which follows/practices an Agile methodology for its project(s) but is not necessarily adapting Scrum as the Agile methodology. Scrum Master has become interchangeable with “Iteration Manager”, “Agile Coach” or “Team Coach”.

Atlassian conducted a Twitter poll seeking to understand whether practitioner of Agile and Scrum adhere strictly to the guides of Scrum or some custom/hybrid version. Ninety-two percent of respondents responded saying that they practiced some customized version of Scrum.[[15]](#footnote-15) Atlassian then pursued finding out what this could potentially translate to for the role of the Scrum Master and how this would fit in a non-strictly adherent Agile realm.

The company observed that various teams that are on-boarding “Agile” in their organisation amongst their team members place significant dependence on Scrum Master as the owner of the process. Synonymous to the Scrum Master in the non-agile realm, the Project Manager whose primary focus is the method of executing work and the resolution of workflow problems that may arise through facilitation.

Atlassian strongly advocates for the use of a Scum Master if and only if the organisation the organisation involved understands, full commits to and implements Scrum properly. Considering this, with regards to the survey results, as the facilitator – the servant leader, should be able to find a place in any version of customised version of Agile

### Comparison to Traditional Methods

While modern project management methods like Scrum have been developed in the last two to three decades, traditional methods still thrive and are especially effective depending on the company or industry. Known as the most commonly used traditional methodology (Long, D. et al., 2020, p. 131-168) “Waterfall” is characteristic of heavy up-front planning and rigid, non-overlapping processes where subsequent process are almost entirely dependent upon completion of preceding processes.

The waterfall model was first popularised by Dr. Winston Royce in the year 1970 pertaining to software development concepts (Van Casteren, W., 2017, p. 1-6). Characteristically, the model generally cascades in the following phases, also known as milestones, shown in the Figure 2. below.

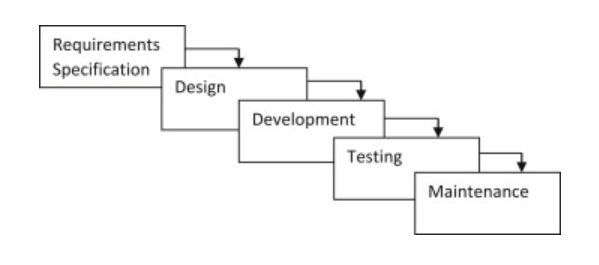


Figure 2. Waterfall Model excerpt from Clinical Engineering 2nd Edition – Chapter 916

|  |  |  |
| --- | --- | --- |
|  | Waterfall | Agile |
| Plan | Detailed, long-term | Iterative, Shorter planning |
| Team | Defined. Rigid team roles | Flexible. Cross-functional |
| Change Adaptability | Costly. Change is discouraged | Expected. Less impactful |
| Product Delivery | Completely. Delivered at project end | Functionally delivered in increments |
| Requirements | Contract-based predefined scope | Collaborative and interactive regarding scope |
| Customer Involvement | Only at beginning and end of project | Involvement throughout the project |
| Dependencies | Heavy dependency upon predecessor stages | Reduction of dependencies |

Table 1. Direct comparison of Waterfall to Agile[[16]](#footnote-16)

### Industry and Evolution

Structure has been brought to organise the skills and techniques under project management – so much so that Project Management has become a thriving profession. According to the Project Management Institute, the self-proclaimed leading association for those who consider project, program or portfolio management their profession, [[17]](#footnote-17) in Germany, the median salary of a Project Manager with a master’s degree is the equivalent of approximately $97,000 United States Dollars (USD) per year. Also, the median yearly salary of a Project Manager with a master’s degree in the United States is $120,000 USD according to Northeastern University[[18]](#footnote-18).

Supporting the profession is an entire Project Management educational industry which has developed over the last couple decades. Today there are many professionally recognised certificates that promote careers in (areas related to) Project Management. The Project Management Institute (PMI) offers 13 certifications[[19]](#footnote-19) and the Scrum Alliance offers 8 Scrum certifications[[20]](#footnote-20) per their respective websites.

For the fiscal year 2018 the Project Management Institute reported an annual revenue of just over 220 million dollars[[21]](#footnote-21)and the Scrum Alliance just under 18.6 million dollars. Figure 3.[[22]](#footnote-22) has the trending of dues and exam fee revenue and advertising cost for the last 4 years obtained from the Project Management Institute. The increase in revenues indicates potentially either an increase in membership, demand for certification or willingness to pay increase price for attaining membership/certification. The PMI’s increase in advertising cost is an indication its going concern (desire to continue its operations).

Figure 3. Project Management Institute Fees and Costs

With respect to courses, a brief internet search will reveal scores of them, the cost of which can be relatively expensive to the aspiring professional. For the certifications mentioned above, courses range from $225 to $800 for members ($300 to $1,000 for non-members) from the PM and just the Scrum Master certification offered by the Scrum Alliance, can cost as much as 2,395 Euros in Germany. 21

Also facilitating this evolution is the development of several project management software tools. There is a plethora of companies that utilize such software tools. “Jira” is currently the most widely use software with over one hundred thousand companies incorporating the software in its business practices for a Project-Management-appropriate purpose; “Microsoft Azure DevOps” is used by over fifty thousand companies worldwide.[[23]](#footnote-23) The hybrid cloud approach – combining private cloud with public cloud services, with proprietary software – that is offered, augments the desire for these software tools[[24]](#footnote-24).

The result of the growing demand for improved process and knowledgeable professionals, the marketing to promote the profession and software that supports it, as well as the investment associated, is an approximate 5-billion-dollar Agile enterprise transformation service market in 2018 projected to be just over 18 billion dollars in the year 2026.[[25]](#footnote-25)

### Scrum Roles

The Scrum process defines three main roles – the Product Owner, the Scrum Master and the Development Team.

The Product Owner is the primary visionary for a project; therefore, it is imperative that Product Owners understand the value that is to be delivered through the project and all the work that needs to be done to ensure a project’s success. As a result, their active involvement in all aspects of a project is critical.[[26]](#footnote-26) The Product Owner’s primary responsibility is prioritising work. In understanding the value that is be delivered, the Product Owner steers the project, solving any conflicts in conflicting priorities that may develop[[27]](#footnote-27).

The Scrum Master is the protector of Scum – protecting the Scrum team and Scrum process.29 With respect to the Scrum team, the Scrum Master is a servant leader, facilitating/supporting the needs of the team.30TheScrum Master also protects the team from distractions e.g., organizational disruptions external to the functioning of the project29. Regarding protecting the Scrum process, the Scrum Master should be an expert on Scrum and its application advising those within the project team when questions of efficiency/effectiveness of the Scrum process arise, ensuring the project team remains within the framework of Scrum.

The Development Team executes the actual work that delivers value to the customer. As work is determined by the priorities driven by the Product Owner, the Development Team autonomously organises to execute the tasks that amass to prioritised items.

### Scrum Artifacts

Cambridge Dictionary defines an “artifact” as “an object that is made by a person, such as a tool or a decoration, especially one that is of historical context”.[[28]](#footnote-28) The Scrum Alliance supports generally accepted universal definition of three Scrum Artifacts – Product Backlog, Sprint, Backlog and Product increment.[[29]](#footnote-29)

The Product Backlog is a list of all work that is to be executed in order to complete a project or project or product. That list is prioritised with the work that is deemed most important at the top of the list and those items being executed first. Work is executed in iterations called Sprints; at the end of which, a version of the product (iteration) is completed. This product is the Product Increment. The work done in a Sprint where the Product Increment is produced, which is a subset of the Product Backlog, is the Sprint Backlog.[[30]](#footnote-30)

### User Stories

A User Story is an explanation of a specific unit of work which reflects the perspective of the end user in order to articulate how that feature will add value to the customer.[[31]](#footnote-31) This is central to the principles of the Manifesto for Agile Software Development regarding customers’ needs being given priority. Technically speaking, with regards to Scrum, a User Story is a functional division of work which constitutes the Product Backlog with each story contributing to the value of the end product. The Agile Alliance suggests that good User Stories adhere to a formula following the acronym “I.N.V.E.S.T.”meaning: [[32]](#footnote-32)

* “I” ndependent (of all others)
* “N” egotiable (not a specific contract for features)
* “V” aluable (or [vertical](http://guide.agilealliance.org/guide/incremental.html) – useable successive versions)
* “E” stimable (to a good approximation)
* “S” mall (so as to fit within an iteration)
* “T” estable (in principle, even if there isn’t a test for it yet)

Project Management software tools bring additional structure to how project teams can manage their Scrum processes using user stories. Evident in the top two Agile software products by market share “Jira” and “Microsoft Azure DevOps” (mentioned above and amongst others) a suggested hierarchy of work units add layers of organisation

These tools describe the following items:[[33]](#footnote-33)

* Theme – A large focus areas spanning an organisation
* Epic – Represents a business initiative
* Feature – Represents a shippable component of software

User Stories, which would roll up to features, can then be further divided into the following sub-categories.[[34]](#footnote-34)

* Task – Used to track work at more detailed level below user stories
* Bug – Track code defects

From a managerial standpoint, when utilising project management software tools such as Jira or Microsoft Azure DevOps, the backlog is planned at Epic level and filtered down to the lower levels in the hierarchy. Features are the bridge between management and the Development Team and can be planned in both directions in the backlog – from the management to Development Team and in the reverse.

User Stories are planned at the Development Team level. This is where the actual work happens. Scrum and planning meetings can occur extensively at this level as the Development team needs to critically understand the unit of work and the priority/order that this is work requires. Tasks divide these units of work in to actionable separate parts. Bug typically result from testing the release of a product – the current form of the product at the end of or at a certain stage of an iteration Additionally, new units of work are continuously created as needed by members of the development team.

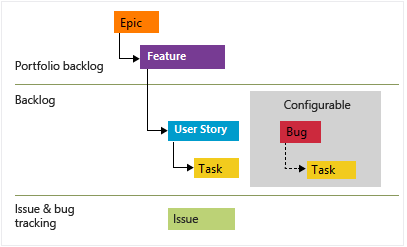


Figure 4. MS DevOps hierarchy of work items37

### Story Points

A Story Point is a numeric estimation/rating of the “size” of a User Story relative to work effort required to execute work item.[[35]](#footnote-35) In general, there are two types of estimation techniques applicable to User Stories – a Relative and an Absolute approach. The relative approach (Story Points) compares size or intensity of a baseline User Stories amongst other User Stories in the Sprint Backlog. 38 The Absolute approach estimates User Stories in traditional time-honoured units such as man-days or man-hours.[[36]](#footnote-36) Some valid arguments advocating for/against the use of both techniques used for estimation include:[[37]](#footnote-37)

Pros

* Ease vs. Difficulty – ease of relating ‘larger’ to ‘smaller’ Stories by ordinality
* Avoids Commitment – Development team held accountable to Absolute estimates
* Relativity mitigates the differences in estimation that are evident in the amount of time that is required by different individuals to do the same job.

Cons

* Arbitrary – estimation on hours is learned based on actual work and experience
* Overestimation – Velocity as a key performance indicatory (KPI) inadvertently encourages overestimation
* Non-Project-related work cannot be separately analysed**.**

**Velocity/Burndown**

Primarily, estimation of Story Points is needed for the purpose of planning future work. This is achieved by measuring Velocity or Burndown.

Scrum Inc., and organisations that “makes change a competitive advantage” founded by cocreator of Scrum, Jeff Sunderland (mentioned previously) defines velocity as a “measure of the amount of work a team can tackle during a single Sprint”.[[38]](#footnote-38)Velocity is often displayed using a Burndown chart. This is a graph which demonstrates the measure of the amount of work that has been completed in a Sprint and conversely the work remaining – representing a team’s likelihood of completing the work by the end of the Sprint.[[39]](#footnote-39)On the vertical axis is the quantity of work remaining and on the horizontal axis is the time elapsed during the Sprint[[40]](#footnote-40).

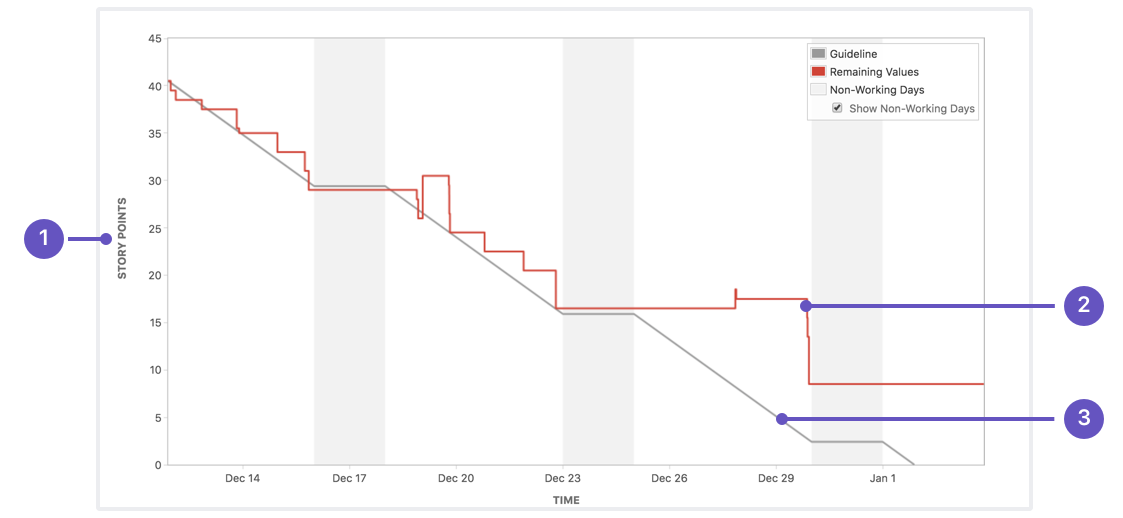


Figure 5. Example of a Burn down chart

While Scrum organisations recommend that estimation is done best with the entire development team is involved, they do not they also refrain from recommending exactly how estimates should be provided. They do, however, offer different estimation techniques/methods[[41]](#footnote-41).

### Story Point Estimation (Planning Poker)

There are several techniques that Scrum teams use for estimating User Stories. The Project Management Institute references the following examples:[[42]](#footnote-42)

1. Planning Poker
2. T-Shirt Size
3. Affinity Grouping using T-Shirt sizes
4. Determining Group Velocity
5. Forecast for Item Completion

For the purposes of this paper, the focus will be on the first and most popular technique from this list – Planning Poker.[[43]](#footnote-43)

Planning Poker stems from a forecasting method known as “Wideband Delphi” which was developed around somewhere between the year 1948 and 1960 (sources vary).[[44]](#footnote-44) Per the Project Management Institute, it was then refined by Extreme Programmer, James Grenning in 2002 in his article, “Planning Poker or How to avoid analysis paralysis while release planning”.[[45]](#footnote-45) The technique was then popularised by Mike Cohn (one of the founders of the Scrum Alliance) amongst the Scrum community in 2005 throughthe release of his book “Agile Estimation and Planning”.[[46]](#footnote-46)

Planning Poker is a gamification-type approach to estimation whereby each member of an Agile/Scrum team uses playing cards to assign Story Point values (based on an ordinal scale) to a User Story upon understanding its intent. The scale initially follows a Fibonacci-like sequence with numbers 1, 2, 3, 5, 8, and 13, then the sequence continues with 20, 40 and 100 with ascending order of value. After gaining an understanding of the User Story, some form of consensus amongst the team is obligatory in order that the value for the Story Point to be assigned (one of the criticisms of this technique). Any discrepancies are then discussed and procession of estimation of other User Stories continues upon consensus on the score given to the current User Story in question.

Experience as a university student and a professional, reveal Story Points an estimation technique in need of development. On one occasion, for example, the class was divided into teams. Each team was tasked with dividing the curriculum into User Stories – progressive units of work encompassing all the knowledge that was to be learned for the class. The Professor urged the teams very earnestly that that each team member should read all of the work material in order to ensure true learning given the fact that the material was unknown to the majority of the class. The User Stories were then to be estimated as Story Point using the Planning Poker gamification method as the project proceeded.

The task of estimating the point values was difficult and time consuming. Reviewing the material, the team was able to form ideas about which User Story would be more intensive/demanding compared to others; However, given the fact that some of the work was unfamiliar, the team chose different values which then had to be discussed until consensus was reached. Furthermore, given the arbitrary nature of the Fibonacci gradation though ordinal (lower to higher ranking) the personal perspective of each individual team member’s subjective view of the value/weight of the Story Point itself (this will be discussed further later in the paper) can make the process quite random.

Continuing with project planning, the work was divided into a set number of weekly Sprints in order to ensure completion within the frame of the semester. As the development work progressed, the team managed the Sprints by assessing a burn down rate. This once more highlighted the random nature of Story Points. While the team was aware of the total number of Story Points that needed to be covered for the project/semester and in each Sprint, in the cases where either a Sprint was completed earlier or later than its duration, this could not necessarily be translated into a direct Story Point value. If estimations and burn down were estimated in more absolute terms such as hours (or using an improved relative technique) the team would have been able to more directly relate this to how advanced or lagged the project was. An absolute measure of estimation would have been especially beneficial in cases where work took longer than the sprint duration to complete the User Stories for that Sprint.

Relating this to a professional environment, deadlines still remain a reality and management wants to be kept abreast of/have controllership over projects and timelines (especially when agreements are made and/or there is fiduciary accountability to stake/shareholders). Mid-to-lover-level managers prudently would not report outstanding project times in Story Points as this would again be an arbitrary measure unrelatable to business and incommunicable to certain stakeholders.

Pros and Cons of Planning Poker

Kiryl Baranoshnik, is a practitioner of Agile for more than five years who has worked with a variety of global clients and international corporations.[[47]](#footnote-47) Based in the Europe, he is a respected speaker at various “Agile” meetups and conferences and is one of the cofounders of the Agile and Lean community in Belarus called “peopleVprocess”. With ten years as an experienced developer and now working as an Agile coach at EPAM Systems,[[48]](#footnote-48) as well as being an Authorised Agile Trained at AgileLAB,[[49]](#footnote-49) he shared his insight about the pros and cons of Planning Poker.[[50]](#footnote-50)

*Pros*

As mentioned above, consensus is incentivised. The fact that the Development team is encouraged to engage in robust discussion. The creates and enhanced understanding amongst the team so that there is a general sensibility about the how the User Story is viewed. This provided a great team benefit.

Given the fact that in the process of Planning Poker, scores are chosen (confidentially) first by each member before it is shared with the extended team, especially when Planning Poker software is used to facilitate the scoring process, this helps to avoid cognitive bias. Members of a team can be influenced by the most senior authority on the team if he or she is involved in the scoring process. Additionally, this avoids some team members awaiting the input from other team members and help ensure more avid participation. This type of cognitive bias (mentioned above) is known as “anchoring” in study of negotiation. According to the Program on Negotiation (PON) at Harvard University[[51]](#footnote-51), “the anchoring effect is a cognitive bias that describes the common human tendency to rely too heavily on the first piece of information offered (the ‘anchor’) when making decisions”.

*Cons*

If you are a newcomer to the game of Planning Poker, you may need time to understand how to participate because the game can be quite complex. One example of this complexity abounds when there is the need to come to in instances where there isn’t a consensus of scores – on can image that due to the nature of human differences, this may occur quite frequently. When there is a conflict regarding the scores given to a User Story, team members are known to make suggestions that do no align with the intent of Planning Poker. These suggestions include. Re-drawing cards, simply choosing the bigger score and taking the average of the scores that may vary. The complexity of the game can be rather time consuming; counter to the intent of timesaving regarding estimation.

Planning Poker, being a relative estimation process, is challenging in that very regard as applicable to Story Points. The game is played by starting with a benchmark User Story as a score of “1” (one) and subsequently assigning scores to other User Stories relative to the benchmark. The nature of the Fibonacci sequence being non-ratio or non-interval (no meaningful distinction between scores as well as non-multiple/divisible) creates a situation where the team member(s) confront obstacles to simple relativity when scoring User Stories.

## Agile and Scrum Critique

While the original signatories of the Manifesto for Agile Software Development have revolutionised software development, project management and the way business in general is conducted in many countries across the globe, some have been rather outspoken in their critique of the way “Agile” is done today. Those (now) antagonists advocate that the way “Agile” is implemented is misaligned with what it was originally intended to represent.

One prominent example of this is Dave Thomas’s presentation at the GOTO conference in 2015 entitled “Agile is dead”.8 In this presentation, Thomas argues that agile was intended to be an adjective – a description of something, versus “Agile” a proper noun (starting with a capital letter). He asserts that, executing programming work agilely is not saleable and posits this a highly suggested reason that “Agile” has taken a life of its own – as a noun, a proper noun. Thus, the industry that has evolved around “Agile” needed it to be objectifiable in order that it could sell training, consultancy, books and conferences.

Thomas also argues that the newness of “Agile” at the time was a draw for it. With new terminology, roles and measure of performance (Story Point velocity) companies feared being left behind of the new way of doing business and as result, the Agile industry became self-sustaining due to the need for specialists with knowledge. Therefore, the trainings, consultancy and others mentioned above are needed to maintain a competitive advantage.

His belief is that agility needs to be reclaimed for its true values as was intent of the original manifesto and suggests a version of it represented be three repeatable steps:

1. Understand your current position/situation
2. Take a small step in the direction of your goal.
3. Evaluate what happened.

And when faced with options along the project’s process, chose that which gives the ability to make future changes easier.

He strongly advocates that sources (referring to the Agile industry) which broadly instruct how to write software or conduct a project are implausible due to the nature of context – his closing mantra being “Agile is not what you do. Agility is how you do it.”

Touching on the topic of Scrum, Thomas thinks the methodology is a “potent” way to get a team focused on and organised around software development. However, he disagrees with the absolute nature of teachings from Scrum sources – positing a specific way to proceed with a project, not teaching flexibility (agility).

Ron Jeffries, another original signatory of the Manifesto for Agile Software Development antagonises over the industry that “Agile” has become in his article entitled “Developers Should Abandon Agile”.[[52]](#footnote-52) Referring to “Agile” coaches, trainers and competing frameworks, Jeffries acknowledges that he supports organisations that make attempts to improve regardless of whether “Agile” ideas are poorly implemented. He highlights, however, the detriment that poorly applied “Agile” translates into for developers – higher pressure and more demand to produce in less time – resulting in more defective work.

Jeffries believes that developers should refrain from any specifically named “Agile” methodology and instead practice software development in ways that work within these frameworks – adhering to the principles that were alluded to the in the original manifesto.

He expounds, calling for developers, regardless of what managers may deem to be the applicable “Agile” method to learn to work using the following:

* Producing running, tested, working and integrated software weekly/biweekly; building skills until a new fully functional operational version can be created multiple times a day.
* Keeping software design clean; resisting the tendency of complexity as it grows.
* Using current iterations of software that’s ready to go for all conversations with leadership/management.

Summarising his view, Jeffries thinks that software developers should avoid specific types of “Agile” (referring to the methodologies girding it – Scrum Kanban) but instead turn to the principles of the Manifesto for Agile Software Development as the leading guide. Martin Fowler, another of the signatories, mentions three challenges stemming from “faux-agile” which is “agile” that is practiced but which does not adhere to the founding principles and values.[[53]](#footnote-53) These challenges are:

1. The “Agile Industrial Complex”
2. Raising the importance of technical excellence
3. Organisation around products and not projects

Fowler refers to the Agile industrial Complex as consultants of agile who push for a way that it is or should be done (ironically including himself to some degree in this category). He states that this is a travesty and claims that there is no “one-size-fits-all” in software development. His belief (per the principles of agility) is that the (development) team should choose how it works and “Agile” should not be forced upon them.

Too many project managers, not enough software developers – alludes to the second problem Fowler believes that decisions are being influenced by business analysts/experts and project managers and there is lack of recognition of the need for developers influencing the way work is conducted. He highlights the need for nurturing the skills of developers to easily cope with rapid changes.

Strong advocacy for getting rid of projects and being more product oriented is the final major point Fowler touts. He purports the idea of long-lasting business capabilities which results in a combined effort of both technical and business personnel. Crucial to this concept is the need for teams to have a customer-centric focus and with that focus, the teams should help the customer enhance what they do.

## Summary of Theoretical Findings

Agile, which began as a qualitative approach towards software development, has evolved into its own nomenclature. Its principles have been and continue to be a revolutionising force on a global scale, the methodologies of which have infiltrated and become an integral part of many business across several industries. Agile has become the new standard for software development and other business process and has mostly replaced traditional methods primarily because of an increasing need for flexibility along the product development cycle as well as an intermittent stream of feedback and reassessment of processes.

Agile has become a critical “skill” or “ability” in the world of software development and general business. As a result, there has been a lot of organisational development around it. Many associations have emerged as special interest groups which advocated for Agile as a principle or Scrum as an applied methodology (most prominently). Some of these organisations have been founded by the original founders of the Manifesto for Agile Software Development. From these associations an Agile industry had emerged with its own system of education, certification and governance from which many specialist project management jobs are supplied globally.

The Scrum process has predefined roles – Scrum Master, Product Owner and the Development Team. Considering the Scrum Master as a leader in respect of to the Development time which executes the deliverable work, the demand for leadership certifications in Agile and Scrum is evident. Amongst the top fifteen most demanded certifications in the year 2020 are the Scrum Master and the Project Management Professional (Agile) certifications[[54]](#footnote-54).

Scrum tools (known as artifacts) and techniques are so widely used that a technology has been developed to supplement their use. There are various applications that companies have integrated into their projects which have become quite ubiquitous with Agile and Scrum – the likes of one (Atlassian) is a primary revenue driver for its parent company[[55]](#footnote-55).

Another change from traditional methods is the way work is estimated. Although not an official Scrum artifact, User Stories have become a preferred method of relative estimation compared to absolute methods of estimation using hours. Furthermore, Planning Poker, a gamification technique has become a popular method of relative estimation. Following Agile, technology has also been developed to facilitate use of this technique.

There is strong critique of what Agile and Scrum have evolved into directly from some of the developers of these concepts, methodologies and principles themselves. This critique is girded around Agile and Scrum’s detraction from their original intent. The industrialisation of these processes, instructing rigid ways of executing their principles and the commercialism that formalised the authority around knowledge and certification is the basis of some of more outspoken founders’ main qualm.

## Professional Experience and Practical Application

Planning Poker, which stemmed from a forecasting method know as Wideband Delphi, was developed, has been refined and finally was popularised over approximately 50 years. Although it is not an official scrum artefact or officially endorsed/recommended by the founders of scrum, it has become the most Popularly used estimation technique for Story Points. The author relates this with Dave Thomas’ (one of the founders of Scrum) overarching view that this technique became most popular because of its newness and draw - not necessarily because it was proven truly effective as there are critical.

Planning Poker requires consensus in assigning Story Points. This rarely ever happens. From experience, the resulting group think usually results in the overcoming effect of the most senior or respected person’s choice on the assignment of a Story Point. Additionally, the length of discussion required at time to arrive at a consensus can make the process feel tedious for developers and typically, one person simply concedes for the sake of moving along rather than engaging in robust discussion. Along with the stipulation of consensus in Planning Poker there are situations that arise, for example, where prevailing scores are assigned and the average of those two scores is taken. In the aforementioned classroom example, teams were actually encouraged by the Scrum Professional, who was the auxiliary administrator of the course, to take the average of differing chosen Story Point values in some instances when consensus could not be reached. These are a couple of the general real-life practices of Planning Poker (the incorrectness of averaging two contending scores for Story Points will be discussed later).

Story Points are defined to be in measure of effort required to complete a unit of work - that unit of work being defined as a User Story. When planning a project, initially all user stories are defined as best as possible. When assigning Story Points, the first User Story is chosen and given a score. Subsequently, another User Story is chosen and assigned a Story Point relative to the size and effort estimated for the previous User Story.

Story Points are considered to be relative estimates. When one considers the (general and customary) process as described above there is a critical element of this process that is missing. If Story Points are relative measures then what are they relative to? What is the base story? What is missing is a pooling of all user stories to assess which User Story is that base relative User Story to be assigned the base Story Point value from which all others are estimated. Simply starting the assignment of Story Points with the first chosen User Story is remiss. Is that User Story assigned the base Story Point? What Story Point value does that User Story get assigned to it?  If the first User Story is assigned a Story Point value of one (or the lowest valued Story Point) what occurs when a subsequent User Story is determined to be or require lesser effort?

Planning Poker and relative estimation is portrayed to be is simple and fast method of estimation and as a result, as mentioned above, the crucial step of pooling all user stories to determine a base has not been a part of the practical application of this game this gamification technique. As well, other practical scenarios discussed have develop as mal practices. As a result of this, antagonists to the use of Planning Poker (and Story Points) view this practice as arbitrary.

# Empiricism

## Outline and Procedure

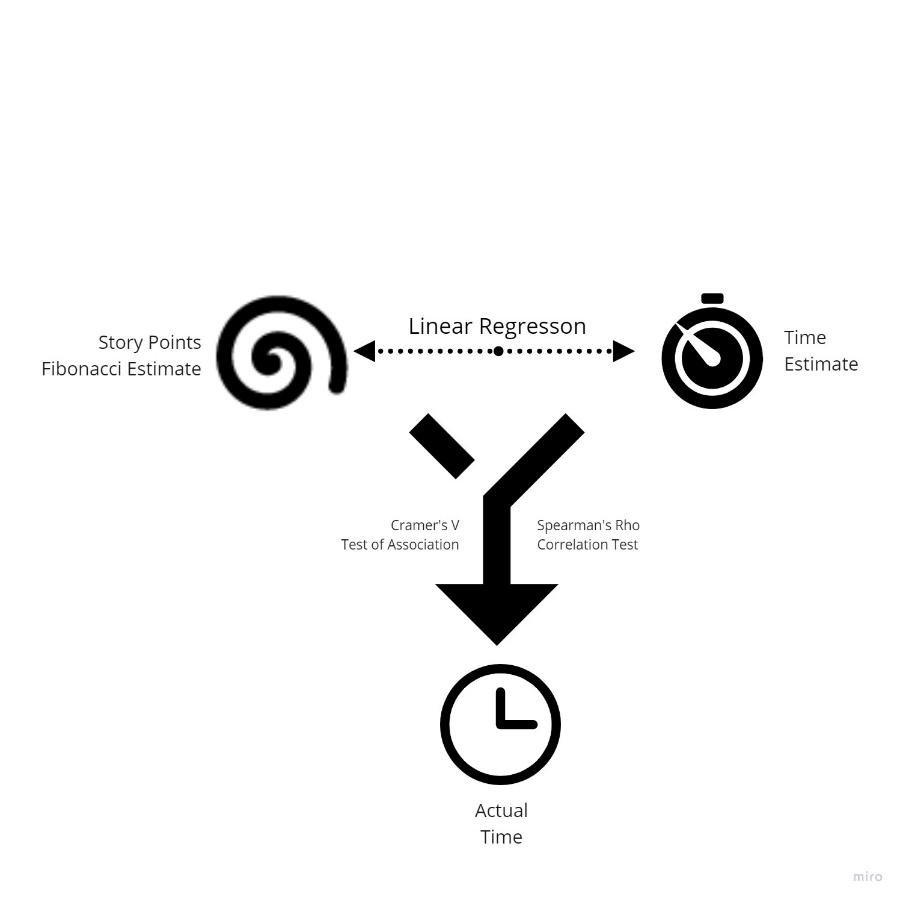


Figure 6. Demonstration of Case Study procedure

The aim of this thesis is to assess whether the relative estimation technique of using Story Points is a feasible/accurate measure of estimation.

Given the data that is available, the most viable approach towards arriving at an outcome is comparing each measure (relative versus absolute) to the only variable that can determine an accurate result – the actual (average) time spent on each work unit.

Additionally, given the scepticism around the actual value and ordinal nature of a Story Point, performing a statistical check to see whether a higher-valued Story Point is legitimately valued higher than that of a lower-valued Story Point (and vice versa) will be an important value addition to this study.

Furthermore, understanding how one estimation method relates to the other is imperative for providing value to those who prefer one method over the other. Bridging this gap can be benefit for organisation where there is a split in favour.

The following are the primary approaches that will be pursued to achieve this aim as well as to add value to the reader:

1. Analysis of Association (Correlation) – Estimates and Actual hours
2. Analysis of Association (Cramer’s V) – Story Points and Actual hours
3. Pairwise Comparison – Ordinal value of Story Point
4. Linear Regression – Modelled value of a Story Point in units of time

Correlation Analysis

Conducting an analysis of association between the Estimated hours and the Actual hours gives the strength of the relationship between these two variables. Obtaining a proportionate statistic value of this will allow a more direct comparison to the Story Point relative estimation association with Actual hours. There are various methods of executing this test of association. The choice of the method used will be discussed and executed in the Feature Generation portion of the case study below.

Analysis of Association (Cramer’s V)

Fulfilling the same purpose as above, gaining a statistical proportionate value that allows comparison of the strength of association is critical to this analysis. As a result of the nature of the variables involved, a different approach, Cramer’s V, will be used (also explained in the Feature Generation section of the case study).

Pairwise Comparison – Wilcox Rank Sum Test & Kruskal-Wallis Test

Initially, the Kruskal-Wallis test will be used to determine if, statistically, a difference exists in the levels of Story Point (categorical variable) as they relate to the Actual hours (continuous variable).

Subsequently, a pairwise comparison of the Wilcox Rank Sum Test will be used to assess if there are any significant differences (similarities) amongst the Story Point levels. This will be beneficial for finding potential groupings within the levels.

Furthermore, testing the nature of the assumption that each successive Story Point is greater than the next will be conducted via a modified pairwise comparison using the Wilcox Rank Sum Test. Essentially, each Story Point value will be check against its ordinal predecessor upon the bases of hours to reveal if it is statistically significant to reject the underlying assumption.

Linear Regression

A linear regression model will be used to translate the value of a Story Points to the number of hours that is required to execute that unit of work. This will be beneficial to both protagonists and antagonists of either (preferred) method and will foster better understanding of what one means in relation to the other.

## Case Study and Findings

### Case Study

#### Introduction

To perform this study, real project data at a consulting firm will provide the raw primary data. The data of the company/companies involved and the information provided will be anonymised removing any employee and client as well as project-specific and business process identifying information.

Five months’ worth of recorded project estimated and actual time spent data using project management software (Microsoft Azure DevOps Server) will be the empirical basis for the data used for this study. More specifically, this includes User Stories and sub work-items to those User Stories, those being Tasks and Bugs. The data primarily includes

1. Estimated Story Points assigned to the User Stories
2. Estimated Hours to execute Tasks and fix Bugs;
3. A time log of actual time spent on each Task and Bug.

Actual time provides the opportunity to assess the accuracy of both estimates.

A variation of the Cross-Industry Standard Process (CRISP) methodology (Huber et al., 2019, p. 403-8) will be applied as the layout this study and will follow the following format:

* Project understanding – define the problem and project objective
* Data mining – describing the data source and collection methods
* Data Understanding & Cleansing – understanding data and treating with discrepancies
* Data Exploration – insights that can be obtained from the data
* Feature Generation – creating new fields/variables
* Modelling – Numerical relationship between dependent and independent variables
* Data Visualisation – demonstrating key findings

#### Project Design & Understanding

Company A , in search of additional software development, project management/development resources, contracted Company B to bring its revolutionary Software as a Service (SaaS) project to fruition. The project timeline was initially slated for a seven-month launch period but the launch was delayed by a month a half. A customised mixture of Agile and Scrum was the methodology used to approach the project. Sprints were scheduled biweekly and time estimates at the Task and Bug level were conducted by the Development Team. User Stories were estimated and tracked using Story Points via Planning Poker.

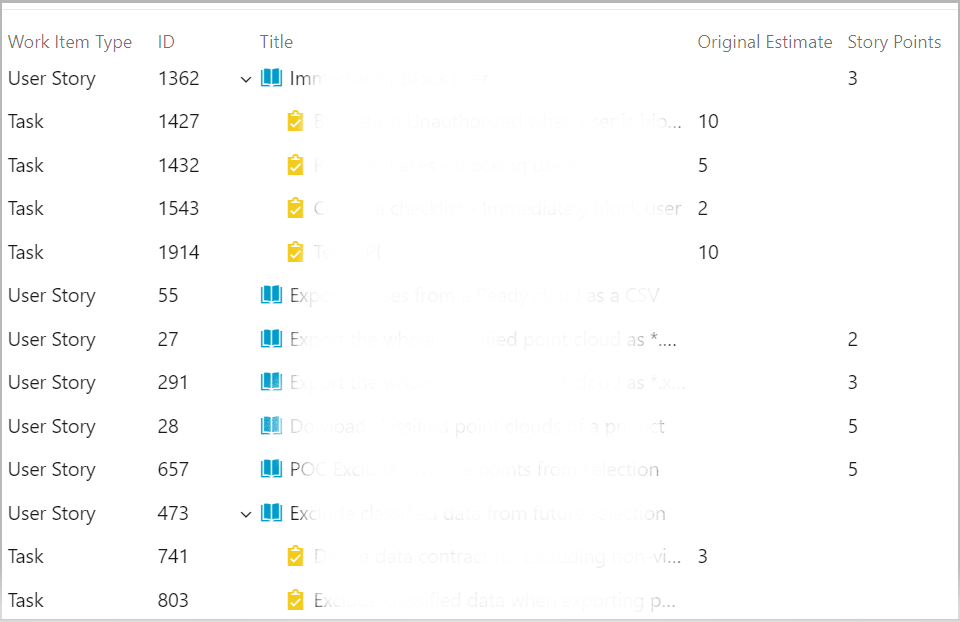


Figure 7. Example of Product Backlog in Microsoft Azure DevOps

As a result of the notification of the delay of the project, management at Company A sought to have tighter controls over and better measures related to estimation and tracking of time, and deemed that an absolute time estimation method would be more feasible.

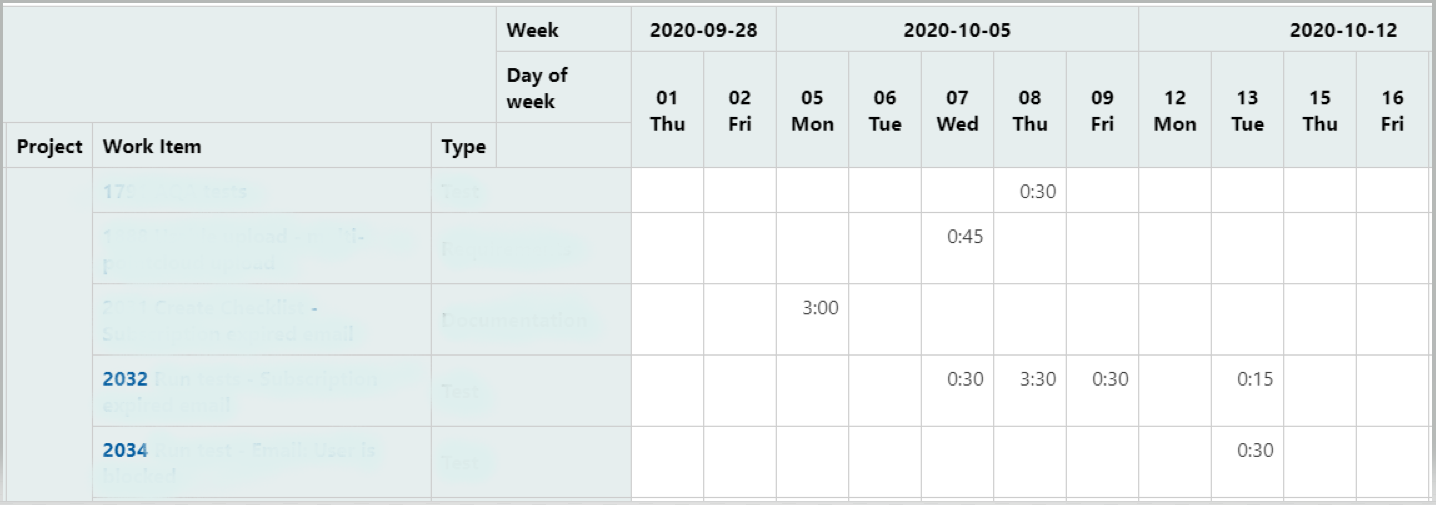


Figure 8. Example of time logged Actual time spent

While both hourly (absolute) and Story Point (relative) methods of estimation were used going forward, hourly estimation became the prominent focal point to be adhered to. Given the method of estimation being identified as a factor under contention, the aim of this project is to assess which method has proven to be more accurate and more feasible.

#### Data Mining

Data for the project was recorded and maintained in and obtained from Microsoft Azure DevOps project management software tool. From inception to launch and through continued development, User Stories were created as units of work and sub-tasked then developed to complete those units, with Bugs as necessary fixes. All Story Point and hourly estimation are recorded and stored in this tool. Additionally, all actual time spent is recorded in this tool at the Task and Bug level. Data is obtained in comma-separated value format via query directly from Azure DevOps and formatted and examined accordingly.

#### Data Understanding & Cleansing

Many fields are available to be queried into a single report in tabular format. For the purposes of anonymity, titles and descriptions of work items i.e., User Stories, Tasks and Bugs, as well as development team names to whom item are assigned to, will be excluded.

Each work item is given a unique identifying number and will be the sole modus of differentiation of work items. The field called “Parent” is used to identify the Tasks and Bugs which are subsets of their respective User Stories.

There are some User Stories that had not been assigned a Story Point value. These were excluded from the data. Additionally, there were some Tasks and that were not given an hour estimate which were also excluded – More specifically, User Stories that were given a Story Point value but for which a Task or Bug estimate was not deemed, were excluded. In summary, the entire subset of a User Story including the Story Point for the User Story must be estimated in order for that work item to be included. The details of the data cleansing are as follows:

After anonymising and cleaning the data and selecting the most relevant five months’ worth of data, the total number of unique User Stories was 202. Of those unique User Stories, 57 were not assigned a Story Point value and were excluded from the data. Of the remaining 145 unique User Stories there were, 28 for which actual worked time data was not entered. The remaining 117 had all completed fields (no estimated fields were missing).

#### Data Exploration & Insight

The cleansed data contains five fields the characteristics of which are summarised below by the following R script excerpt.

|  |
| --- |
| > head(data)  UserStoryID StoryPoint Estimate Actual WIT  1 236 7 5.00 1.00 User Story  2 256 5 6.80 3.54 User Story  3 259 5 8.67 3.50 User Story  4 285 9 7.27 8.21 User Story  5 356 1 3.00 3.13 User Story  6 477 4 5.67 5.57 User Story  > str(data)  'data.frame': 117 obs. of 5 variables:  $ UserStoryID: int 236 256 259 285 356 477 486 1029 1030 1055 ...  $ StoryPoint : Factor w/ 12 levels "0.1","0.25","0.5",..: 10 9 9 12 5 8 5 6 6 9 ...  $ Estimate : num 5 6.8 8.67 7.27 3 5.67 3.94 3.5 3.3 9.2 ...  $ Actual : num 1 3.54 3.5 8.21 3.13 5.57 2.52 2.1 1.83 6.63 ...  $ WIT : chr "User Story" "User Story" "User Story" "User Story" ... |
| |  | | --- | |  | |

Figure 9. Script showing summary of Case Study data

The excerpt reveals the “Estimate” and “Actual” fields being the only numeric variables which are the average of the sub work-items rolled up to the User Story level. The User Story ID, though it may seem to be a numeric field, is not truly numeric and was changed to a factor – essentially a categorical variable.

Normality Tests

Additionally, both numeric variables were tested for normality both visually and hypothetically. Both do not follow a normal distribution.

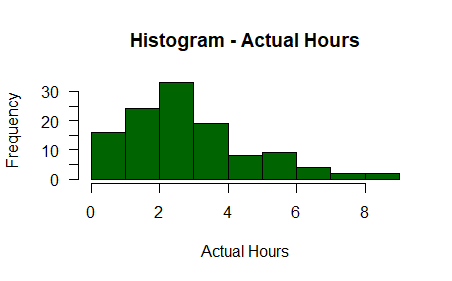
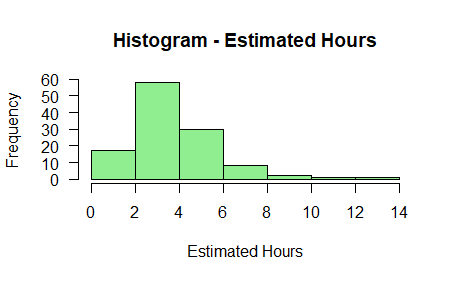


Figure 10. a & b) Showing histograms of Estimate and Actual Hours

The histogram shows that the Estimated Hours and Actual hours are both right tail skewed and asymmetrical as (symmetry about the mean) is expected of a normal distribution.

Normality was also assessed visually using Quantile-Quantile plots. These can be seen in diagrams below.

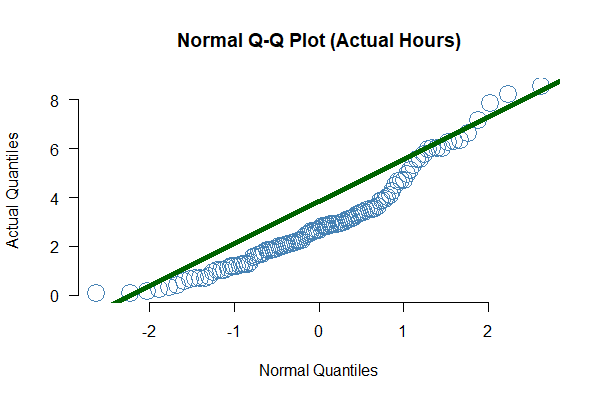
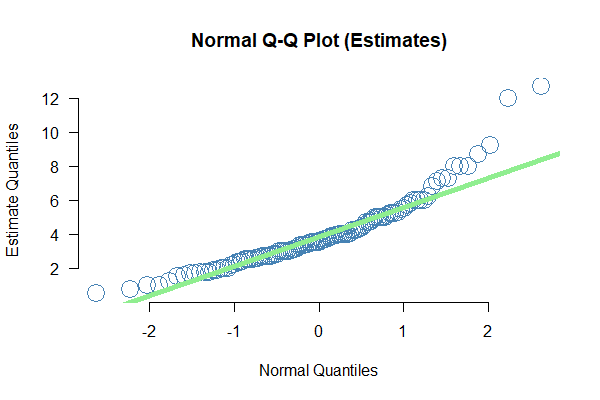


Figure 11. a & b) Showing Q-Q Plots of Estimate and Actual Hours

The Q-Q plots reveal both the quantiles of the estimated hours and actual hours are not equal or close to their respective quantiles of their respective theoretical normal distributions – the plotted quantile point deviate significantly from the line of normality. Hence, visually, one may state with some certainty that these variables are not normally distributed.

Furthermore, normality was tested hypothetically using the Shapiro-Wilk test of normality. The null hypothesis (H0) of the Shapiro-Wilk test of normality is the following:

*H0: The variable is normally distributed*

The alternative hypothesis (H1) of the Shapiro-Wilk test of normality is the following:

*H1: The variable is not normally distributed*

The results of the tests of normality for the numeric fields – Actual Hours and Estimated Hours. can be seen in the Figure 12. below.

> shapiro.test(data$Estimate)

Shapiro-Wilk normality test

data: data$Estimate

W = 0.89539, p-value = 1.538e-07

> shapiro.test(data$Actual)

Shapiro-Wilk normality test

data: data$Actual

W = 0.93762, p-value = 3.714e-05

Figure 12. Script extract showing results of Shapiro-Wilk normality test

At an alpha or significance level of 0.05, the probabilities of getting a sample statistic or a more extreme sample statistic in the direction of the alternative hypothesis when the null hypothesis is true (Bluman, A., 2006, p. 379) of both variables are significantly lower than the alpha. In other words, the likely hood of both variables following a normal distribution is less than 5%. As a result, one may safely reject the null hypothesis that the variables are normally distributed in favour of the alternative hypothesis that the variables are not normally distributed.

Pairwise Comparison

Testing the ordinality of the Story Points would provide valuable insight into the degree to which each translates to actual time spent. This was examined using different types of analyses of association. The type of analysis that can be conducted is determined by the nature of the variables under consideration. Story Points being ordinal and Actual hours being numeric; this narrows the scope of analysis options. Furthermore, it is conclusive that both numerical variables (“Estimate” and “Actual”) do not follow a normal distribution (non-parametric). As such, a Kruskal-Wallis test was used to examine the dependencies between the Actual Hours and the Story Point variables per level – an Eta ANOVA test could not be used as assumptions of normality were not met (Liu, H., 2015, p. ii). A summary of the results of that analysis is show below.

> kruskal.test(x = data$Actual, g = data$StoryPoint, Actual ~ StoryPoint, data = data)

Kruskal-Wallis rank sum test

data: data$Actual and data$StoryPoint

Kruskal-Wallis chi-squared = 36.527, df = 11, p-value = 0.0001383

Figure 13. Script extract showing result of Kruskal-Wallis test

Interpreting the results, the null hypothesis (H0) shown in Figure 13. above of the Kruskal-Wallis H Test is the following:

*H0: The mean ranks of the levels are equal*

The null hypothesis (H1) of the Kruskal-Wallis H Test is the following:

*H1: The mean ranks of the levels are not equal*

Given an alpha or significant level of 0.05 and the determined probability-value (p-value) of 0.0001383 (markedly less that the alpha) the null hypothesis can convincingly be rejected for the alternative that the mean ranks are not equal – Essentially the Story Point levels are not representatively the same with respect to the Actual hours worked.

In lieu of the Kruskal-Wallis test, a two-sided Wilcox Rank Sum test will reveal specifically which Story Point levels have statistically significant distributions (if there are any) – essentially determining whether there may be any potential grouping of Story Point levels.

> pairwise.wilcox.test(data$Actual, data$StoryPoint,

+ p.adjust.method = "BH", alternative = "two.sided")

Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data: data$Actual and data$StoryPoint

0.1 0.25 0.5 0.75 1 2 3 4 5 7 8

0.25 0.7943 - - - - - - - - - -

0.5 0.2875 0.4241 - - - - - - - - -

0.75 0.8627 1.0000 0.7943 - - - - - - - -

1 0.3288 0.2635 0.1033 0.5593 - - - - - - -

2 0.3690 0.2880 0.0471 0.6857 0.3690 - - - - - -

3 0.1612 0.0484 0.0367 0.3905 0.3718 0.8627 - - - - -

4 0.9600 0.8885 0.7182 1.0000 0.6959 0.6857 0.5302 - - - -

5 0.1571 0.0471 0.0021 0.3866 0.0364 0.3690 0.6154 0.3866 - - -

7 0.3866 0.6947 0.8392 1.0000 0.9766 0.9766 0.7182 0.6947 0.6154 - -

8 0.8627 0.7500 0.3718 1.0000 0.4187 0.9057 1.0000 1.0000 0.9766 0.9600 -

9 0.8627 0.7500 0.3718 1.0000 0.3690 0.3690 0.6154 0.7500 0.3866 0.6947 1.0

P value adjustment method: BH

Figure 14. Script showing results of pairwise Wilcox Rank Sum (two-sided) Test

Interpreting the results, the null hypothesis (H0) of the Wilcox Rank Sum (two-sided) Test is the following:

*H0: The distributions of the levels are equal*

The alternate hypothesis (H1) of the Wilcox Rank Sum Test is the following:

*H1: The distributions of the levels are not equal*

Given an alpha or significant level of 0.05, with respect to the p-values highlighted in red in Figure 14. above, the null hypothesis can be rejected for the alternative that the distributions of the Story Point level are not equal – the Story Point levels are not representatively the same with respect to the Actual hours worked. That is, considering, for example, the comparison pairing of StoryPoint\_0.5 and StoryPoint\_5: these two levels are distinctly different as they pertain to the number of Actual hours worked. However, the distributions for StoryPoint\_2 and StoryPoint\_3 are statistically not different. These results can mostly be seen in the categorical boxplot below.

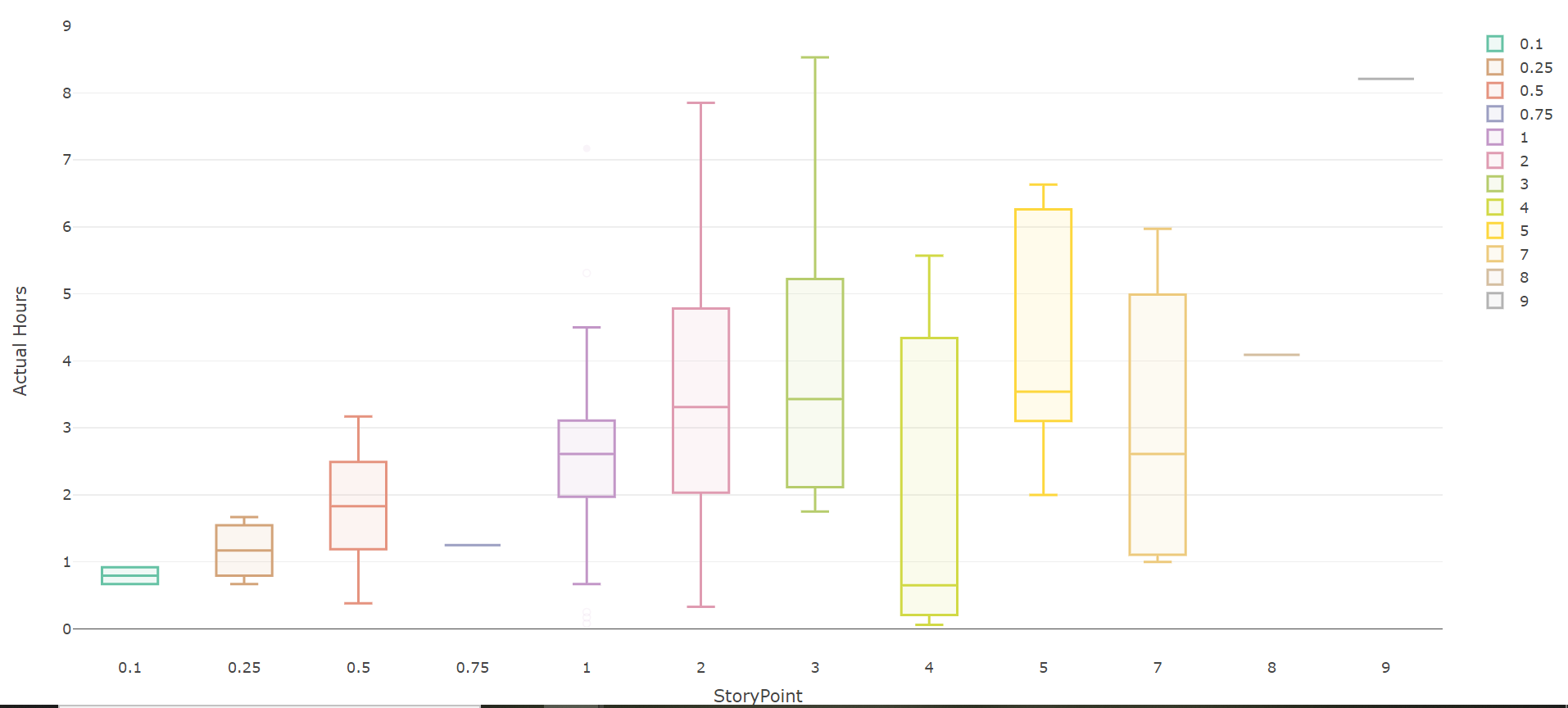


Figure 15. Story Point Categorical Box Plot

Given the non-numeric yet ordinal nature of a Story Point in this regard, an insightful procedure would be testing how the value of each Story Point compares to the others. More specifically, in testing both the difference and the ordinal nature, the successive levels of Story Points were examined to see if there are any direct differences between each level of Story Point, and whether each succeeding level is greater than the next (as it should). This was performed using one-sided a (greater than) pairwise comparison test.

> pairwise.wilcox.test(data$Actual, data$StoryPoint,

+ p.adjust.method = "BH", alternative = "greater")

Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data: data$Actual and data$StoryPoint

0.1 0.25 0.5 0.75 1 2 3 4 5 7 8

0.25 0.468 - - - - - - - - - -

0.5 0.144 0.212 - - - - - - - - -

0.75 0.500 0.611 0.849 - - - - - - - -

1 0.164 0.132 0.052 0.289 - - - - - - -

2 0.185 0.144 0.024 0.375 0.185 - - - - - -

3 0.081 0.024 0.018 0.195 0.186 0.500 - - - - -

4 0.849 0.849 0.849 0.849 0.849 0.871 0.910 - - - -

5 0.079 0.024 0.001 0.193 0.018 0.185 0.328 0.193 - - -

7 0.193 0.377 0.493 0.707 0.589 0.707 0.849 0.377 0.891 - -

8 0.500 0.434 0.186 0.611 0.209 0.533 0.611 0.611 0.589 0.574 -

9 0.500 0.434 0.186 0.611 0.185 0.185 0.328 0.434 0.193 0.377 0.611

P value adjustment method: BH

Figure 16. Script showing results of pairwise Wilcox Rank Sum (greater than) Test

Interpreting the results, the null hypothesis (H0) of the Wilcox Rank Sum (one-sided greater than) Test is the following:

*H0: The distribution of the row level is greater than that of the column level*

The alternate hypothesis (H1) of the Wilcox Rank Sum (one-sided greater than) Test is:

*H1: The distribution of the row level is not greater than that of the column level*

Given an alpha or significant level of 0.05, with respect to the p-values highlighted in red in the Figure 16. above, the null hypothesis can be rejected for the alternative that the distributions of the Story Point level corresponding to the row is not greater than the distribution of the Story Point level corresponding to the column. In other words, the Story Point of the row is not statistically greater than the Story Point of the corresponding column with respect to the Actual hours worked. That is, considering, for example, the comparison pairing, StoryPoint\_3 is not statistically greater than StoryPoint\_0.5 pertaining to the number of Actual hours worked. Conversely, however, the distributions of StoryPoint\_5 is statistically greater than that of StoryPoint\_3. While there may be other factors that contribute to the seeming lack of successive disorder of the Story Point variable (this will be elaborated upon further in the “Discussion” section of this paper) there is statistical evidence which suggests that Story Point are not ordinal (with respect to Actual hours spent on a unit of work).

#### Feature Generation

Analysis of Association

In order to determine the feasibility of using Story Points (in this case) direct comparisons of Estimates Hours to Actual Hours and Story Point to Actual Hours was conducted to see which is more closely associated. While Estimate and Actual data are continuous variables, the Story Point field is ordinal. In order to be able to determine the degree of association through a correlation coefficient, the Actual (numeric) column was binned to become a categorical variable in order that a Cramer’s V analysis of association could be conducted.

Varying binning groupings were made in order to attain the best accuracy. Binning was done on the following bases:

* Five equal lengths: 0-2], (2-4][[56]](#footnote-56), (4-6], (6-8], > 8
* Nine equal lengths: 0-1], (1-2] through (6-7], (7-8], > 8
* Seventeen equal lengths: 0-0.5], (0.5-1.0] through (7.0-7.5], (7.5-8.0], > 8.0

Cramer’s V was then used to check the association of the Story Point compared to the newly binned fields. The result is shown below:

|  |
| --- |
| > dataBins1 <- data.frame(data$ActualBins1, data$StoryPoint)  > cramersV(table(dataBins1))  [1] 0.4551752  Warning message:  In chisq.test(...) : Chi-squared approximation may be incorrect |
|  |
|  |

Figure 17. Script extract showing results of Cramer's V Test of Association

The Cramer’s V (coefficient) resulted in a 45.52% association between the Story Point and Actual hours fields. The bin corresponding to five groups of equal length returned the best result; the others closely behind.[[57]](#footnote-57)

Comparing the numeric Estimate to the Actual variables, a Spearman’s Rho/Rank test was conducted. This is the appropriate test of association given the fact that there is no assumption/requirement of normality of the numeric variables (which was established about the variables). The result of which are the following:

> ## Spearman's Rank for numeric variables

>

> cor.test(x = data$Estimate, y = data$Actual, method = c("spearman"))

Spearman's rank correlation rho

data: data$Estimate and data$Actual

S = 87412, p-value < 2.2e-16

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.6725109

Figure 18. Script extract showing results of Spearman's Rho Test of Association

The null hypothesis (H0) of the Spearman’s Rank Test shown in Figure 18. is the following:

*H0: There is no association between the two variables*

The alternative hypothesis (H1) of the Spearman’s Rank Test is the following:

*H1: There is an association between the two variables*

At an alpha/significance of 0.05, the p-value of 2.2 x 10-16, the null hypothesis can be confidently rejected for the alternative; meaning that there is an association between the variables. Furthermore, there is a resultant 67.25% strength of positive linear association between the two variables.

Comparing both correlation association percentages, while some data is lost as a result of the binning process (Lin, Hewett and Altman, 2002, p. 454) the association of Estimate to Actual hours is greater than that of Story Points to Actual hours. While this does not provide an absolute conclusion, it does support the notion that regarding estimation, using an absolute measure – estimate hours, is a more feasible choice in this specific case.

Dummy Variables

Seeking to add further value, an understanding of the translation of Story Points to Actual hours was pursued. This translation would best be demonstrated using a regression model. In order to proceed executing this, a determination would need to be made about dependent and independent variables. Since the understanding of a Story Point is that it is a measure of intensity of work, the time required to perform that work (User Story) was assumed to be the dependent variable. For this to be achieved, however, given the nature of the Story Point being non-numeric, the Story Point variable was encoded into binary output (dummy variables). Given the fact that there were twelve levels, eleven dummy variables were created – if there are “j” number of levels, “j – 1” number of dummy variables is required (Hardy, M., 1993, p. 8).

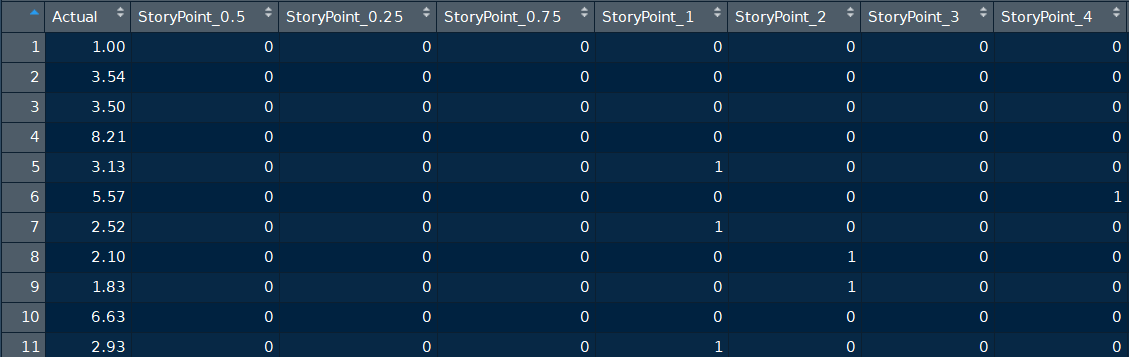


Figure 19. Extract demonstrating sample of created dummy variables

The naming convention for the dummy variables is “StoryPoint\_(value)” so that the column can easily identify to which Story Point level it corresponds. The Story Point level “0.1” (StoryPoint\_0.1 column) was not created as a result of this variable being the reference group (Alkharusi, H., 2012, p. 202). The reference group acts as the column/variable that is the baseline for which the coefficients of the (other) created variables are compared – With respect to dummy variables the interpretation of the coefficient is always in light of the reference group.

#### Modelling

The model used will be a linear regression model with the created dummy variables as the independent variables and the Actual hours as the dependent variable – the formula of which if the following:

*Actual ~ StoryPoint\_0.5 + StoryPoint\_0.75 +… + StoryPoint\_8 + StoryPoint\_9*

Before modelling can take place, however, aside from cleaning the data since which has already taken place, the follow procedures are general considerations that should be factored in to ensure the building of a nonbiased model:

* Outliers’ treatment
* Data Split - Train and Test
* Multicollinearity – Correlation of independent variables.

Outliers

Quite simply, categorical (ordinal) variables explicitly cannot have outliers as outliers pertain to numeric variables. When categorical variables are converted to dummy variables as was done for this case, therefore, there will be no consideration of outliers as each entry of each variable contain data that will be pertinent to building the regression model.

Train & Test Sets

Generally, data is split into train and test sets in the portion of 80/20 or 70/30 respectively in order to avoid overfitting. The train set is usually the larger set so that a model can be trained with a significant and representative portion of the population or sample data. The test set is kept separate and untouched in order that it remains unbiased to the model. Subsequently, after applying the model to the test set, the resulting predicted values are tested for accuracy and certain assumptions.

In this case, however, given the nature of dummy variables splitting the dataset may produce a less accurate model because splitting the data may cause exclusion of certain levels. For example, there is only one data point for Story Points 0.75, 8 and 9. If the data is split into train and test sets, the record for these Story Points with only one record will force that single record into one of the data sets.

Consider if that one record falls in the train set, the model factors that variable into the algorithm. Subsequently, applying the model to predict on the test set, there may not necessarily be any adverse effect. Given the opposite scenario, however, that one record may fall into the test set (most importantly out of the training set) and therefore, the model will be trained without influence of that record. Subsequent application of the model to the test set would have adverse effects on the predictions and accuracy of the model given the fact that the model would not have been trained to deal with the unseen variable in the test set. Hence, in this case, the data will not be split. Testing of the model’s accuracy will be discussed later in greater detail.

Multicollinearity

Another factor that can affect model performance is multicollinearity amongst independent or explanatory variables. Multicollinearity refers to redundancy as a result of high correlation between variables. Essentially, if one explanatory variable is highly correlated with another (or combination of) explanatory variable(s), then that variable is influenced by the other with respect to the corresponding coefficients of the regression function. This results in potentially unreliable coefficients as well as the tests of significance of the coefficients. In testing for the presence of multicollinearity amongst variables, the following approaches were used:

Correlation Matrix

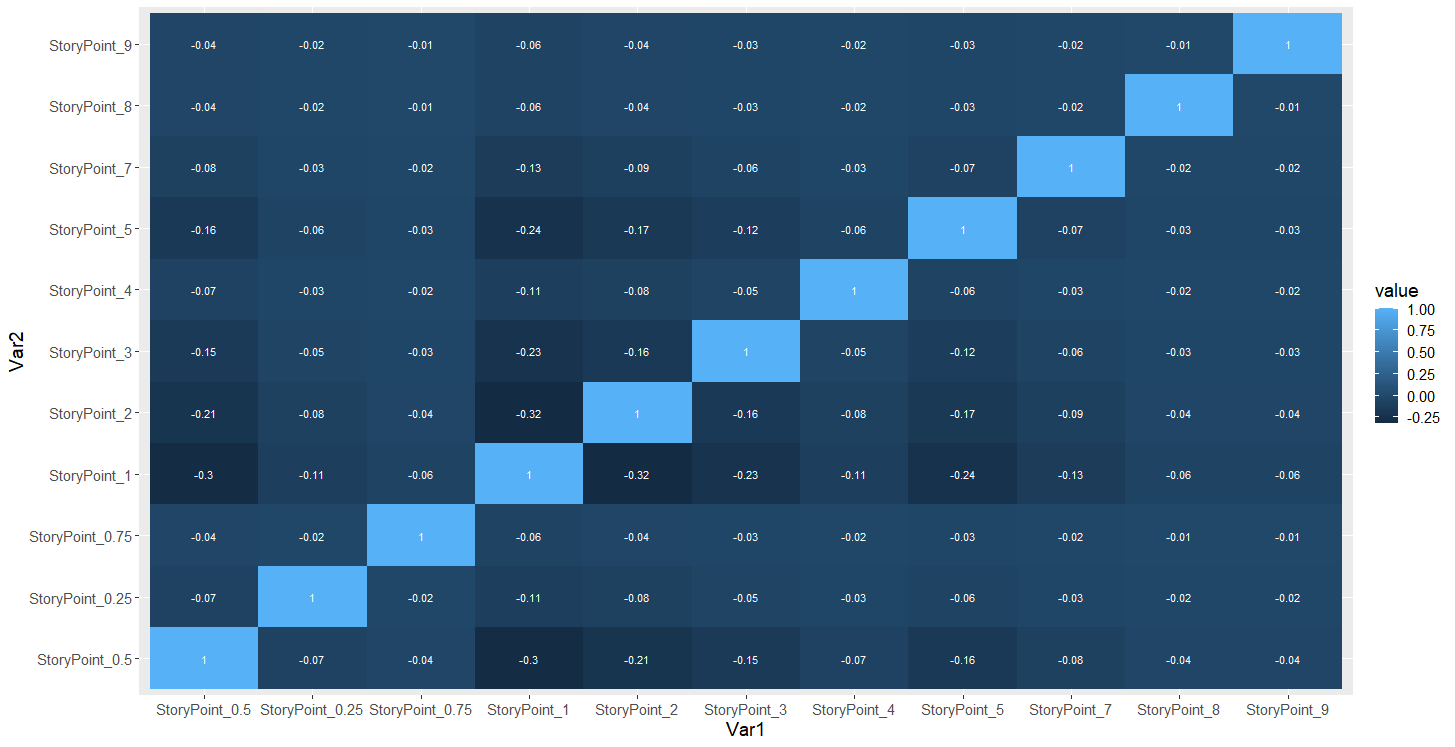


Figure 20. Correlation Matrix of Dependent Variables

From the correlation matrix of the Story Points, one will observe that all fields are not only negatively correlated but lowly negatively correlated – the highest negative correlation being -0.32.

Fortifying comforts around predictors and concerns around multicollinearity another method of ensuring that multicollinearity would not affect model performance is checking the variance inflation factor. While pairwise correlation, such as what was executed by the correlation matrix, is valuable, it is inherently limiting by its nature of pairing.[[58]](#footnote-58) Some many or all (as in this case) pairwise correlations may be small; however, correlations between variables and combination of variables and vice versa, or between combinations of variable and combinations of variables cannot be determined from pairwise correlations alone.

The variance inflation factor is a statistic which determines the degree to which the variance of an estimated coefficient in a regression function increases when explanatory variables (single or multiple) are correlated (Akinwande, et al., 2015, p. 754-67).

In multiple linear regression, the variance inflation factor measures of the amount of multicollinearity amongst regression coefficients. It is represented by a ratio of the comprehensive model variance to that of a model containing only the single respective variable. The mathematical formula for its calculation is the following in Figure 21. and Figure 22.

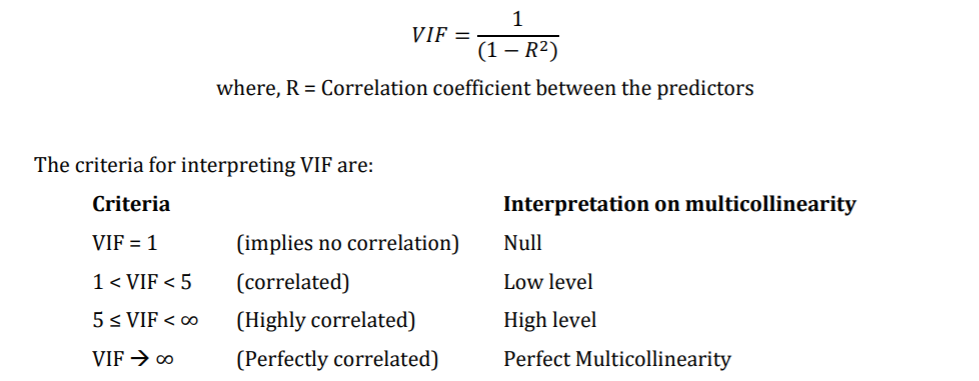


Figure 21. Variable Inflation Factor Formula

and where:

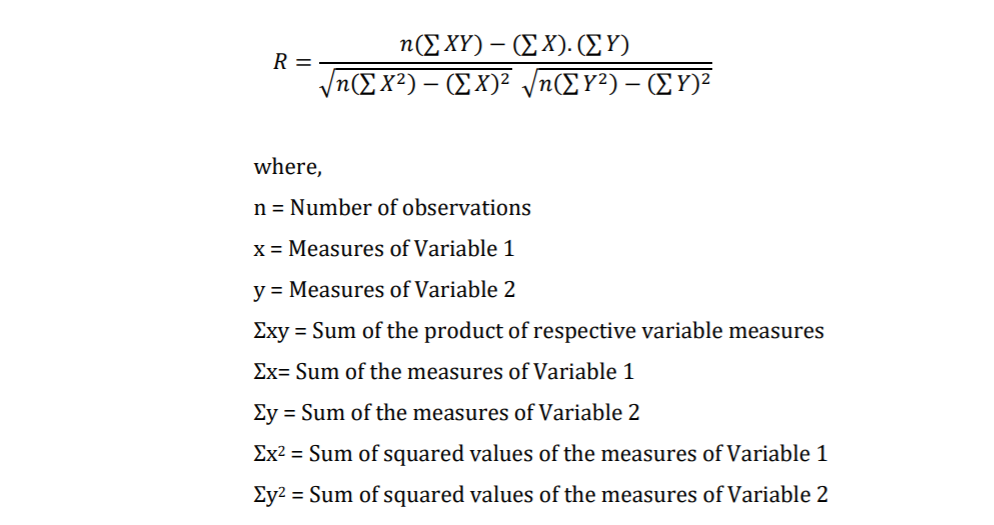


Figure 22. Variable Inflation Factor Formula (continued)

The higher a variance inflation factor, the higher the correlation of that variable with other variables. A variance inflation factor greater than 10 indicates a poorly estimated correlation coefficient due to multicollinearity.

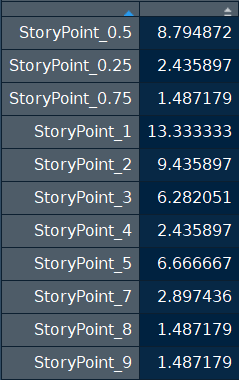


Figure 23. Variable Inflation Factor Results

Result of Variance Inflation Factor testing of Story Point independent variables:

While StoryPoint\_1 is the only variable with a factor greater than 10, other variables that should be remarked include StoryPoint\_0.5, StoryPoint\_2, StoryPoint\_3 and StoryPoint\_5. These can affect the reliability of the coefficients in a linear regression model.

While, statistically, there appears to be multicollinearity amongst the Story Point variable, for the purpose of obtaining a translation from Story Point to Actual time required, all Story Points will be included in model building for an initial iteration and subsequently assessed for rebuild in another iteration after assessing the model.

Model Building

The resulting model demonstrated Actual hours as a function of Story Points in the following liner function:

*Actualhours = 0.795 + (0.375 \* StoryPoint\_0.25) + (1.027 \* StoryPoint\_0.5) + (0.455 \* StoryPoint\_0.75) + (1.791 \* StoryPoint\_1) + (2.589 \* StoryPoint\_2) + (3.076 \* StoryPoint\_3) + (1.298 \* StoryPoint\_4) + (3.550 \* StoryPoint\_5) + (2.252 \* StoryPoint\_6) + (\* StoryPoint\_7) + (3.295 \* StoryPoint\_8) + (7.415 \* StoryPoint\_9)*

The result and summary of the linear regression model is shown on the following script extract below:

Call:

lm(formula = Actual ~ ., data = dumm.data)

Residuals:

Min 1Q Median 3Q Max

-3.0543 -0.8454 -0.0108 0.8143 4.6592

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.795 1.130 0.704 0.483202

StoryPoint\_0.5 1.027 1.188 0.865 0.389161

StoryPoint\_0.25 0.375 1.459 0.257 0.797604

StoryPoint\_0.75 0.455 1.957 0.233 0.816592

StoryPoint\_1 1.791 1.160 1.544 0.125654

StoryPoint\_2 2.589 1.182 2.190 0.030745 \*

StoryPoint\_3 3.076 1.220 2.520 0.013224 \*

StoryPoint\_4 1.298 1.459 0.890 0.375426

StoryPoint\_5 3.550 1.214 2.925 0.004215 \*\*

StoryPoint\_7 2.252 1.384 1.628 0.106553

StoryPoint\_8 3.295 1.957 1.684 0.095191 .

StoryPoint\_9 7.415 1.957 3.789 0.000252 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.598 on 105 degrees of freedom

Multiple R-squared: 0.3115, Adjusted R-squared: 0.2394

F-statistic: 4.32 on 11 and 105 DF, p-value: 2.601e-05

Figure 24. Script extract showing results of Linear Regression Model

Interpreting the results of the regression model, the null hypothesis (H0) is the following:

*H0: There is no relationship between the dependent and independent variables*

The alternative hypothesis (H1) is the following:

*H1: There is relationship between the dependent and independent variables*

With a p-value of 2.601-05 the null hypothesis is rejected for the alternative hypothesis. There is a linear relationship between the dependent and independent variables.

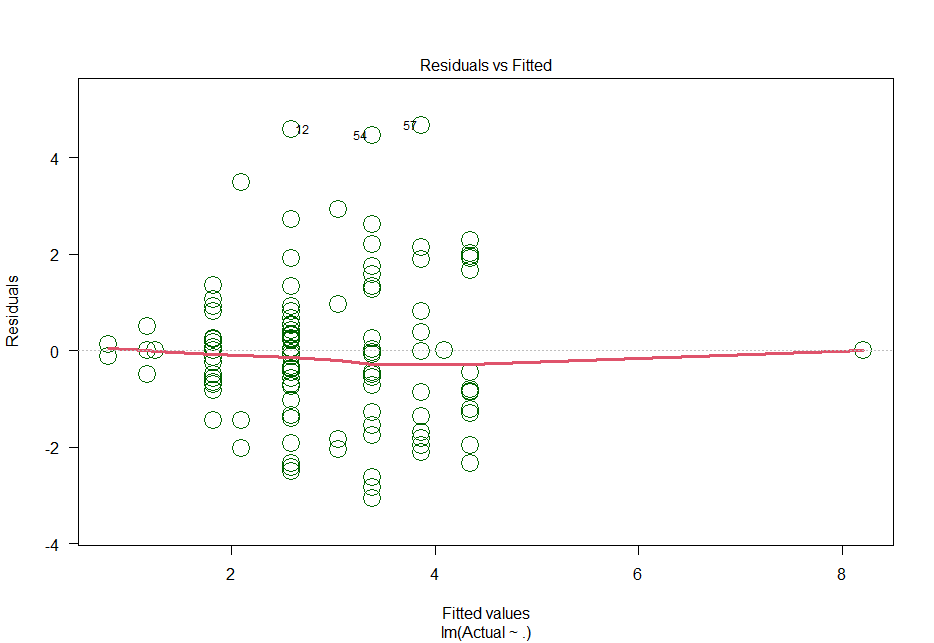


Figure 25. Graph showing Linear Model Residuals (Homoscedasticity)

With an adjusted R-squared value of 0.2394, there is a low proportion of variation around the regression line which is explained by the model. However, observing the graph of the residuals around the in Figure 25. above, the variation is homoscedastic.

Testing the normality of the residuals visually and using a Kolmogorov-Smirnov Test. The results are the following:

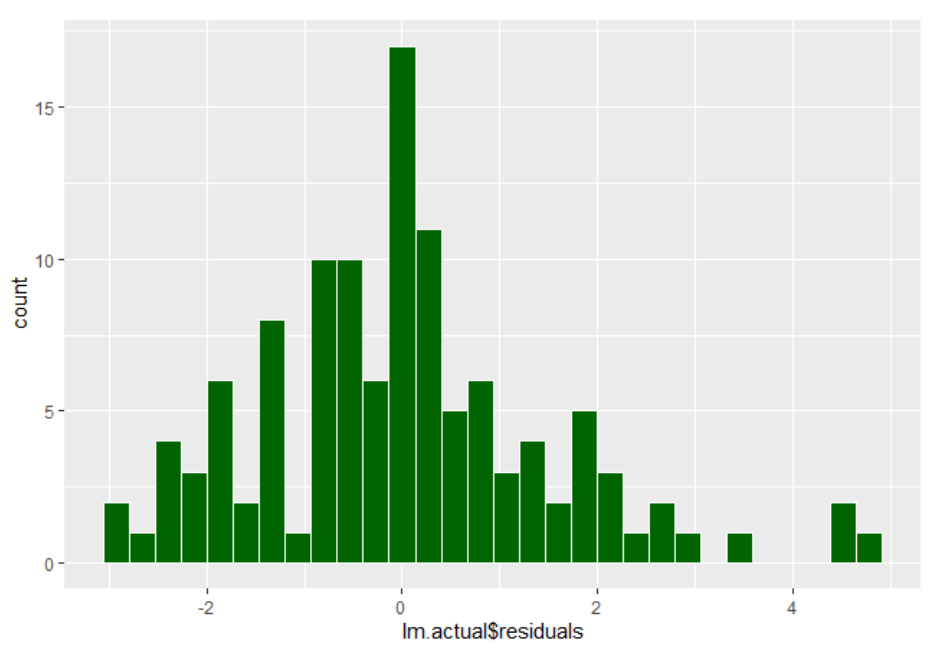


Figure 26. Histogram of Linear Model Residuals

> # test residuals for normality

> ks.test(lm.actual$residuals, pnorm)

One-sample Kolmogorov-Smirnov test

data: lm.actual$residuals

D = 0.11415, p-value = 0.09477-/

alternative hypothesis: two-sided

Figure 27. Script extract showing test and results of Kolmogorov-Smirnov Test

Interpreting the results of the Kolmogorov-Smirnov test, the null hypothesis (H0) is the following:

*H0: There is no difference between the distribution of the residuals and a normal distribution.*

The alternative hypothesis (H1) is the following:

*H1: There is a difference between the distribution of the residuals and a normal distribution.*

The residuals visually appear to be normally distributed and at a significance level of 5%, at a p-value of 0.09477, the null hypothesis should not be rejected for the alternative hypothesis, meaning that the residuals are normally distributed.

The final assumption of a linear regression function is independence of residuals. While the most effective check for this is using the Durbin-Watson test (Akter, J., 2014, p. 137-44). This test, however, is best when the sequencing of the data records is germane (for example with time series data) which it is no in this instance. Hence, in order to test independence, the correlation between the function residuals and the predicted values. Results are shown in the script extract below.

|  |
| --- |
| >  > # test for independence  >  > cor(x = lm.actual$residuals,  + y = predict(lm.actual),  + method = c("pearson"))  [1] -2.921256e-16  > |
|  |

Figure 28. Script showing results of Test of Independence (Pearson’s Correlation Test)

Given the negligible correlation at -2.921256-16, there is some statistical comfort that the residuals of the regression function are independent. The assumptions of a linear regression function have been met.

Note that there is no coefficient for “StoryPoint\_0.1” because this is the reference category. All other dummy variable categories are compared to the reference category.

The primary objective is to establish a linear relationship between the Story Point variables and the Actual hours. Despite only 23.94% (R-squared) of variation around the regression line explained, as a result of meeting all assumptions, the model is unbiased.

While the main focus of the translation is to provide great value to those who prefer one estimation method over the other to bridge gaps in misunderstanding, there are concerns around the significance of the coefficients in function. This will be discussed in further in the “Discussion” section on the paper.

## Summary of Empirical Findings

This case study was aimed at assessing the absolute and relative methods of estimation of the units of work in an “Agile” project and drawing a relationship between those methods. Through statistical methods including analysis of association and linear regression modelling, the feasibility and relationship of each method was tested.

The company’s major product launch was delayed because of a misestimation of time needed to bring the product to a ready-to-launch iteration. Since the project’s burndown/velocity (amortisation) was estimated in Story Points at that point, a more direct method of time management was sought. Therefore, there was the need for time Estimates versus estimation in Story Points.

The data was housed using Microsoft Azure DevOps software. User Stories, (measured in Story Pints) however, were still being assessed (though deemed inadequate for estimation of time to complete the project).  Tasks and bugs, which roll up to the user stories and are estimated in time which became the primary estimator and velocity metric.

After extracting and cleaning the data from thousands of User Stories in Microsoft Azure DevOps, there were 117 records (User Stories) that remained. As some User Stories were “larger” than others, meaning that they encompassed more tasks or bugs, the mean of those tasks/bugs were taken to represent average time on the User Story. This became the sample data which was used for the testing and experimentation that followed. The fields of specific interest were the User Stories, Estimated time and Actual time – an ordinal variable, and two numeric variables respectively. The numeric variables were tested for normality which determined some of the types of testing that would be used to going forward. Each of the numeric variables did not follow the model of a normal distribution (non-parametric).

A check for outliers had revealed nothing majorly significant, however, all data was considered necessary to be included for building the most parsimonious model.

Initially a Kruskal-Wallis H test was conducted to examine whether the main ranks of the levels (Story Points) were equal. Results were significant to reject the null hypothesis for the alternative hypothesis that the levels were not equal. Essentially the Story Point levels were not representatively the same with respect to the actual hours worked. Additionally, in lieu of the results of the Kruskal-Wallis test, a two-sided Wilcox rank sum test was conducted using a pairwise comparison. This revealed that the specific areas or levels where the Story Points were not equal. Furthermore, the nature of the Story Point levels being ordinal was examined using a one-sided (greater than) Wilcox rank sum test using a pairwise comparison. This also revealed some but not all Story Points that were assumed to be successively greater than its predecessors were as not assumed with respect to actual hours worked. One may conclude, statistically speaking, that the assumed ordinal nature of Story Points is not a reliable assumption.

Due to the non-normal nature of the numerous variables – Estimated and Actual hours, in assessing the association of those two variables, a Spearman’s Rho correlation test was used. Spearman’s Rho is better suited for non-numeric variables in comparison to, for example, a Pearson’s correlation. Additionally, in order to assess the association between the Story Points (ordinal/categorical) variable and the Actual hours (numerical) variable, the Actual hours were binned/grouped and a Creamers V test was performed. Results revealed that there is a greater association between the estimated and actual hours at a rate of approximately 67% versus the association between the Story Points and the binned and Actual hours at a rate of approximately 46%.

In order for a linear model to be built for the purposes of establishing a relationship between the two forms of estimation, the Story Point of variable was converted into binary dummy variables.  Subsequently, as would be customary in building a linear regression function, dependent variables were tested for multicollinearity. Although the data revealed some multicollinearity amongst the variables, as a result of the nature of dummy variables and, more specifically, the need for a model that represented a relationship between each User Story and the actual hours, no variables were excluded form model building. The linear regression function was then built and tested for accuracy in representation. The assumptions of a linear regression function, however, were not met because of non-normality of the function residuals.

The p-value of the rebuilt model was significant enough to reject the null hypothesis that there was no relationship between the dependent and independent variables and other assumptions were met around homoscedasticity, normality and independence of residuals. Addressing concerns around the significance of the model coefficient, given the fact that all predictor variables are derived from a single non-metric/qualitative ordinal variable, the aim is to attain a coefficient that can be related to the dependent (Actual time) variable. The adjusted R-squared value, however, was relatively low at approximately at approximately 24% however, while this meant that not much of the residual variation could be explained by the function, given the statistical confidence in the model meeting all linear regression assumptions, the model would be used going forward for the usefulness of translating Story Points to estimate time.

# Summary and Contribution

Agile and Scrum frameworks and methodologies (and the reiterative approaches that undergird them) have overwhelmingly replaced traditional product management methods and have proven to be very attractive to product/software development processes. The quickness and adaptability of these methods applicable to making change have brought great benefit to businesses. Story Points estimation, although not a part of any official Agile or Scrum guidance have been widely adopted and attached to practical Scrum process. Touted for its ease and quickness as a result of its relative characteristics, the affiliation is understandable. This supposed ease and quickness of estimation is precisely what is advocated for by its proponents and while, from a matter of principle, the competitive advantage of being able to execute a process in a relatively shorter space of time is desirable, there are elements of Story Point estimation that can and perhaps should be given more detailed consideration. More specifically, in the context of a development team, a deeper understanding of what this relative estimation artefact means to individual developers should be sought.

Theoretically, there is a dichotomy that exists in organisations that already (or aim to) implement Agile methodologies. In the practical world of business, deadlines exist and customer need, management demand or requirement around industry and innovation. In the Agile world project timelines are ideally prevalent over deadlines which essentially do not fall into the framework. How would a middle, senior or project manager respond to executive management when a deadline is given? It would not be wise for that manager to respond by retorting that there are no deadlines in Agile. How would that manager respond to an executive or senior when asked for the time remaining for a project given an established deadline? That manager or senior manager cannot respond with the number of Story Points that remain. Therefore, for the sake of deeper understanding amongst the development team, the wider project team and the wider organisation, an even deeper investment of time and appreciation, especially in more critical projects should be spent on estimation versus making that process as short as possible as is suggested by the Story Point estimation process.

In order for project teams to understand and determine which estimation method would be best applicable to their needs, similar to the case study, both Story Points and time estimates, as well as actual time spent initially are recommended to be recorded and subsequently assessed per the recommendations that follow.

Story Points Elements

Story Points are defined as a measure of the effort that is required to perform a unit of work but what defines this effort? Experts remark four primary elements of effort:

1. Complexity
2. Risk/Uncertainty
3. Amount of work
4. Time required to perform work

These are the elements that comprise a Story Point according to various sources including the official documentation of Scrum as developed by the founders of the Agile framework as it is known today. A primary data poll targeting professionals who use “Agile” and Scrum highlights the different weights of importance that each professional places on the different elements without any significant weight being given to one specific element. The results can be seen in the Figure 29. below.

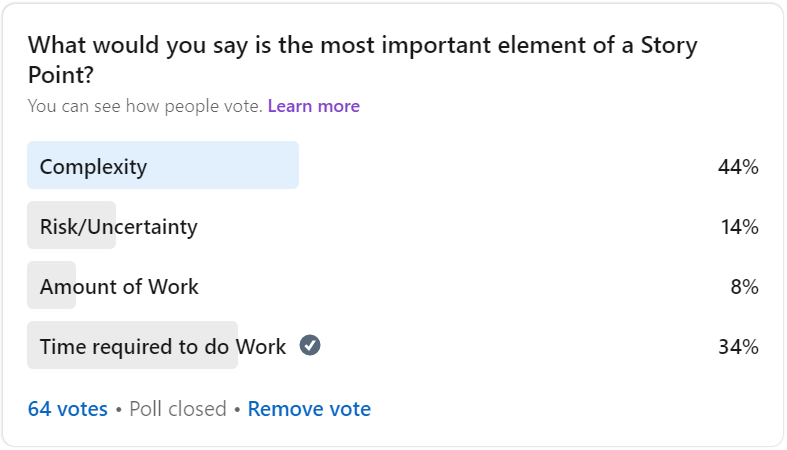


Figure 29. Poll about the Elements of Story Points

Understanding these elements would be critical to bridge any misunderstanding in the ‘black box’ that is the relativity of Story Points.

User Story Measurement

Despite the various techniques that exist for measuring Story Points, the most popularly used method is Planning Poker (also known as Agile Poker or Scrum Poker). Briefly discussed earlier in this paper, the non-scalar/interval nature of the Fibonacci scale which Planning Poker uses hampers the very essence of the relativity that is the aim of using Story Points. If a project team prefers relative estimation over absolute, a deeper sense of relativity should be ideally inherent in the characteristic of a Story Point – if a development team estimates a User Story to be twice or three times the size of the other that relativity should be able to be derived from the scores. This was evident in the challenges of the need for binning in the checks of association (resulting in data loss) as well as in the need for the use of dummy variables as a result of the categorical nature of this popularly used Fibonacci sequence.

Still, the ordinal scale of the Fibonacci sequence creates significantly large gaps at the higher end of the scale and very closely gathered intervals at the lower end. This, along with the customarily missed opportunity of taking an inventory of all User Stories (at either the product backlog or Sprint levels) in order to establish a base/anchor User Story, leaves significant room for misinterpretation and error. Therefore, the practice of analysing the baseline and “highest valued” User Stories before assigning Story Points should be a norm that is advocated by all Scrum Masters, trainers and promoters of Story Points.

Moreover, a more ‘suitable’ scale of measure might be used instead of the popularly adopted Fibonacci scale. Ideally a scale that captured the characteristics of an interval variable would provide a better assessment of relativity - that being:

* Giving a sense of order from lowest to highest or vice versa
* Meaning derived from the distance between the intervals

A numeric scale of perhaps “1” through “10” provides the filling characteristics that would make estimation simple enough:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **User Story ID** | **Complexity** | **Risk/Uncertainty** | **Amount of Work** | **Time required for Work** |
| Team Member 1 | 10 | 6 | 5 | 8 |
| Team Member 2 | 6 | 3 | 4 | 9 |
| Team Member 3 | 5 | 7 | 3 | 4 |
| Team Member 4 | 4 | 3 | 8 | 6 |
| Team Member 5 | 8 | 6 | 7 | 8 |
| Total | 33 | 25 | 27 | 35 |
| *Weight (Vote)* | *40%* | *20%* | *10%* | *30%* |
| **Story Point** | **13.2** | **5** | **2.7** | **10.5** |

Table 2. Suggested method for using Story Points

1. Difference between smallest and largest intervals is not so large that relativity is lost.
2. Clumped grouping at the lower end of the scale and exponential grouping at the higher end of the scale is avoided due to are even intervals between scores.
3. ‘Multiple’ relativity can be expressed (4 is twice as impactful as 2)
4. Concerns around forced consensus is mitigated as averaging points would be possible.
5. The Story Point axis of a velocity/burndown chart is more representative and to scale.

Following the example in Table 2., based on the “1” through “10” numeric/metric scale for each User Story, each development team member would vote on each separate element of a Story Point (the team can agree upon which elements they deem important for voting). Subsequently, each team member would vote on which element is believed to carry the heaviest weight in descending order of weight. Discrepancy can be discussed with some deference given to the team member who may be most responsible for the User Story and/or due to the metric nature of measurement, averages can be taken where necessary without the need for consensus. The sum of the elements can then be taken as a Story Point or the elements can stand individually.

Furthermore, as is evident in the poll above, given the different weights of importance placed on each element of what constitutes a Story Point, consideration may be given to assigning a score based on the same “1” to “10” scale to each element. Hence, alternatively/additionally, a weight, based on the rank of importance of each element to the team can then be applied and a weighted average can subsequently be calculated and assigned as the value of a Story Point (also illustrated in Table 2.).

Translation

One of the main reasons, aside from the difficulty that comes with providing absolute estimates, for the innovation behind Story Points as an absolute measure is that development team members were held accountable by the estimates provided for User Stories if those estimates weren’t met.  A culture of understanding that as an estimate is simply that – an estimate – should be encouraged throughout businesses because regardless of a company’s use of Story Points or not, employees should not be held to task by an estimate. Senior/Upper management now also holds project/development team members accountable to Story Points.[[59]](#footnote-59) This further reiterates the need for efforts around gaining a better understanding of relative estimation using Story Points what this may mean in time - the metric that Senior managers more practically relate to.

In business, however, measuring results/output is critically necessary. For example, reviewing demand of forecasts against actual sales is a key measure used to understand business operations and model predictability, and it is imperative for any business to remain competitive. As previously advocated, estimates should be treated as such, but similar checks of actuals against estimates provides the ability to reveal development opportunities which can drive a company forward.

Development teams’ choosing to track actuals against time estimates, similar to the practice of Company A in the case study, would provide an opportunity to model actual time as it relates to Story Point elements estimated in the manner suggested above. And with this translation, (and improved estimation) there can be greater understanding and communication of the different perspectives/values of time as viewed amongst senior level management, product managers and the development team.

Retrospective

At the end of each sprint, the development team conducts a retrospective (review). In a similar manner, the development team can (less) periodically use the processes conducted in the case study to enhance the estimation process which will enhance overall project planning. More specifically, these processes include test of association between the Story Point estimation (in the suggested numerically interval form) compared to actual recorded time, as well as time estimates compared to actual time. The correlation percentages of each would reveal potentially which, statistically, would be the more accurate/feasible estimation method. The liner regression function gives the modus for translation between the two measures of time.

It may not be necessary that these checks occur every sprint as is the case of retrospectives but as agility suggests, they may be executed as frequently as the project team sees fit.

# Discussion

This paper refers to a real-life business case in a specific company for its specific project.  While empirical results directly pertain to this direct situation, inherent in the customary practices that led to this business case are similarities which may be shared commonly shared in various projects across different industries and businesses. Therefore, recommendations found in this study are transferrable and can be and applied to various projects.

The project management software used to house data obtained in the case study contained the ability to store a great number of fields. While some of this data was recorded and many fields were used, in cleaning the data many fields that may have been pertinent to model building could not be included due to a significant of amount of missing data values per field. This includes numerically rated fields like “Risk” and “Priority” As well as fields categorical fields such as “Business Purpose”. These fields may have been critical dependent variables which potentially may have influenced the regression model.

Referring to the research question, the answer can only be given pertaining to this specific case. As a result of the nature of the categorical Story Point variable (in its current popularly used manner) being based on the Fibonacci sequence and the challenges associated with it:

* Inability to represent ‘multiples’ of effort (“13” is not 13 times as much as “1”)
* Closely grouped lower scores and increasingly widely spread higher scores
* Lost relativity due to differing intervals between scores

Coupled with the measure of association between time and actual estimates being significantly larger than the association between (binned) Story Point and actual time there is an inclination towards **affirming that Story Point as a not feasible estimation measure (in this specific case)**. In this light, however, there are considerations that should be taken into account.

Story Points We’re not used as generally recommended using the Fibonacci sequence. Some user stories were too large and then divided into smaller units with scores of 4, 6, 9 and even 0.25 and 0.75. This may explain some discrepancies in the ordinal aspect of some of the Story Points.

With respect to the association of Story Points to actual time, given the numeric nature of the Actual time variable, information was lost in the binning process. Assessment of the number of hours records could not be individually assessed per values of Story Points. If the Story Point variable were numeric as well, a Pearson’s correlation test or Spearman’s Rho would have been an appropriate test of association but the binning process was the only way to then perform a Cramer’s V (or any association) test due to the categorical nature of the Story Point variable.

Story Points, although they are an indirect measure of time measured by the amount number of Story Points that can be completed in a Sprint, they are not a direct measure of time as is evident in the poll conducted. While the linear regression function is used as a translator between Story Points and actual time (providing a potentially beneficial perspective) The analysis of association is based on solely one element that constitutes a Story Point – that being time. This may provide an explanation for the moderately low Cramer’s V in comparison to the higher calculated percentage association between two like (time) variables – Actuals versus Estimates. This may explain some discrepancies in the pairwise comparisons as well. Furthermore, this also alludes to the suggestion that assessing each Story Point element will be beneficial.

A concerning outcome of the linear regression model is the significance of the coefficients, many of which seemed to be insignificant. Statistically, there is no significant difference in effect on the dependent variable from many of the predictor variables. There can be a range of reasons for this outcome – from multicollinearity amongst the predictor variables, covariates or the sample size being too small to non-customary use of Story Points potentially affecting the data-collection process. The Wilcox Rank Sum two-sided test alludes to this statistical insignificance (see Figure 14.).

However, given the aim of demonstrating the process of creating a relationship between Story Points and time as precedent focus, as well as the fact that all the predictor variables are derived from a single categorical variable (Story Points) all predictors were included and used for translation between the two fields.

Applying a stepwise variable selection technique reveals the subset of variables – StoryPoint\_1, StoryPoint\_2, StoryPoint\_3 and StoryPoint\_5, StoryPoint\_7, StoryPoint\_8, and StoryPoint\_9, as the best function that can be produced reducing prediction error and that meets all linear model assumptions.

> #stepwise selection

> lm.step <- stepAIC(lm.actual, direction = "both", trace = F)

> summary(lm.step)

Call:

lm(formula = Actual ~ StoryPoint\_1 + StoryPoint\_2 + StoryPoint\_3 +

StoryPoint\_5 + StoryPoint\_7 + StoryPoint\_8 + StoryPoint\_9,

data = dumm.data)

Residuals:

Min 1Q Median 3Q Max

-3.0543 -0.8708 -0.0657 0.8192 4.6592

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.6875 0.2982 5.658 1.25e-07 \*\*\*

StoryPoint\_1 0.8982 0.3953 2.272 0.025043 \*

StoryPoint\_2 1.6968 0.4556 3.724 0.000312 \*\*\*

StoryPoint\_3 2.1833 0.5445 4.010 0.000112 \*\*\*

StoryPoint\_5 2.6579 0.5297 5.018 2.04e-06 \*\*\*

StoryPoint\_7 1.3600 0.8436 1.612 0.109814

StoryPoint\_8 2.4025 1.6061 1.496 0.137585

StoryPoint\_9 6.5225 1.6061 4.061 9.23e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.578 on 109 degrees of freedom

Multiple R-squared: 0.3027, Adjusted R-squared: 0.258

F-statistic: 6.761 on 7 and 109 DF, p-value: 1.166e-06

Figure 30. Script of Stepwise function results

With a p-value of 1.166-06, this function also meets the hypothesis that there is a (linear) relationship between the dependent and predictor variables. The improved R square value of approximately 26% was also produced. Most of the coefficients of this newly produced model as well are statistically significant as seen in Figure 30. above.

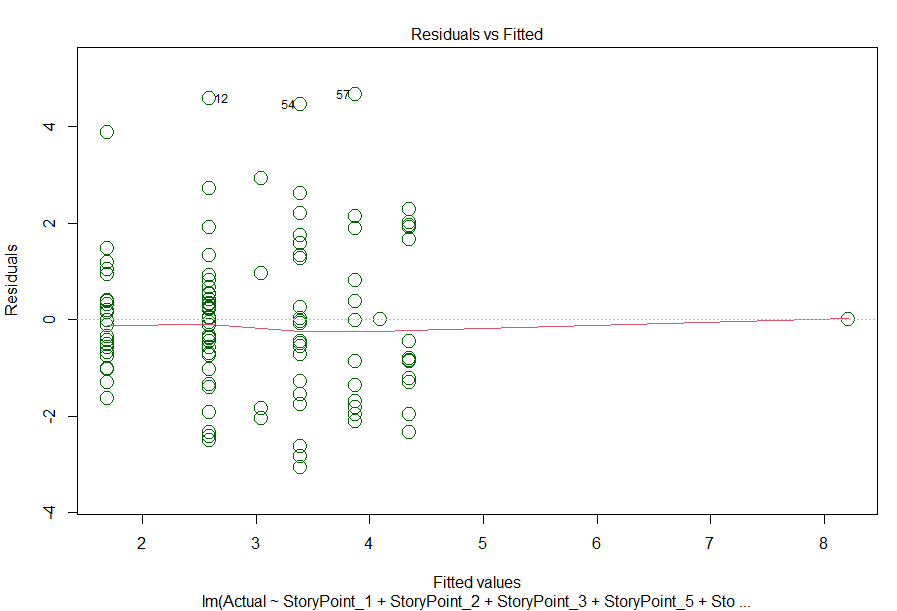


Figure 31. Stepwise Selection Lineal Model Residual Plot

Meeting the assumption that the linear regression model is homoscedastic, the residual plot seen in Figure 31. above demonstrates the constant nature of variability of the residuals.

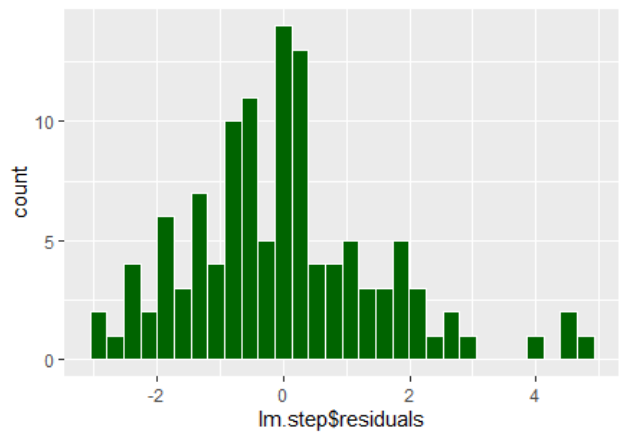


Figure 32. Histogram of Stepwise Selection Linear Model Residuals

Visually, the residuals appear to be normally distributed as seen in Figure 32. This will be further confirmed using a Kolmogorov-Smirnov Test in Figure 33. below.

> # test residuals for normality

> ks.test(lm.step$residuals, pnorm)

One-sample Kolmogorov-Smirnov test

data: lm.step$residuals

D = 0.10561, p-value = 0.147

alternative hypothesis: two-sided

Figure 33. Script of Kolmogorov-Smirnov Test for Residuals of Stepwise Selection Model

Interpreting the results of the Kolmogorov-Smirnov test, the null hypothesis (H0) is the following:

*H0: There is no difference between the distribution of the residuals and a normal distribution.*

The alternative hypothesis (H1) is the following:

*H1: There is a difference between the distribution of the residuals and a normal distribution.*

The residuals visually appear to be normally distributed and at a significance level of 5%, at a p-value of 0.147, the null hypothesis should not be rejected for the alternative hypothesis and therefore, the residuals are normally distributed.

Lastly, in assessing the independence of the residuals of the stepwise selection linear model, the results can be viewed in Figure 34.

> # test for independence

> cor(x = lm.step$residuals,

+ y = predict(lm.step),

+ method = c("pearson"))

[1] -9.097819e-18

Figure 34. Script showing test of Independence of Residuals of Stepwise Selection Model

As previously mentioned, while the most effective check for this is using the Durbin-Watson test (Akter, J., 2014, p. 137-44) for data where sequencing of the records is relevant (as with time series data). The data in this case study is not sequentially pertinent. Therefore, the correlation between the function residuals and the predicted values which shows a negligible correlation of approximately -0.9098-18, suggesting the residuals do not correspond with the predicted values (as is desired) and therefore assumed to be independent.

Although this model has more desirable results, the stepwise selection technique excludes levels of the Story Point variable and data is lost, reiterating the prudence of maintain all variables/levels in the function. Furthermore, reiterating the aim of having the ability of translating Story Points to Time and vice versa for the benefit of being able to be able to communicate estimates more effectively when there are protagonists and antagonists on a single team advocating for use of the opposing estimation methods, the coefficient values of the initial model are included.

While both estimation types (Story Points and time) were being recorded as well as actual time, upon the acknowledgment of the delay, Time estimates replaced Story Points as the primary metric of velocity. At this juncture, estimated times are being checked against actual times per task. The development team was provided feedback on discrepancies between time estimated time and actual time for a task. This was not done with the aim of reprimanding the development team members but to ensure budget limitations would be followed, as well as to provide developmental feedback. Being aware that they were being studied, the development team's consequential awareness may have had an impact on performance as prescribed per the “Hawthorne effect” (McCambridge et al, 2014, p. 267-77). As a result, both estimation and actual time records prior to and after the acknowledgement of the delay may have had an impact on modelling as well as association testing. In order to mitigate this going forward in future projects, the practice of checking actuals against estimates should clearly be communicated to the development team before project initiation, assuring them that the intent not to hold them accountable and this will not be held against their performance but for the purposes of identification of developmental opportunities and ultimately providing better estimates.

Adhering to the base principles of “Agile” (agility) this paper is not intended to superimpose its recommended technique held within over that which might currently be used in any project team. The reader should view this is a data-driven approach to assessing estimation. The process is transferable, and the recommendations and considerations may be regarded and applied to similar “Agile” and Scrum projects.

# List of Literature

Agile Alliance. 2020. *Sprint Backlog | Agile Alliance*. [online] Available at: <https://www.agilealliance.org/glossary/sprint-backlog/#q=~(infinite~false~filters~(postType~(~'page~'post~'aa\_book~'aa\_event\_session~'aa\_experience\_report~'aa\_glossary~'aa\_research\_paper~'aa\_video)~tags~(~'backlog))~searchTerm~'~sort~false~sortDirection~'asc~page~1)> [Accessed 1 February 2021].

Agile Alliance. 2020. *What are Story Points?*. [online] Available at: <https://www.agilealliance.org/glossary/points-estimates-in/> [Accessed 2 February 2021].

Agile Alliance. 2020. *What is a Burndown Chart?*. [online] Available at: <https://www.agilealliance.org/glossary/burndown-chart/> [Accessed 2 February 2021].

Agile Alliance. 2020. *What is Planning Poker?*. [online] Available at: <https://www.agilealliance.org/glossary/poker/> [Accessed 2 February 2021].

Agile Alliance. 2021. *About Agile Alliance*. [online] Available at: <https://www.agilealliance.org/the-alliance/> [Accessed 25 January 2021].

Agile Alliance. 2021. *What does INVEST Stand For?*. [online] Available at: <https://www.agilealliance.org/glossary/invest> [Accessed 1 February 2021].

Agile Alliance. 2021. *What is a Scrum Master?*. [online] Available at: <https://www.agilealliance.org/glossary/scrum-master/> [Accessed 6 February 2021].

Agile Alliance. 2021. *What Is Agile Software Development?*. [online] Available at: <https://www.agilealliance.org/agile101/> [Accessed 25 January 2021].

Akinwande, M., Dikko, H. and Samson, A., 2015. Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open Journal of Statistics*, [online] 05(07), pp.754-767. Available at: <https://www.scirp.org/html/11-1240578\_62189.htm?pagespeed=noscript> [Accessed 13 February 2021].

Akter, J., 2014. Bootstrapped Durbin– Watson Test of Autocorrelation for Small Samples. *ABC Journal of Advanced Research*, [online] 3(2), pp.137-142. Available at: <https://i-proclaim.my/journals/index.php/abcjar/article/view/39/39> [Accessed 11 March 2021].

Alkharusi, H., 2012. Categorical variables in regression analysis: A comparison of dummy and effect coding. *International Journal of Education*, *4*(2), p.202.

Atlassian Marketplace. 2021. *Agile Poker for Jira-planning&estimation*. [online] Available at: <https://marketplace.atlassian.com/apps/700473/agile-poker-for-jira-planning-estimation?hosting=cloud&tab=overview> [Accessed 28 February 2021].

Atlassian. 2020. *Story Points & The Evolution Of Agile Estimation*. Video. https://www.youtube.com/watch?v=\_N5gj9gzOjg.

Atlassian. 2021. *What Is Agile? | Atlassian*. [online] Available at: <https://www.atlassian.com/agile> [Accessed 25 January 2021].

Baranoshknik, K., 2021. *About – Kiryl Baranoshnik*. [online] Baranoshnik.com. Available at: <https://baranoshnik.com/about/> [Accessed 6 February 2021].

Baranoshnik, K., 2021. *Why I Stopped Using Planning Poker*. [online] Medium. Available at: <https://medium.com/agilelab/why-i-stopped-using-planning-poker-f6e7f1e6dbda> [Accessed 6 February 2021].

Bluman, A., 2006. *Elementary statistics*. 3rd ed. Boston: McGraw-Hill Higher Education, p.379.

Datanyze. 2021. *Project Management Market Share Report | Competitor Analysis | Jira, Microsoft Project, Smartsheet*. [online] Available at: <https://www.datanyze.com/market-share/project-management--217> [Accessed 25 January 2021].

Davidson, D., 2021. *Why do we use Story Points for Estimating?*. [online] Scrum.org. Available at: <https://www.scrum.org/resources/blog/why-do-we-use-story-points-estimating> [Accessed 2 February 2021].

Denning, S., 2021. *What Is Agile?*. [online] Forbes. Available at: <https://www.forbes.com/sites/stevedenning/2016/08/13/what-is-agile/> [Accessed 25 January 2021].

Discovery.hgdata.com. 2021. *Companies Using Atlassian JIRA Software, Market Share, Customers and Competitors*. [online] Available at: <https://discovery.hgdata.com/product/atlassian-jira-software> [Accessed 29 January 2021].

Discovery.hgdata.com. 2021. *Companies Using Microsoft Azure DevOps Server, Market Share, Customers and Competitors*. [online] Available at: <https://discovery.hgdata.com/product/microsoft-azure-devops-server> [Accessed 29 January 2021].

Ditigal.ai, 2018. *13th Annual State Of Agile Report*. State of Agile Report. [online] p.9. Available at: <https://stateofagile.com/> [Accessed 26 January 2021].

Ditigal.ai, 2021. *14Th Annual State Of Agile Report*. State of Agile Report. [online] p.10. Available at: <https://stateofagile.com/> [Accessed 25 January 2021].

Ditigal.ai, 2021. *14Th Annual State Of Agile Report*. State of Agile Report. [online] p.10. Available at: <https://stateofagile.com/> [Accessed 26 January 2021].

Docs.microsoft.com. 2021. *Define features and epics, organize backlog items - Azure Boards*. [online] Available at: <https://docs.microsoft.com/en-us/azure/devops/boards/backlogs/define-features-epics?view=azure-devops&tabs=agile-process> [Accessed 2 February 2021].

En.agilelab.org. 2021. *AgileLAB – Agile and Lean trainings*. [online] Available at: <https://en.agilelab.org/> [Accessed 6 February 2021].

Epam.com. 2021. *EPAM | Software Engineering & Product Development Services*. [online] Available at: <https://www.epam.com/> [Accessed 6 February 2021].

Fair, J., 2012. Agile versus Waterfall: approach is right for my ERP project?. In: *PMI® Global Congress*. [online] Newtown Square: Project Management Institute. Available at: <https://www.pmi.org/learning/library/agile-versus-waterfall-approach-erp-project-6300> [Accessed 28 January 2021].

Fortunebusinessinsights.com. 2021. *DevOps Market Size, Share, Growth, Trends | Research Report, 2026*. [online] Available at: <https://www.fortunebusinessinsights.com/devops-market-102040> [Accessed 29 January 2021].

Fowler, M., 2018. *The State of Agile Software in 2018*. [online] martinfowler.com. Available at: <https://martinfowler.com/articles/agile-aus-2018.html> [Accessed 2 February 2021].

GOTO Conferences, 2015. *GOTO 2015 • Agile is Dead • Pragmatic Dave Thomas*. [video] Available at: <https://www. youtube.com/watch?v=a-BOSpxYJ9M> [Accessed 2 February 2021].

Green, P., 2017. *Scrum Foundations eLearning Series Educational Videos*. [online] Scrumalliance.org. Available at: <https://www.scrumalliance.org/learn-about-scrum/scrum-elearning-series/scrum-artifacts#:~:text=In%20archaeology%2C%20the%20term%20%E2%80%9Cartifact,was%20made%20by%20a%20human.&text=Scrum%20describes%20three%20primary%20artifacts,below%20to%20view%20the%20videos> [Accessed 1 February 2021].

Grenning, J., 2002. *Planning Poker or How to avoid analysis paralysis while release planning*. [ebook] Available at: <https://wingman-sw.com/papers/PlanningPoker-v1.1.pdf> [Accessed 2 February 2021].

Hardy, M., 1993. *Regression with dummy variables*. Newbury Park, CA: Sage Publications, p.8.

Hartman, B., 2021. *An Introduction to Planning Poker*. [ebook] Sacramento. Available at: <http://athena.ecs.csus.edu/~buckley/CSc231\_files/Introduction%20to%20Planning%20Poker.pdf> [Accessed 2 February 2021].

Highsmith, J., 2021. *History: The Agile Manifesto*. [online] Agilemanifesto.org. Available at: <https://agilemanifesto.org/history.html> [Accessed 25 January 2021].

Huber, S., Wiemer, H., Schneider, D. and Ihlenfeldt, S., 2019. DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. *Procedia CIRP*, [online] 79, pp.403-408. Available at: <(http://www.sciencedirect.com/science/article/pii/S2212827119302239)> [Accessed 3 February 2021].

In: *Cambridge Dictionary*. 2021. [online] Cambridge University Press. Available at: <https://dictionary.cambridge.org/dictionary/english/artifact> [Accessed 1 February 2021].

Jeffries, R., 1998. *Developers Should Abandon Agile*. [online] Ron Jeffries. Available at: <https://www.ronjeffries.com/articles/018-01ff/abandon-1/> [Accessed 2 February 2021].

Joubert, S., 2020. HOW MUCH DOES A PROJECT MANAGER MAKE?. [Blog] *Northeastern University Graduate Programs Blog*, Available at: <https://www.northeastern.edu/graduate/blog/ways-to-increase-pmp-salary/#:~:text=In%20the%20U.S.%2C%20the%20median,earning%20between%20%2493%2C000%20and%20%24140%2C000.> [Accessed 28 January 2021].

Lin, Z., Hewett, M. and Altman, R.B., 2002. Using binning to maintain confidentiality of medical data. In *Proceedings of the AMIA Symposium* (p. 454). American Medical Informatics Association. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2244360/> [Accessed 8 March 2021]

Liu, H., 2015. Comparing Welch's ANOVA, a Kruskal-Wallis test and traditional ANOVA in case of Heterogeneity of Variance.

Long, D., Axell, R., Ganney, P. and Taktak, A., 2020. *Clinical Engineering*. 2nd ed. Academic Press, pp.131-168.

McCambridge, J., Witton, J. and Elbourne, D., 2014. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, [online] 67(3), pp.267-277. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3969247/ pdf/main.pdf> [Accessed 8 March 2021].

Misra, A., 2016. *Story Points - Advantages and Disadvantages*. [online] Linkedin.com. Available at: <https://www.linkedin.com/pulse/advantages-disadvantages-using-story-points-anshika-misra/> [Accessed 2 February 2021].

O'Connor, S., 2020. AGILE VS. SCRUM: WHAT’S THE DIFFERENCE?. [Blog] *North Easter University Graduate Programs Blogs*, Available at: <https://www.northeastern.edu/graduate/blog/agile-vs-scrum/> [Accessed 26 January 2021].

Online.stat.psu.edu. 2018. *10.7 - Detecting Multicollinearity Using Variance Inflation Factors | STAT 462*. [online] Available at: <https://online.stat.psu.edu/stat462/node/180/> [Accessed 13 February 2021].

PON - Program on Negotiation at Harvard Law School. 2021. *What is the Anchoring Effect?*. [online] Available at: <https://www.pon.harvard.edu/tag/anchoring-effect/> [Accessed 6 February 2021].

Project Management Institute Inc., 2020. *Consolidated Financial Statements*. [online] Newtown Square: Project Management Institute Inc. Available at: <https://www.pmi.org/-/media/pmi/documents/public/pdf/microsites/2019-annual-report/pmi-2019-financial-statements.pdf?v=174c161d-1805-4ec4-8788-fa2601f111ee> [Accessed 29 January 2021].

Project Management Institute. 2021. *About Us*. [online] Available at: <https://www.pmi.org/about> [Accessed 28 January 2021].

Project Management Institute. 2021. *Certifications*. [online] Available at: <https://www.pmi.org/certifications> [Accessed 28 January 2021].

Radigan, D., 2021. *What are Story Points and how do you estimate them?*. [online] Atlassian. Available at: <https://www.atlassian.com/agile/project-management/estimation> [Accessed 2 February 2021].

Rehkopf, M., 2021. *Epics, Stories, Themes, and Initiatives | Atlassian*. [online] Atlassian. Available at: <https://www.atlassian.com/agile/project-management/epics-stories-themes> [Accessed 2 February 2021].

Rehkopf, M., 2021. *Learn burndown charts with Jira Software | Atlassian*. [online] Atlassian. Available at: <https://www.atlassian.com/agile/tutorials/burndown-charts> [Accessed 2 February 2021].

Rehkopf, M., 2021. *User Stories | Examples and Template | Atlassian*. [online] Atlassian. Available at: <https://www.atlassian.com/agile/project-management/user-stories> [Accessed 1 February 2021].

Rehkopf, M., 2021. *What is a Scrum Master? | Atlassian*. [online] Atlassian. Available at: <https://www.atlassian.com/agile/scrum/scrum-master> [Accessed 8 February 2021].

Research and Markets. 2021. *U.S. Enterprise Agile Transformation Services Market: Opportunity Analysis And Industry Forecast, 2019-2026*. [online] Available at: <https://www.researchandmarkets.com/reports/4844593/u-s-enterprise-agile-transformation-services?utm\_source=BW&utm\_medium=PressRelease&utm\_code=wh3mzn&utm\_campaign=1303074+-+United+States%27+%2418%2b+Billion+Enterprise+Agile+Transformation+Services+Market%2c+2019-2026&utm\_exec=joca220prd> [Accessed 29 January 2021].

Resources.scrumalliance.org. 2021. *Scrum Roles Demystified*. [online] Available at: <https://resources.scrumalliance.org/Article/scrum-roles-demystified> [Accessed 1 February 2021].

Schwaber, K., 1995. SCRUM Development Process. In: *OOPSLA Business Object Design and Implementation Workshop*. [online] London: Springer, p.17. Available at: <http://www.jeffsutherland.org/oopsla/schwapub.pdf> [Accessed 26 January 2021].

Schwaber, K., 1995. SCRUM Development Process. In: *OOPSLA Business Object Design and Implementation Workshop*. [online] London: Springer, p.16. Available at: <http://www.jeffsutherland.org/oopsla/schwapub.pdf> [Accessed 26 January 2021].

Scrum Inc. 2019. *Scrum Velocity*. [online] Available at: <https://www.scruminc.com/velocity/> [Accessed 2 February 2021].

Scrumalliance.org. 2021. *Agile and Scrum Training & Certification | Scrum Alliance*. [online] Available at: <https://www.scrumalliance.org/get-certified> [Accessed 28 January 2021].

Scrumalliance.org. 2021. *The Difference Between Agile And Scrum | Scrum Alliance*. [online] Available at: <https://www.scrumalliance.org/about-scrum/definition> [Accessed 26 January 2021].

Sliger, M., 2012. *Agile Estimation Techniques*. [online] Project Management Institute. Available at: <https://www.pmi.org/learning/library/agile-project-estimation-techniques-6110> [Accessed 2 February 2021].

Takeuchi, H. and Nonaka, I., 1986. The New New Product Development Game. *Harvard Business Review*, [online] (64). Available at: <https://hbr.org/1986/01/the-new-new-product-development-game> [Accessed 25 January 2021]. Takeuchi, H. and Nonaka, I., 1986. The New New Product Development Game. *Harvard Business Review*, [online] (64). Available at: <https://hbr.org/1986/01/the-new-new-product-development-game> [Accessed 25 January 2021].

United States Securities and Exchange Commission, n.d. *Form 20-F ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934*. Atlassian Corporation Plc., p.15. Available at: <https://www.sec.gov/Archives/edgar/data/1650372/000104746915008450/a2226437zf-1.htm> [Accessed 29 January 2021].

UNITED STATES. (2019). Return of Organization Exempt From Income Tax. [Washington, D.C.], Dept. of the Treasury, Internal Revenue Service. Available at: <https://www.scrumalliance.org/ScrumRedesignDEVSite/media/ScrumAllianceMedia/Files%20and%20PDFs/About%20Us/Tax%20Forms%20and%20Bylaws/2018-Form-990.pdf> [Accessed 29 January 2021]

Van Casteren, W., 2017. *The Waterfall Model and the Agile Methodologies: A comparison by project characteristics*. Research Gate, pp.1-6. Available at: < https://www.researchgate.net/profile/Wilfred\_Van\_Casteren/publication/313768756\_The\_Waterfall\_Model\_and\_the\_Agile\_Methodologies\_A\_comparison\_by\_project\_characteristics/links/58a56a5ca6fdcc0e07648bb6/The-Waterfall-Model-and-the-Agile-Methodologies-A-comparison-by-project-characteristics.pdf> [Accessed 28 January 2021].

West, D., 2021. *Agile Scrum Roles | Atlassian*. [online] Atlassian. Available at: <https://www.atlassian.com/agile/scrum/roles> [Accessed 1 February 2021].

White, S., 2021. *Top 15 IT certifications in demand for 2021*. [online] CIO. Available at: <https://www.cio.com/article/3562331/top-15-it-certifications-in-demand-for-2021.html> [Accessed 2 February 2021].

World Rugby, *A Beginners’S Guide To Rugby Union : The Scrum*. [online] Available at: <https://passport.worldrugby.org/?page=beginners&p=12> [Accessed 25 January 2021].

# Statutory Declaration

I herewith formally declare that I have written the submitted thesis independently. I did not use any outside support except for the quoted literature and other sources mentioned in the paper.

I clearly marked and separately listed all of the literature and all of the other sources which I employed when producing this academic work, either literally or in content.

I am aware that the violation of this regulation will lead to failure of the thesis.

Damian H. Spencer 

Student’s name Student’s signature

s0567487 16 March, 2021

Matriculation number Berlin, date

1. Highsmith, J., 2021, *History: The Agile Manifesto*, viewed 25 January 2021, <https://agilemanifesto.org/history.html>. [↑](#footnote-ref-1)
2. Datanyze. 2021, *Project Management Market Share Report | Competitor Analysis | Jira, Microsoft Project, Smartsheet*, 25 January 2021, <https://www.datanyze.com/market-share/project-management--217>. [↑](#footnote-ref-2)
3. Atlassian, 2021, *What Is Agile?*, viewed 25 January 2021, <https://www.atlassian.com/agile>. [↑](#footnote-ref-3)
4. Denning S., 2021. *What Is Agile?*, Forbes, viewed 25 January 2021, <https://www.forbes.com/sites/stevedenning/2016/08/13/what-is-agile/>. [↑](#footnote-ref-4)
5. Agile Alliance, 2021, *About Agile Alliance*, viewed Accessed 25 January 2021, <https://www.agilealliance.org/the-alliance/>. [↑](#footnote-ref-5)
6. Agile Alliance, 2021, *What Is Agile Software Development?*, viewed Accessed 25 January 2021, <https://www.agilealliance.org/agile101/>. [↑](#footnote-ref-6)
7. Ditigal.ai, 2021. *14Th Annual State Of Agile Report*, viewed 25 January 2021, <https://stateofagile.com/> [↑](#footnote-ref-7)
8. World Rugby, 2021. *A Beginner’s Guide To Rugby Union: The Scrum*, viewed 25 January 2021, <https://passport.worldrugby.org/?page=beginners&p=12>. [↑](#footnote-ref-8)
9. Thomas, D., 2015, *Agile is Dead Pragmatic*,GOTO Conferences, viewed 2 February 2021, <https://www.youtube.com/watch?v=a-BOSpxYJ9M>. [↑](#footnote-ref-9)
10. O'Connor, S., 2020. *AGILE VS. SCRUM: WHAT’S THE DIFFERENCE?*, North Easter University Graduate Programs Blogs, viewed 26 January 2021, <https://www.northeastern.edu/graduate/blog/agile-vs-scrum/>. [↑](#footnote-ref-10)
11. Scrum Alliance, 2021. *The Difference Between Agile And Scrum*, viewed 26 January 2021,<https://www.scrumalliance.org/about-scrum/definition>. [↑](#footnote-ref-11)
12. Ditigal.ai, 2021. *14Th Annual State Of Agile Report*. State of Agile Report, viewed 26 January 2021, <https://stateofagile.com/>. [↑](#footnote-ref-12)
13. Ditigal.ai, 2021. *14Th Annual State Of Agile Report*. State of Agile Report, viewed 26 January 2021, <https://stateofagile.com/>. [↑](#footnote-ref-13)
14. Agile Alliance, 2021, *What is a Scrum Master?*, viewed 6 February 2021, <https://www.agilealliance.org/glossary/scrum-master/>. [↑](#footnote-ref-14)
15. Rehkopf, M., 2021, *What is a Scrum Master?*, Atlassian, viewed 8 February 2021, <https://www.atlassian.com/agile/scrum/scrum-master>. [↑](#footnote-ref-15)
16. Fair, J., 2012. *Agile versus Waterfall: approach is right for my ERP project?*, Project Management Institute, viewed 28 January 2021, <https://www.pmi.org/learning/library/agile-versus-waterfall-approach-erp-project-6300>. [↑](#footnote-ref-16)
17. Project Management Institute, 2021, *About Us*, viewed 28 January 2021, <https://www.pmi.org/about>. [↑](#footnote-ref-17)
18. Joubert, S., 2020, *HOW MUCH DOES A PROJECT MANAGER MAKE?*, Northeastern University GraduatePrograms, viewed 28 January 2021, <https://www.northeastern.edu/graduate/blog/ways-to-increase-pmp-salary/#:~:text=In%20the%20U.S.%2C%20the%20median,earning%20between%20%2493%2C000%20and%20%24140%2C000.>. [↑](#footnote-ref-18)
19. Project Management Institute, 2021, *Certifications*, viewed 28 January 2021, <https://www.pmi.org/certifications>. [↑](#footnote-ref-19)
20. Scrum Alliance, 2021, *Agile and Scrum Training & Certification viewed* 28 January 2021, <https://www.scrumalliance.org/get-certified>. [↑](#footnote-ref-20)
21. Project Management Institute Inc., 2020, *Consolidated Financial Statements*, viewed as 29 January 2021 <https://www.pmi.org/-/media/pmi/documents/public/pdf/microsites/2019-annual-report/pmi-2019-financial-statements.pdf?v=174c161d-1805-4ec4-8788-fa2601f111ee>. [↑](#footnote-ref-21)
22. UNITED STATES, 2019, *Return of Organization Exempt From Income Tax*, Dept. of the Treasury, Internal Revenue Service, viewed 29 January, 2019, <https://www.scrumalliance.org/ScrumRedesignDEVSite/media/ScrumAllianceMedia/Files%20and%20PDFs/About%20Us/Tax%20Forms%20and%20Bylaws/2018-Form-990.pdf [↑](#footnote-ref-22)
23. HG Insights, 2021, *Companies Using Atlassian JIRA Software, Market Share, Customers and Competitors*. 29 January 2021 <https://discovery.hgdata.com/product/atlassian-jira-software>. [↑](#footnote-ref-23)
24. Fortune Business Insights, 2021, *DevOps Market Size, Share, Growth, Trends | Research Report, 2026*, viewed 29 January 2021, <https://www.fortunebusinessinsights.com/devops-market-102040>. [↑](#footnote-ref-24)
25. Research and Markets, 2021, *U.S. Enterprise Agile Transformation Services Market: Opportunity Analysis And Industry Forecast, 2019-2026*, viewed 29 January 2021, <https://www.researchandmarkets.com/reports/4844593/u-s-enterprise-agile-transformation-services?utm\_source=BW&utm\_medium=PressRelease&utm\_code=wh3mzn&utm\_campaign=1303074+-+United+States%27+%2418%2b+Billion+Enterprise+Agile+Transformation+Services+Market%2c+2019-2026&utm\_exec=joca220prd>. [↑](#footnote-ref-25)
26. Scrum Alliance, 2021, *Scrum Roles Demystified*, viewed 1 February 2021, <https://resources.scrumalliance.org/Article/scrum-roles-demystified>. [↑](#footnote-ref-26)
27. West, D., 2021, *Agile Scrum* Roles, Atlassian. viewed 1 February 2021, <https://www.atlassian.com/agile/scrum/roles>. [↑](#footnote-ref-27)
28. Cambridge Dictionary, 2021, Cambridge University Press, viewed 1 February 2021, <https://dictionary.cambridge.org/dictionary/english/artifact>. [↑](#footnote-ref-28)
29. Green, P., 2017, *Scrum Foundations eLearning Series Educational Videos*, Scrum Alliance, viewed 1 February 2021, <https://www.scrumalliance.org/learn-about-scrum/scrum-elearning-series/scrum-artifacts#:~:text=In%20archaeology%2C%20the%20term%20%E2%80%9Cartifact,was%20made%20by%20a%20human.&text=Scrum%20describes%20three%20primary%20artifacts,below%20to%20view%20the%20videos>. [↑](#footnote-ref-29)
30. Agile Alliance, 2020, *Sprint Backlog*, viewed 1 February 2021, <https://www.agilealliance.org/glossary/sprint-backlog/#q=~(infinite~false~filters~(postType~(~'page~'post~'aa\_book~'aa\_event\_session~'aa\_experience\_report~'aa\_glossary~'aa\_research\_paper~'aa\_video)~tags~(~'backlog))~searchTerm~'~sort~false~sortDirection~'asc~page~1)>. [↑](#footnote-ref-30)
31. Rehkopf, M., 2021, *User Stories | Examples and* Template, Atlassian, viewed 1 February 2021, <https://www.atlassian.com/agile/project-management/user-stories>. [↑](#footnote-ref-31)
32. Agile Alliance, 2021, *What does INVEST Stand For?*, viewed 1 February 2021, <https://www.agilealliance.org/glossary/invest>. [↑](#footnote-ref-32)
33. Rehkopf, M., 2021, *Epics, Stories, Themes, and Initiatives*, Atlassian, viewed 2 February 2021, <https://www.atlassian.com/agile/project-management/epics-stories-themes>. [↑](#footnote-ref-33)
34. Microsoft, 2021, *Define features and epics, organize backlog items - Azure Boards*, viewed 2 February 2021, <https://docs.microsoft.com/en-us/azure/devops/boards/backlogs/define-features-epics?view=azure-devops&tabs=agile-process>. [↑](#footnote-ref-34)
35. Radigan, D., 2021, *What are Story Points and how do you estimate them?*, Atlassian, viewed 2 February 2021, <https://www.atlassian.com/agile/project-management/estimation>. [↑](#footnote-ref-35)
36. Agile Alliance, 2021, *What are Story Points?*, viewed 2 February 2021, <https://www.agilealliance.org/glossary/points-estimates-in/>. [↑](#footnote-ref-36)
37. Misra, A., 2016, *Story Points - Advantages and Disadvantages*, viewed, 2 February 2021, <https://www.linkedin.com/pulse/advantages-disadvantages-using-story-points-anshika-misra>]. [↑](#footnote-ref-37)
38. Scrum Inc, 2019, *Scrum Velocity*, viewed 2 February 2021, <https://www.scruminc.com/velocity/>. [↑](#footnote-ref-38)
39. Rehkopf, M., 2021, *Learn burndown charts with Jira Software*, Atlassian, 2 February 2021, <https://www.atlassian.com/agile/tutorials/burndown-charts>. [↑](#footnote-ref-39)
40. Agile Alliance, 2020, *What is a Burndown Chart?*, viewed 2 February 2021, <https://www.agilealliance.org/glossary/burndown-chart/>. [↑](#footnote-ref-40)
41. Davidson, D., 2021, *Why do we use Story Points for Estimating?*, Scrum.org., 2 February 2021, <https://www.scrum.org/resources/blog/why-do-we-use-story-points-estimating>. [↑](#footnote-ref-41)
42. Sliger, M., 2012, *Agile Estimation Techniques*, Project Management Institute, 2 February 2021, <https://www.pmi.org/learning/library/agile-project-estimation-techniques-6110>. [↑](#footnote-ref-42)
43. Atlassian Marketplace, 2021, *Agile Poker for Jira-planning&estimation*, viewed 28 February 2021, <https://marketplace.atlassian.com/apps/700473/agile-poker-for-jira-planning-estimation?hosting=cloud&tab=overview>. [↑](#footnote-ref-43)
44. Hartman, B., 2021, *An Introduction to Planning Poker*, viewed 2 February 2021, <http://athena.ecs.csus.edu/~buckley/CSc231\_files/Introduction%20to%20Planning%20Poker.pdf>. [↑](#footnote-ref-44)
45. Grenning, J., 2002, *Planning Poker or How to avoid analysis paralysis while release planning*. Viewed 2 February 2021, <https://wingman-sw.com/papers/PlanningPoker-v1.1.pdf>. [↑](#footnote-ref-45)
46. Agile Alliance. 2020. *What is Planning Poker?*, viewed 2 February 2021 <https://www.agilealliance.org/glossary/poker/>. [↑](#footnote-ref-46)
47. Baranoshknik, K., 2021, *About – Kiryl Baranoshnik*. Baranoshnik.com, viewed 6 February 2021, <https://baranoshnik.com/about/>. [↑](#footnote-ref-47)
48. Epam, 2021, *Software Engineering & Product Development Services*, viewed 6 February 2021, <https://www.epam.com/>. [↑](#footnote-ref-48)
49. AgileLab, 2021, *AgileLAB – Agile and Lean trainings*. viewed 6 February 2021, <https://en.agilelab.org/>. [↑](#footnote-ref-49)
50. Baranoshnik, K., 2021, *Why I Stopped Using Planning Poker*, Medium. Viewed 6 February 2021, <https://medium.com/agilelab/why-i-stopped-using-planning-poker-f6e7f1e6dbda>. [↑](#footnote-ref-50)
51. Program on Negotiation at Harvard Law School, 2021, *What is the Anchoring Effect?*, viewed 6 February 2021 <https://www.pon.harvard.edu/tag/anchoring-effect/>. [↑](#footnote-ref-51)
52. Jeffries, R., 1998, *Developers Should Abandon Agile*, viewed 2 February 2021, <https://www.ronjeffries.com/articles/018-01ff/abandon-1/>. [↑](#footnote-ref-52)
53. Fowler, M., 2018, *The State of Agile Software in 2018*, viewed 2 February 2021, <https://martinfowler.com/articles/agile-aus-2018.html>. [↑](#footnote-ref-53)
54. White, S., 2021, *Top 15 IT certifications in demand for 2021*, CIO, viewed 2 February 2021, <https://www.cio.com/article/3562331/top-15-it-certifications-in-demand-for-2021.html>. [↑](#footnote-ref-54)
55. United States Securities and Exchange Commission, n.d. *Form 20-F ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURITIES EXCHANGE ACT OF 1934*. *Atlassian Corporation Plc.*, viewed 2 Feb 2021, <https://www.sec.gov/Archives/edgar/data/1650372/000104746915008450/a2226437zf-1.htm> [↑](#footnote-ref-55)
56. (x-y] translates to greater than x but less than or equal to y [↑](#footnote-ref-56)
57. Nine equal-length groups were 41.013% while Seventeen equal-length groups were 45.146% associated. [↑](#footnote-ref-57)
58. Pennsylvania State University Eberly College of Science, 2018, *10.7 - Detecting Multicollinearity Using Variance Inflation Factors*, viewed 13 Feb 2021 <https://online.stat.psu.edu/stat462/node/180/>. [↑](#footnote-ref-58)
59. Atlassian, 2020, *The Evolution Of Agile Estimation*, viewed 23 February 2021, <https://www.youtube.com/watch?v=\_N5gj9gzOjg>. [↑](#footnote-ref-59)