Towards a Data-Driven Pricing Decision

With the Help of A/B Testing

EYGLO MYRRA OSKARSDOTTIR

Master’s Degree Project
Stockholm, Sweden May 2016
Towards a Data-Driven Pricing Decision
With the Help of A/B Testing

Eyglo Myrra Oskarsdottir
Abstract
An A/B test is implemented on a SaaS firm’s product page to examine the difference in conversion rates from website visitors who are randomly assigned to two different product-landing pages that show different prices. To count as a successful conversion the visitors that view a product-landing page have to click on a “Free Trial” button. Half of the group will be assigned the treatment page, which will state higher prices and the other half will be assigned the controlled page, which will state today’s current price. The only variant that will differ from the two pages will be the stated price of the product and all other factors will be kept constant. The controlled experiment is executed to get a sense of customers’ price sensitivity, hence this thesis contributes to microeconomic research of the private sector, more specifically to the ICT industry by using a novel approach with the help of A/B testing on prices. The results showed no statistical significance difference between the two variations, which can be translated to accepting the null hypothesis; the demand for a particular Software-As-A-Service product will hold unchanged after the proposed price increase. At first, this could be a surprising result but when looking into the industry, which the firm participates in and their early mover advantages this result could have been assumed.

Keywords
Microeconomics, Controlled Experiment, A/B Testing, Price Sensitivity
“There are no secrets to success. It is the result of preparation, hard work, and learning from failure.”

Colin Powell
Acknowledgements

Firstly I would like to thank my thesis advisors professor Hans Lööf and professor Ali Mohammadi at the Royal Institute of Technology. They provided important guidance and valuable feedback during the thesis process. I would also like to thank Tempo Software for allowing me to implement a controlled experiment on their website and gather real time data on current and potential customers. Thereafter, I would also like to thank the A/B testing expert at Spotify, Evan Shrubsole, for introducing me to the most common pitfalls in A/B testing that I was able to account for during my experiment. Next, my thanks go to Bjorn Kristmannsson who helped me patiently to format my thesis exactly the way I wanted it. Finally, my thanks go to my friends that kept me company during the long process and my family that motivated me on tough days. Without everyone’s input and help I would not have been able to proudly turn in my thesis.
# Contents

Abstract 2

Acknowledgements 4

List of Figures 7

List of Tables 8

Abbreviations 9

1 Introduction 11

1.1 Study Objective and Problem statement .......................... 12
1.2 Background ......................................................... 12
1.3 Purpose and Scope of the study ................................. 14
1.4 Limitations ......................................................... 15

2 Literature Review 17

2.1 Definition and Characteristics of SAAS .......................... 17
2.2 Theory of Pricing and Output Decisions .......................... 18
2.3 Theory of Demand and Price elasticity .......................... 18
2.4 Controlled Experiments on the Web ............................. 19
2.5 Economic Field Experiments on prices ........................... 21

3 Empirical Model, Methodology and Data 23

3.1 Empirical Model .................................................... 23
  3.1.1 Recommendations ............................................. 23
  3.1.2 Preparing an A/B Test ....................................... 24
  3.1.3 Implementation Architecture ............................... 25
3.2 Methodology ...................................................... 27
  3.2.1 Phase 1-A/A Test ............................................. 27
  3.2.2 Phase 2-Effective Sample Size ............................ 29
  3.2.3 Phase 3-A/B Test ............................................. 30
3.3 Data .............................................................. 32
# List of Figures

3.1 Page client-side assignment method ................................. 26  
3.2 Flow chart of A/A test .............................................. 27  
3.3 Appearance of variations ............................................ 30  
3.4 Flow chart of A/B test .............................................. 31  

4.1 A/B Test Result Graph .............................................. 34  
4.2 Total number of evaluation signups across years over the same time period .............................................. 37  
4.3 Monthly evaluation signups for 2014 ............................... 38  
4.4 Monthly evaluation signups for 2015 ............................... 38  
4.5 Monthly evaluation signups for 2016 ............................... 38
List of Tables

1.1  Server Product Tiers .............................................. 13
1.2  Cloud Product Tiers ............................................... 13
3.1  Differences in prices between groups .......................... 31
4.1  A/B testing results ............................................... 34
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>SaaS</td>
<td>Software-As-A-Service</td>
</tr>
<tr>
<td>OEC</td>
<td>Overall Evaluation Criterion</td>
</tr>
<tr>
<td>A/B Test</td>
<td>Controlled Experiment where two different variations are assigned to users</td>
</tr>
<tr>
<td>Tempo</td>
<td>Tempo Software, the Saas firm implementing A/B test on their website</td>
</tr>
</tbody>
</table>
I dedicate this thesis to my two sisters, Tinna Osk Oskarsdottir and Thorunn Dia Oskarsdottir. Tinna Osk played a role model for me during my master education.

She was the one that changed my attitude towards writing the thesis, which kept me positive during the process and allowed me to set high goals. Thorunn Dia, is the reason why I started the program first of all. She is the one that motivated me to apply for this program, which led me to a great city, valuable education and growth.
Chapter 1

Introduction

The use of real time data from an A/B testing experiment can lead the way to data-driven pricing decisions for the novel business industry, SaaS (for definition of A/B testing see D). The creation of the Internet has assisted the formation of modern markets, which separate themselves from current markets by large scale, increased customization, rapid innovation and the collection and use of detailed consumer and market data (Levin, 2016). It has facilitated the creation of the SaaS model, which allows software vendors to tie together software application, IT infrastructure, support services and deliver them to users across networks that pay for final computing utility on demand (Ma and Seidman, 2008).

Software vendors often base their pricing strategy from the necessities to cover costs and achieve profit objectives (Harmon et al., 2004). The circular logic of the pricing approach where costs determines price and price influences sales volume causes trouble to the pricing process. According to, Harmon et al. (2004), the key to a long-term success pricing strategy is to have a deep knowledge of the customer, which can result in more appropriate approaches to pricing strategy. The recognition of the price the customer is willing to pay depends on the customer’s value requirements, not the vendor’s is the secret of successful strategy, which is often referred as to value-based pricing.

To gain a better understanding about the market value of a firm’s product, the maximum willingness to pay of the customers and their price sensitivity a controlled experiment\(^1\) also called randomized experiment using real data could lead

\(^1\)An experiment where users are randomly assigned to two variations and their interactions with the site are tracked and analyzed. The users are not aware of the presence of the experiment Kohavi et al. (2012).
This thesis contributes to microeconomic research, more specifically to the ICT sector by using a novel approach with the help of A/B testing on prices. It demonstrates the importance of the role of economists in the private ICT sector. This thesis starts with presenting the technicalities of the SaaS industry and the current pricing methods in the SaaS industry. Subsequently, important economic factors are introduced and discussed. Thereafter, previous researches using controlled experiments are analyzed. Based on previous research, a quantitative case study with help of A/B testing is conducted and empirical evidence from the study is presented and examined.

1.1 Study Objective and Problem statement

The thesis will investigate firstly, how A/B testing can be used as a tool to measure price sensitivity of demand of SaaS products. Secondly, how this type of method can replace educated guesses taken by management for decision-making in regards to prices with pure data driven decision-making.

By the quantitative analysis of this research, this thesis seeks to answer the following research questions:

- How a controlled experiment (A/B test) on prices can lead to valuable findings in regards to a customer’s value perception of the firm’s product, which should be taken into account when a firm chooses a pricing strategy?

- How customer valuation of a product and their price sensitivity plays a vital role in price strategy decisions for managers in the software industry?

The thesis will be tangential to questions regarding optimal pricing and profit maximization for SaaS firms and in what type of market structure SaaS firms participate in.

1.2 Background

Tempo participates in the SaaS industry and is a spinoff from another Icelandic software development company called TM Software (Kamallakharan, n.d.). The

\(^2\)An enterprise created in collaboration with an incumbent firm (Parker, 2009).
later firm focuses on building custom web solutions for customers and uses a product and issue-tracking platform across its firm called, JIRA. The platform was missing necessary features such as accurate time tracking, internal payroll and a billing tool, which made the firm devote a couple of developers to develop these necessary functions. The solution was unique and worked well internally. Due to the fact that there was no similar solution to be found on the market, TM Software launched the product externally a year later with great success. Later on, a decision was made to go from solely developing web solutions to product development; creating Tempo.

Tempo has now over 7,500 customers and has been doubling sales every year from its establishment in 2012 (Kamallakharan, n.d.). It is part of the Atlassian Ecosystem, which serves over 35,000 customers by selling 2,153 different JIRA add-on solutions that add functionality and crucial features to the Jira platform, which is used by IT and business teams of both small and large firms (Atlassian, n.d.). Currently, Tempo has four add-ons on the marketplace and one of them is the most sold add-on. Sellers on the Marketplace are obligated to sell their products via a tier-based-licensing model and receive 75% of the sale, whereas the Marketplace collects 25%. The pricing tiers depend on if the product is a cloud or server product.

For a server product the tiers are split up as table 1.1 shows:

| 10 | 25 | 50 | 100 | 250 | 500 | 1,000 | 10,000 | 10,000+ |

For a cloud product the tiers are split up as 1.2:

| 10 | 15 | 25 | 50 | 100 | 500 | 2,000 |

In this thesis a pricing experiment will only be implemented on the server product.

Although Tempo is a rapidly growing company and has had impressive yearly returns, it is second-guessing its pricing model. The firm has never implemented a pricing experiment prior to this thesis and the management has based their pricing decisions on gut feelings, pricing models of its relative competitors in the
Introduction

market and pricing on JIRA. Its knowledge about their customers and their maximum willingness to pay is limited, but with this research Tempo will get closer to understanding their potential better.

1.3 Purpose and Scope of the study

In this research the relationship between price and quantity demanded will be conducted with the help of controlled web experiment, A/B testing. Having the top selling add-on on the Marketplace, Tempo questions if the demand for the product is high enough so price increases will not have an effect on the demand. The test will gather real time data from the controlled experiment, which could lead Tempo towards more data-driven decision-making in the future.

There is lot of literature to be found regarding optimal pricing and explaining how it depends on market competition and customer price sensitivity. In addition, optimal pricing in microeconomics is dependent on marginal costs, marginal revenue, and price elasticity of demand (Besanko et al., 2013). However, in an industry where marginal cost is very low or close to zero the decision making in regards to price setting gets a whole lot more complex. Firms within the software industry have relatively low marginal cost since the same product can be sold multiple times to multiple customers without any additional cost to produce one more unit. In this case, it is challenging to estimate what is the optimal price. For these types of firms, an important factor is to estimate how much the market values their product to make a relative pricing decisions.

In the past Tempo has sent out surveys to current customers in regards to the usage of their products with a response rate around 2-3%. No real conclusion can be drawn from the surveys due to the fact that it could lead to biased results since Tempo cannot account for multiple factors within the undersized group that responded. Instead of sending out surveys to customers asking them to value the product, a controlled experiment is an option that could eliminate the biased results. Field experiments allow the researcher to build less controlled experiments in exchange for increased realism compared to laboratory experiments (Carpenter et al., 2005). More economists have turned to experimental methods to understand human behavior (Levitt and List, 2009). However, many of these methods were made in the form of laboratory experiments, including volunteers and a controlled
environment. Over time, more economists have explored economic phenomena with the use of field experiments that include randomization in naturally occurring settings with participant subjects unaware of the presence of the experiment. In the literature review, example of these field experiments will be discussed and explained further.

The experiment in this thesis aims to evaluate the price sensitivity of Tempo’s customers by making use of a controlled experimental method, A/B testing. Large international firms such as Microsoft and Google have successfully developed and implemented A/B testing methodologies and systems, but even the smaller firms can benefit from this type of tests (Brodovsky and Rosset, 2011). After the completion of the experiment, Tempo will be able to repeat the experiment again in the future to follow changes in demand for their products. The experiment will also aim to become an important tool for Tempo’s decision making over the years.

The data for this study is gathered through an A/B test experiment, which is used to test the hypothesis of the thesis:

- The null hypothesis: The demand for a particular SaaS product will hold unchanged after a proposed price increase. Hence, there will be no significant difference on conversion rate between the control and the treatment group.

1.4 Limitations

There are some limitations to the experiment due to time constraints, secondly the demand for Tempo’s products, thirdly the medium traffic on the product pages and lastly the limited access to different routes that current and potential customers view the Tempo products.

The field experiment will be implemented and run until a minimum sample size is reached. The experiment will only be done on Tempo’s website visitors who are looking at the Timesheets Server product. The reason why this product was chosen over the others to experiment on is due to the fact that the product-landing page for this particular product is the most viewed landing page out of all the four

---

3The treatment group is the group that is assigned to the variation that is being evaluated and the control group is the group assigned to the existing version.
products, hence potentially the most demanded product. In addition, the management team has discussed a proposal of a price increase for Timesheets Server with additional features. With more website traffic on the product landing page the more likely a quality result will be reached due to the probability to obtain a large enough sample size. This is a drawback, since with that approach we exclude any other visitors who could be looking at other products. This limits the thesis to research the price sensitivity of demand on Timesheets Server rather than the aggregated price sensitivity on all Tempo products.

It is important to note that Tempo’s product-pages are not the only route where potential customers can view Tempo’s products and pricing. As mentioned above, Tempo is part of the Atlassian Marketplace, which lists their products and pricing details. Hence, this experiment will only take into account the current and potential customers that are viewing the Timesheets Server product page and pricing details via the Tempo’s website. This is a limitation since the whole populations of current and potential customers are not being evaluated.

In regards to sustainability factor, this thesis presents a sustainable business practice, which firms can use on continuous bases to follow fluctuations in demand for their products and to get closer to an optimal pricing model. However, there is a risk involved that a customer discovers that he/she is a subject of an A/B test. This might raise confusion and anger towards the firm that is testing. Good preparation of Customer Success/Service Teams is important to deal with these customers gracefully.
Chapter 2

Literature Review

2.1 Definition and Characteristics of SAAS

The SaaS industry is about delivering software solutions remotely to a group of clients via the Internet (Sun et al., 2008). The end-users use credentials from the SaaS vendor to access and use the SaaS solution via the Internet without needing to download or install a software package. The vendor does not only deliver a standalone software solution but also services to the users (Ma, 2007). The vendors store the software system and users’ data and maintain the software, data backups, upgrades and security management. It is easy to implement, easy to update, easily accessible and can be fairly inexpensive compared to other services (Ma and Seidman, 2008). Rather than buying software at a comparatively high price, the user follows a pay-per-use model, which can reduce total cost (Gong et al., 2010). In addition to the service orientation and the accessibility of SaaS products, it helps organizations to cut down capital expenditures and pay for the SaaS product as an operational expenditures (Godse and Mulik, 2009). These might be a few reasons why the SaaS industry is growing; the number of vendors in this industry has increased significantly as well as the number of successful players (Sun et al., 2008).

Additional characteristics of the SaaS industry, according to Sun et al. (2008), is that the SaaS firms take advantages brought by economies of scale by delivering software solutions to a big group of clients over the Internet with one single instance of software application. Economies of scale is achieved when a firm’s average cost decreases as the number of output increases, the marginal cost of the last unit produced must be lower than its average cost (Besanko et al., 2013). “The most common source of economics of scale is the spreading of fixed costs
over an ever-greater volume of output” (Besanko et al., 2013, p.64). Fixed cost can include research and development expenses such as the cost of building new software.

2.2 Theory of Pricing and Output Decisions

According to economic theory, firm’s fundamental goal is to maximize profits and set price as high as it possibly can for a given amount of output (Besanko et al., 2013). Nevertheless, the firm is constrained by the demand curve that it faces, which sets a limit regarding price setting. To recognize firms’ optimal price and output the concept of marginal cost and marginal revenue are practical. When marginal cost equals marginal revenue the firm has reached its optimal level of price and output and is unable to increase profits by changing the output level. An alternative way of thinking about these concepts is to articulate marginal revenue as the price elasticity of demand. This can guide pricing decisions without knowing the firm’s demand curve or marginal cost function.

One of the most important decisions a firm makes when launching a new product is the selection of a pricing strategy. Empirical evidence shows that traditionally, managers within the software industry have developed a cost-based pricing strategy that is highly based on cost-related criteria rather than on the customer’s value perception of the product (Harmon et al., 2004). The logic of cost-based pricing is that price is dependent on the volume and the sales volume is dependent on the price, which makes it difficult for managers to model the price. The key to developing a pricing model specifically for software service is to incorporate customer value and understand the price that the client is able and willing to pay for the service (Kamdar and Orsoni, 2009). One reason why value-based pricing is employed so infrequently is that it features complicated customer specificity, which creates obstacles for marketers.

2.3 Theory of Demand and Price elasticity

A price sensitivity of a customer is obtained from calculating price elasticity of demand for a product, which could possibly give firms a sense of the customers’ value perception. The microeconomic theory of price elasticity measures the percentage change in the quantity demanded per one per cent change in price and measures
how sensitive consumers are to price change (Jehle and Reny, 2011). According to Besanko et al. (2013), price elasticity can be estimated using a statistical approach, but for the most part managers will not have the fortune of a precise numerical estimate of elasticity based on those techniques. Therefore, a manager must rely on his or her knowledge of the product and the nature of the market to estimate price sensitivity. The major determinants of price elasticity according to empirical evidence include the availability of product substitutes, closeness of product substitutes, information in regards of the product and its possible substitutes, the market structure, which the firm participates in, the degree to which the product is considered a necessity and the percentage of buyers' income spent on the product (Pagoulatos and Sorensen, 1986). The greater availability of close substitutes of a particular product the greater the price elasticity of demand should be.

According to Porter, when a firm has a sense for its price elasticity of demand it can decide on its strategy approach in regards to pricing (Besanko et al., 2013). A firm with a high price elasticity could gain market share by lowering prices and implement a share strategy\(^1\) to underprice competitors. On the other hand, if a firm has a low price elasticity of demand, they might benefit from margin strategy\(^2\) by charging higher prices than its competitors. The increased price might allow the firm to exploit an advantage through higher profit margins.

### 2.4 Controlled Experiments on the Web

What if numerical estimates can be achieved by implementing controlled experiments on the web gathering real time data? Controlled experiments on the web are used to make data-driven decisions at multiple large international firms such as at Amazon, Microsoft, eBay, Google, Yahoo, Zynga and other companies (Kohavi et al., 2012). Microsoft has implemented multiple controlled experiments in recent years (Kohavi et al., 2009). One of them was on their MSN Real Estate site where they tested different designs for their ”Find a home” widget. The winning design increased revenues from referrals by almost 10%.

\(^1\) A strategy where a firm makes use of its benefit or cost advantage through a higher market share rather than through high price-cost margins (Besanko et al., 2013).

\(^2\) A strategy where a firm retains price parity with its competitors and profits from its benefit or cost advantage mainly through high price-cost margins, rather than through a higher market share (Besanko et al., 2013).
A different experiment was made on the MSN home page where comparison on page views and clicks was done between population groups to assess if increasing real-estate ads would increase revenue and how it would affect user experience (Kohavi et al., 2008). To assess this metric a monetary value was assigned to page views and clicks. The experiment was only run on 5% of the MSN US home page users for 12 days. The experiment resulted in a statically significant decrease in click-through rate and page views per user-day. If this feature had implemented on the MSN Home Page the expected revenue loss would have been millions of dollars per year.

Additional experiments were carried through on the MSN home page but now in the UK, which included one million visitors that clicked on a Hotmail module over a 16 day period (Kohavi et al., 2009). The experiment wanted to detect if visitor engagement would increase on the MSN home page if Hotmail would open in a new tab/window when clicked on a Hotmail module compared to when it would open in the same tab/window. The experiment resulted in more user engagement when Hotmail opened in a new tab/window and this feature implemented shortly thereafter in the UK and US on the MSN home page.

Controlled experiments can make huge impacts on businesses; this was seen after an experiment was implemented on Amazon’s website. Greg Linden at Amazon implemented a controlled experiment after a marketing senior vice-president did not see the advantages in a personalized recommendation feature that was based on items in the users shopping cart (Kohavi and Longbotham, 2007). The experimental results showed that with Amazon not having this feature live and up and running, was costing the firm a significant amount of money and was launched soon after the results documented.

A field experiment using an A/B testing based on both Poisson and Negative Binomial distribution was implemented on SweetIM’s website to research user optimization experience and usage maximization when testing conversions, search activity and content (Brodovsky and Rosset, 2011). Specifically, the test on conversion; “ratio between successful installs and unique visits on Landing Page” (Brodovsky and Rosset, 2011, p.737) was implemented to examine if changes in user interface on installation dialogue had an effect on the conversion ratio. The results gave highly significant results between groups, which meant that if specific
feature would be implemented, a change in conversion rate of daily installation would occur, more specifically a decrease of 9%. According to the authors, test results can be analyzed as soon as an effective sample size has been reached.

As the above historical examples have shown the controlled experiments can lead to a new hypothesis and improvements by achieving a deeper understanding of variables that influence outcomes form large amounts of data that is typically gained from theses type of experiments (Crook et al., 2009). A/B tests experiments are the simplest form of controlled experiment where users are randomly divided to one of two variants: Control (A) or Treatment (B). The Treatment version is usually a new version being evaluated (Kohavi et al., 2008). "It is important that the users receive a consistent experience throughout the experiment, and this is commonly achieved thought randomization based on user IDs stored in cookies” (Kohavi et al., 2008, p.150). The null hypothesis in an A/B test is stated as there is no difference between variations and if observed it would be due to chance. With A/B testing a null hypothesis can be tested by collecting data on metrics of interest and then using statistical techniques to evaluate if a tactical significant difference is to be found between the two groups on metrics of interest (Kohavi et al., 2008).

2.5 Economic Field Experiments on prices

Field experiments in the private sector has not been fully explored even though the sector includes central economic issues that can benefit from field experimentation, such as price setting. The limitation is that the naturally occurring data are often unavailable and requires partnership with firms participating in the private sector (Levitt and List, 2009). Field experiments in relation to prices have been tested by using direct mails and by using door-to-door salespeople. A field experiment to test prices with real time data with the use of A/B testing has not been undertaken before, as far as my research goes.

Anderson and Simester (2003) tested the effect of $9 price endings on retail sales via a field experiment. The authors used the method of direct mail and sent out a retail catalog with different stated prices to randomly selected customers and followed their behavior. The empirical evidence shows that the power of $9 had a positive effect and produced a higher number of quantity sold than other prices.
The authors Ashraf et al. (2007) used a different method of field experimentation, when they tested the impact of product prices on purchase and usage by using door-to-door salespeople. The consumer was randomly quoted a price for a product, either a lower or a higher price. After they purchased, a survey was sent out and a test was done in regards to the usage of the product. The empirical evidence shows, those who were willing to pay more appeared to value the good more highly.
Chapter 3

Empirical Model, Methodology and Data

3.1 Empirical Model

This section will discuss and explain the empirical model by Kohavi et al. (2009), which the experiment of this thesis will be following. Since the experiment will be looking for changes in conversion rate between the two variations the experiment can be modeled, as a Bernoulli trial, which will be explained further below. Thereafter, implementation architecture to conduct a controlled experiment introduced also by Kohavi et al. (2009) will be explained thoroughly.

3.1.1 Recommendations

The model recommends the experimenter to define the Overall Evaluation Criterion (OEC), which also can be called the evaluation metric (Kohavi et al., 2009). The well-defined metric can then be compared between the two groups after an A/B test is completed. Kohavi et al. (2009) recommends implementing an A/A test where users are assigned to two different groups, but will view the exact same website variation. The test can be used to gather data and measure variability and to test the experimentation system. Thereafter, implement a single factor A/B experiment with two very different variations, assign one half of the users to a control variant and the other half to the treatment group. The authors also suggested to slowly increase the fraction of users assigned to the treatment group. Additional recommendations are to be aware of the day-of-the-week effect and lastly to calculate the minimum sample size.
3.1.2 Preparing an A/B Test

The importance of planning a controlled experiment is vital (Brodovsky and Rosset, 2011). Each experiment should start with defining the sample size; hence calculating how long the test will run for. Thereafter, a statistical test can be done to evaluate whether the treatment population is different from the control population (Kohavi et al., 2009). To reject the null hypothesis; the metrics of interest will differ between groups, the alternative hypothesis is accepted; the treatment population is statistically significantly different from the control population.

Two formulas are defined as useful in the empirical model. Firstly the t-test:

\[ t = \frac{\bar{O}_B - \bar{O}_A}{\sigma^2} \]

Where \( \bar{O}_A \) is estimated conversion rate (baseline) for the control group and \( \bar{O}_B \) is the estimated conversion rate for the treatment group. The standard deviation of the difference between the two rates is written as \( \sigma^2 \)

Secondly, a formula to calculate sample size:

\[ n = \frac{16\sigma^2}{\Delta^2} \]

The formula above assumes desired confidence level \(^1\) is 95% and desired power \(^2\) is 80%. Where n is the sample size in each variation group. The figure 16 is obtained from the Z score of alpha and beta. The standard deviation is presented as \( \sigma^2 \) and the difference in the mean or the sensitivity one wants to detect is represented as \( \Delta^2 \). When detecting percent change in conversion rate a Bernoulli trial\(^3\) model can be used:

\[ n = \frac{16\sqrt{p(1-p)}}{\Delta^2} \]

Where the figure 16 represents the respective Z-scores for alpha and beta for desired 95% confidence level and desired 80% power hence, 1.96 and .84 multiplied by two.

\[ 2 \times (Z_\alpha + Z_\beta)^2 \rightarrow \]

\[ 2 \times (1.96 + .84)^2 \approx 16 \]

\(^1\)Type I error- Probability of finding an effect that is not there (Kohavi et al., 2008).

\(^2\)Type II error- Probability of finding an effect that is there (Kohavi et al., 2008).

\(^3\)An experiment where n trials are made, with probability p of success in any given trial (Kohavi et al., 2008).
The standard deviation of a Bernoulli is a measure of variability written as:

$$\sqrt{\bar{p}(1 - \bar{p})}$$

Where \(\bar{p}\) is the probability of success. Calculated:

$$\bar{p} = \frac{p_1 + p_2}{2}$$

Where \(p_1\) and \(p_2\) is the probability of success in each variation group.

Therefore, the sample size in each group is presented as \(n\) assuming that the groups are at equal size and calculated as:

$$n = \frac{2 \times \bar{p}(1 - \bar{p})(Z_{\beta} + Z_{\alpha})^2}{(p_1 - p_2)^2}$$

Where the probability in one group is presented as \(p_1\) and \(p_2\), hence, the difference in proportions is \((p_1 - p_2)^2\)

### 3.1.3 Implementation Architecture

Implementing an experiment on a website involves three components; a randomization algorithm to map users to a variation, an assignment method to determine the users’ web experience and a data path that collects the data from the website and applies statistical techniques prepares reports (Kohavi et al., 2009).

A good randomization algorithm has to have three properties; splits users evenly over population groups, assigns users to the same variant on each visit to the site, and if multiple experiments are being made there should be no correlations between experiments. Techniques that satisfy the three properties are pseudorandom with a caching method and hash and partition method. The first accomplishes the three properties by caching on the client side by storing a users’ assignment in a cookie. The later one assigns a unique identifier each experiment and to each user that is maintained either through a database or a cookie. Then a hash function is used to combine the identifiers to obtain an integer that is distributed on a range of values. In this thesis the first technique will be used simply because the software used to implement the A/B test has this technique integrated. In addition, (Kohavi et al., 2009) mentions that the hash and partition method is more sensitive than the pseudorandom with catching method and therefore recommends using the first technique.
There are several assignment methods available for A/B testing such as traffic splitting, page rewriting, client-side assignment and server-side assignment. According to Kohavi et al. (2009), an assignment method is the piece of software that allows the website being used for the experiment to show different code paths for different users. In this thesis a client-side assignment is used, again simply because the software used for the A/B test uses this technique. The client-side method is the most common assignment technique found in a third-party experimentation platforms and defined to be very easy to implement. It is a method that does not require a large involvement of and IT assistance, all the developer needs to do is to add snippet of JavaScript to a page. “That code instructs the end user’s browser to invoke an assignment service at render time” (p.168). Then the service call returns the correct version to the user, and triggers a JavaScript callback that tells the browser to modify the page presented to the user (see figure 3.1).

![Figure 3.1: Page client-side assignment method](image)

To compare results between population groups a good data path is essential to secure collection of data, page views, clicks etc. The system will then need to convert the data into metrics to compare the treatment and control groups. This can be done by event-triggered filtering, raw data collection, using existing (external) data collection, local data collection or by service-based collection. This thesis will be using the service-based data collection method, which implements a service that is designed to record and store observation data from the website. This method has the advantage of consolidating all the observation data, which makes it easier for analysing. Kohavi et al. (2009) states that the service-based collection method is the most flexible and preferred method.
3.2 Methodology

The A/B test for this thesis will be examining the difference in conversion rates from website visitors who are assigned two different landing pages that show different prices. To count as a successful conversion the visitors that view a landing page have to click on the “Free Trial” button. As stated above the thesis will be using three types of methods to implement the experiment; pseudorandom with caching randomization algorithm method, client-side assignment method and service-based collection data path method.

3.2.1 Phase 1-A/A Test

An A/A test will be implemented on Tempo’s pricing pages for their most sold product, Timesheets (Server), with the use of the software, Optimizely. The test will have two variations, treatment page and the control page, that will be randomly distributed to the subject, the website visitors with using the client-side assignment method. The two variations are identical with matching prices (see figure 3.2).

![Flow chart of A/A test](image)

**Figure 3.2:** Flow chart of A/A test

The software will track the conversions for five defined goals. However, the main goal and evaluation metric of our interest is the conversion rate on how many
website visitors click on the “Free Trial” button, located above the pricing card. This particular button is the most clicked button compared to other buttons on the product page. Below are the five listed goals:

- The conversion rate on “Free Trial” located on the top of the page
- The conversion rate on “Try it Now” located below the pricing card
- The conversion rate on “Try for Free” located on the bottom of the page
- The conversion rate on “Try for free” located above the pricing card
- Total engagement on the product-pricing page

Optimizely will track the two variations on the product-pricing page using service-based collection data paths. The A/A test is implemented to detect if the users are distributed equally over the two population groups, to calculate the variability, to test the experiment’s system and finally if the data flows into the software correctly and as expected and to receive a current conversion rate baseline that will be used later on to compare the A/B test conversion rates to.

Optimizely uses pseudorandom with caching method as a randomization algorithm to map variations to users and to account for unique visitors. When a new website visitor enters the product-pricing page, Optimizely will assign him/her a cookie if they are grouped in the treatment group. The cookie will expire after 10 years, which will allow us to see conversion rates from unique visitors only (Optimizely, n.d.a). Optimizely, will run on Tempo’s product-pricing page where one line of JavaScript code is loaded to change the experience of treatment group visitors but not the controlled group visitors, with using the client-side assignment method (Optimizely, n.d.b).

The software collects data on the page visitors and conversions and runs it through a statistical framework to declare the outperforming variation, it uses the service-based collection method as a data path. If a visitor enters the product-pricing page and clicks on the “Free Trial” button, this activity will be accounted as a conversion. However, if the same visitors enters the page again during the experimental period and repeats the same action it will not be accounted for in the experiment; only if the visitor has cleared their cookies in between their visits.
To know if the test is set up correctly the distribution of the variations should be 50/50 and no detectable difference in conversion rates should be observed between the two population groups. If they differ slightly it is due to chance that more people click on the “Free Trial” button for one variation and cannot be stated as statistical significantly different.

3.2.2 Phase 2-Effective Sample Size

After gathering a baseline from the A/A test the effective sample size is calculated with using the minimum sample size formula presented in the empirical model above.

However, the formula is adapted slightly due to the fact that this thesis is detecting differences in proportions rather than means using the Bernoullii method. In addition, the confidence level will be relaxed to 90% due to time limitation but the desired power of 80% will be kept unchanged. Hence, Z-score for beta will equals .84 and Z-score for alpha 1.28 for a one tail test (for Z-score table see A).

\[
n = \frac{2 \times \bar{p}(1 - \bar{p})(Z_{\beta} + Z_{\alpha})^2}{(p_1 - p_2)^2}
\]

The sample size in each group is presented as n assuming that the groups are of equal size. The probability of success in one group is presented as \(p_1\) and \(p_2\), hence, the difference in proportions is:

\[(p_1 - p_2)^2\]

A measure of variability, which is similar to standard deviation, is written as:

\[(\bar{p})(1 - \bar{p})\]

Where \(\bar{p}\) is the probability of success. Calculated:

\[\bar{p} = \frac{p_1 + p_2}{2}\]

The baseline conversion rate, \(p_1\), obtained from the A/A test is used to calculate \(\bar{p}\). As well as \(p_2\), which is calculated by multiplying \(p_1\) with the desired power 0.8. The thesis assumes a one tailed alternative due to the fact that the interest lies in learning whether the treatment variation results in lower (worse) conversion rates than the control variation.
To make sure the calculations are correct a proportions power calculation is done using the software R (for details see B and C).

The R method:

```r
power.prop.test(p1=0.07685, p2=0.06148, power=0.8, alternative='one.sided', sig.level=0.1)
```

Two-sample comparisons of proportions power calculation

```r
n =
p1 = 0.07685
p2 = 0.06148
sig.level = 0.1
power = 0.8
alternative = one.sided
NOTE: n is number in *each* group
```

### 3.2.3 Phase 3-A/B Test

After having calculated the minimum sample size for the experiment, an A/B test will be implemented on Tempo’s pricing pages with the use of the same software and method as the A/A test. With the help of the software Optimizely, two different pricing pages will be created and distributed randomly and evenly to the website visitors (see figure 3.3).

![Figure 3.3: Appearance of variations](image)

Half of the group will be assigned the treatment page, which will state higher prices and the other half will be assigned the controlled page, which will state today’s current price. The only variant that will differ from the two pages will
be the stated price of the product and all other factors will be kept constant (see figure 3.4).

![Flow chart of A/B test](image-url)

The price that the treatment group will see is the future price, which Tempo is planning to charge for Timesheets Server. It is important to note that no website visitor will be charged the higher price, the two groups will only view the different prices. The percentage price increase differs across tiers (see table 3.1). The price for the lowest user tier license will be held constant but the largest license tier price will increase by 25%. The highest percentage change in prices will occur on the third highest tier, 2,000-User license tier, which will have an increase of 38%. The table below presents the old and new prices, where in the experiment the old prices will be used on the controlled group and the new prices will be used on the treatment group.

<table>
<thead>
<tr>
<th>Tier</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>500</th>
<th>2000</th>
<th>10000</th>
<th>10000+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Price</td>
<td>$10</td>
<td>$600</td>
<td>$1100</td>
<td>$2000</td>
<td>$4000</td>
<td>$6000</td>
<td>$8000</td>
<td>$12000</td>
<td>$16000</td>
</tr>
<tr>
<td>New Price</td>
<td>$10</td>
<td>$650</td>
<td>$1250</td>
<td>$2400</td>
<td>$5000</td>
<td>$8000</td>
<td>$11000</td>
<td>$16000</td>
<td>$20000</td>
</tr>
<tr>
<td>Increase</td>
<td>0%</td>
<td>8%</td>
<td>14%</td>
<td>20%</td>
<td>25%</td>
<td>33%</td>
<td>38%</td>
<td>33%</td>
<td>25%</td>
</tr>
</tbody>
</table>

**Table 3.1: Differences in prices between groups**
The A/B test is planned to run for six to eight weeks or when the minimum sample size has been reached. After the experiment, real data will be received from the software that will show the conversion rates; how many visitors from the treatment group clicked on the “Free Trial” button and how many visitors clicked on that same button from the control group. Historically, most of Tempo’s customers do an evaluation of the product before buying it, which is the reason a measure of demand is done by looking at the successful conversion of the visitors who click on the button “Free Trial” rather than “Buy It”. When signing up for an evaluation license, one has to enclose detailed information such as credit card information to receive a license. Post evaluation license’s end-date, the credit card used when signing up is charged for a regular customer license if the evaluation license is not canceled before the specified end-date. This approach, eliminates trial customers from the purchasing process that do not have a pure interest in the product.

The test is implemented to get a sense of a customer’s price sensitivity on Tempo’s most popular product. The data will be used to calculate the change in demand and price sensitivity of Tempo’s customers. It is important to keep in mind that the test is not made to test two different sets of prices and pick the price that leads to more profits.

### 3.3 Data

The software Optimizely collects the data for this thesis by using a service-based data path method. Thereafter, Optimizely runs it through a statistical framework to declare the outperforming variation. The software converts the raw data collected from the website experiment into metrics that can be compared between the Treatment and Control groups. In addition to the experiment data, historical and current data on evaluations will be used to detect if any difference in signup rates occurred on Timesheets Server during the experimental period compared to the historical rates.
Chapter 4

Empirical Results

Two A/A tests had to be implemented due to the fact that a bug was captured during the first attempt of the test. A very low conversion rate appeared for one of the goals, which led to the bug discovery, that was connected to wrongly tracked events. The second attempt of the A/A test allowed for gathering a conversion rate baseline, which was used as a comparison for the A/B test conversion rate. When the baseline for all the goals had been gathered a calculation for the minimum sample size was calculated as follows:

\[
2449.70 = \frac{2 \times 0.069165(1 - 0.069165)(0.84 + 1.28)^2}{(0.07685 - 0.06148)^2}
\]

The minimum sample size is calculated to be 2450 in each variation. To make sure the calculations were correct, a proportions power calculation was computed using the software R:

```r
power.prop.test(p1=0.07685, p2=0.06148, power=0.8, alternative='one.sided', sig.level=0.1)
```

Two-sample comparison of proportions power calculation

n = 2456.145
p1 = 0.07685
p2 = 0.06148
sig.level = 0.1
power = 0.8
alternative = one.sided

The alternative method gave similar results but differed by about 6 (for more
Empirical Results

Details see C). The sample size computed with the statistical formula presented in the empirical model was determined to be the minimum effective sample size in the experiment.

After clarifying the minimum sample size in each variation an A/B test was set up and ran. The test was implemented just prior to the Easter weekend, which could have been the reason why a drop in demand occurred during the first week of the test for both variations (see 4.1). The orange line represents the conversion rate for the treatment group and the blue line represents the conversion rate for the control group.

![Figure 4.1: A/B Test Result Graph](image)

Optimizely, gathered data and created conversion metrics from the collection to be able to compare between the two variations. To reach the minimum effective sample size the experiment had to be run for 48 days. During those days, the Customer Success Team of Tempo did not receive one complaint from website visitors in regards to the A/B test. After the test was completed, the treatment and the control group were statistically analyzed to examine whether the effects between the groups differed.

<table>
<thead>
<tr>
<th></th>
<th>Product Page Visits</th>
<th>Clicks on &quot;Free Trial&quot;</th>
<th>Conversion Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td>2.528</td>
<td>152</td>
<td>6.01%</td>
</tr>
<tr>
<td>Treatment Group</td>
<td>2.476</td>
<td>156</td>
<td>6.30%</td>
</tr>
<tr>
<td>Difference %</td>
<td>0.5%</td>
<td>2.6%</td>
<td>0.29%</td>
</tr>
</tbody>
</table>

The results from the A/B test showed a slightly higher conversion rate for the treatment group, which was assigned the website showing higher prices (see table
Empirical Results

4.1. This was an interesting result and not expected. However, the results showed no statistical significance difference between the two variations, which tells us that it could have been due to chance that more successful conversions occurred in the treatment group. This can be translated to accepting the null hypothesis; the demand for a particular SaaS product will hold unchanged after the proposed price increase. Hence, there will be no significant difference on conversion rate between the control and treatment group.

This might be a surprising result at first because an assumption could have been made for the group viewing prices that differed by up to 38% that it would make statistically significantly fewer successful conversions compared to the visitors viewing lower prices. On the contrary, when looking into the industry which Tempo participates in and their early mover advantages, this result could have been assumed.

The economic theory states that monopolistic competition is a form of imperfect competitive market structure with two main features (Besanko et al., 2013). Firstly, the market involves many sellers where their actions do not materially affect others. Secondly, the sellers offer a differentiated product, where consumers make purchasing choices among competing products on other factors than just price. Both those factors apply to the SaaS industry, which Tempo participates in. In addition, monopolistic competition characterizes of a powerful competitive dynamic due to the fact that optimal pricing for firms participating in differentiated product markets imply setting prices above marginal costs. This results in firms earning positive economic profits, which pushes under market entry that results in loss of market share and profits for incumbents. The competitive structure of a firm’s market is important to recognize to be able to lead the firm towards profit maximization and a correct pricing model (Jehle and Reny, 2011).

Empirical evidence shows that when developing a pricing model for SaaS firms an understanding in regards to market type plays a significant role. For oligopolistic pricing, game theory may be the main approach. On the other hand, game theory may not be applicable in a monopolistically competitive model where each firm could be defined as a monopolist with demand dependent on prices of other products. As software applications of the same type are to some extent different from each other, each application is associated with a unique demand curve, which
Empirical Results

best characterizes the SaaS market and the definition of monopolistic competition (Xie, 2011).

In addition to the market structure, early mover advantages can be a reason why a statistically significant difference between the two variations was not detected. Tempo can be described as the early mover since it introduced the first accurate time tracking product, Timesheets Server, to the JIRA add-on marketplace. The advantages of being an early mover in an industry include building a reputation and limiting buyer uncertainty, increase buyers switching costs and creating a network effect Besanko et al. (2013). If a firm has brand awareness and positive customer experiences, price increase might not have a big effect on current or potential customers.

In regards to current customers, when price increase is put into place the risk of losing current customers might not be as high as the risk of attracting new ones (Besanko et al., 2013). The theory of switching cost, describes when a current customer has developed a brand specific ”know-how” in regards to a product or a system that is not fully transferable to substitute brands. Due to that fact, a customer might be less attracted to switch to a new brand.

Continuously, it could also be the case that increasing prices will not affect current nor potential customers if a company in a specific market has a great virtual network effect. A virtual network effect, or indirect network effect as it is also called, takes place when an increase in number of consumers in a network results in increasing demand for complementary goods. It is tangential to high-tech products, where demand for the infrastructure depends on the availability of applications and vice versa (Varian, 2001). A good example of this could be a computer operating system, where, with more users the more features or other complimentary products will be demanded. This will increase supply for complementary goods, which will consequently add value to the network (Besanko et al., 2013). Hence, the firm that is first to build a large customer base will have an advantage. Current customers will raise supply for complementary products and new customers will detect the large customer base and get attracted to the same firm. “SaaS platforms exhibit indirect network effects to the extent that the popularity of one platform over another with developers makes the platform more attractive to other
developers or users” (Cusumano, 2010, p.28). With more applications on a platform the more attractive it gets.

There are possibly additional factors, which could potentially be the reason why the higher price did not result in lower demand in the experiment such as searching cost and the timing of the experiment but that goes beyond the scope of this paper.

To note, the breakdown of the experimented visitors did not match the set traffic percentage split precisely. When the sample size was reached there was 0.5% more visitors assigned to the controlled group compared to the treatment group. Optimizely explains this by when one denotes the assignment percentage for each group, they are specifying the probability that any given visitor will be placed in that particular group (Optimizely, n.d.b). As a result, the number of visitor in each group might not add up exactly with the denoted percentage splits one has specified.

When looking at Tempo’s historical Timesheet Server evaluation data over the past three years for the same time period as the controlled experiment was implemented this year, March 23rd to May 10th, an increase is detected in total number of evaluation signups for 2016 (see figure 4.2).

![Figure 4.2: Total number of evaluation signups across years over the same time period](image)

At first, this might come of as the experiment had positive effect on evaluation signups rates. However, when looking at the first four months separately for each individual year a trend can be observed. The number of evaluation signups are consistently higher for the month of January than for the month of April the same
Empirical Results

year, for past three years (see figures 4.3, 4.4 and 4.5).

**Figure 4.3:** Monthly evaluation signups for 2014

**Figure 4.4:** Monthly evaluation signups for 2015

**Figure 4.5:** Monthly evaluation signups for 2016

When taking a deeper look at the graph for 2016, a peak in evaluation signups for the month of March is not visible like it is for both 2014 and 2015. Nevertheless, the total evaluation signups for the month of March in 2016 is higher than for the month of March in 2014 and 2015. The first four months of 2016 beats the monthly evaluation signup number for the first four months in both 2014 and 2015.
The data results show interesting trends and shifts in evaluation signup rates over the years and months. But without statistical significant results these results could simply have happened due to chance and without a specific reasons. It is difficult to define the steep decline in evaluation signups for the month of March in 2016 compared to past years without any significant results.
Chapter 5

Conclusion

A/B testing can lead to valuable findings with regards to price sensitivity of customers. The thesis focused on an A/B test, which was implemented on Tempo’s website to evaluate if price increases on a SaaS product will affect demand, hence the price sensitivity of a customer.

Some past research by Kohavi et al. (2009) and Kohavi et al. (2008) used A/B testing on the web to evaluate users experience design. Other studies by, Kohavi et al. (2009), Kohavi et al. (2008), Kohavi and Longbotham (2007) and Brodovsky and Rosset (2011) used A/B testing to new features. Other research have used controlled experiments on prices via direct email such as Anderson and Simester (2003) and Ashraf et al. (2007) by using door-to door salespeople to detect any changes in usage or purchase. Multiple large companies use controlled experiments on the web to make data-driven decisions, which Tempo management is now able to do as well.

The experiment was implemented to gain insight on how customers will react to higher prices on a new version of Timesheets Server that will be released in the near future. The experiment led to valuable findings and showed no statistical significant difference in conversion rates between the treatment and control group though the two groups viewed prices that differed up to 38%. The observed difference in conversion rates between the two groups was 0.29% higer conversion rates among the treatment group. This can be translated to accepting the null hypothesis; the demand for a particular SaaS product will hold unchanged after the proposed price increase. Hence, Tempo can assume low price sensitivity among their website visitors on the experimented product up to a certain point
and can make confident pricing decisions respectively. This can result in pricing decisions that make Tempo closer to its optimal pricing model than before and potentially increase profits. In addition to that, this thesis contributes to microeconomic research by demonstrating the importance of the role of economists in the private ICT sector. Also showing, that by having an expert that understands the importance of demand and prices and how to apply microeconomic theories to real practices is very valuable.

As mentioned above, limitations were found in this research due to time, the demand for Tempo’s products, the medium traffic on the product pages and lastly the limited access to different routes that potential customers view the products. If one had access to all possible routes a potential SaaS customer could view a product, hence, had the opportunity to capture all potential customers for a particular product, it would be interesting for future research to implement an A/B test on prices and evaluate price sensitivity. Another future research proposal would be, if one had access to an e-commerce website and able to track every unique website visitor by assigning them an anonymous ID, it would enrich an A/B test further. The researcher would then be able to track if a visitor was placed in a treatment or controlled variation group and if that particular visitor signed up for an evaluation license or not.

The experiment of the thesis resulted in valuable findings in regards to Tempo’s customers reaction to price changes, their price sensitivity and a data driven decision to increase prices for the upcoming product release. In addition, as stated before this thesis contributes to studies within microeconomics and more specifically the ICT sector.
Appendix A

Z-Score Table

![Z-Score Table Diagram]

Table entry for \( z \) is the area under the standard normal curve to the left of \( z \).
Appendix B

R-Method: Power Prop Test Calculation

```r
> power.prop.test(p1=0.07685, p2=0.06148, power=0.8, alternative='one.sided', sig.level=0.1)
Two-sample comparison of proportions power calculation

  n = 2456.145
  p1 = 0.07685
  p2 = 0.06148
  sig.level = 0.1
  power = 0.8
  alternative = one.sided

NOTE: n is number in *each* group
```
Appendix C

Details for R-Method Sample Size Calculation

**Description** Compute the power of the two-sample test for proportions, or determine parameters to obtain a target power.

**Usage**

```
power.prop.test(n = NULL, p1 = NULL, p2 = NULL, sig.level = 0.05, power = NULL, alternative = c("two.sided", "one.sided"), strict = FALSE, tol =
```

**Arguments**

- **n**: number of observations (per group)
- **P1**: probability in one group
- **P2**: probability in other group
- **Sig.level**: significance level (Type I error probability)
- **Power**: power of test (1 minus Type II error probability)
- **Alternative**: one- or two-sided test. Can be abbreviated.
- **Strict**: use strict interpretation in two-sided case
- **Tol**: numerical tolerance used in root finding, the default providing (at least) four significant digits.
Details Exactly one of the parameters n, p1, p2, power, and sig.level must be passed as NULL, and that parameter is determined from the others. Notice that sig.level has a non-NULL default so NULL must be explicitly passed if you want it computed.

If strict = TRUE is used, the power will include the probability of rejection in the opposite direction of the true effect, in the two-sided case. Without this the power will be half the significance level if the true difference is zero.

Value

Object of class "power.htest", a list of the arguments (including the computed one) augmented with method and note elements.

Note

uniroot is used to solve power equation for unknowns, so you may see errors from it, notably about inability to bracket the root when invalid arguments are given. If one of them is computed p1 \( \neq \) p2 will hold, although this is not enforced when both are specified.

Author(s)

Peter Dalgaard. Based on previous work by Claus Ekstrøm (Dalgaard and Ekstrøm, n.d.)
Appendix D

Definition of A/B Testing

A/B tests experiments are the simplest form of controlled experiment where users are randomly divided to one of two variants: Control (A) or Treatment (B). The Treatment version is usually a new version being evaluated (Kohavi et al., 2008). The web offers a great opportunity to assess ideas promptly using controlled experiments, also called A/B, Control/Treatment tests, split tests, MultiVariable Tests (MVT) and parallel flights. Controlled experiments represent the most effective scientific design for determining a causal relationship between changes and their effect on user-observable behavior.
Bibliography


**URL:** http://link.springer.com/article/10.1023/A:1023581927405


Atlassian (n.d.), ‘Software development and collaboration tools — atlassian’.

**URL:** https://www.atlassian.com/(Accessed: 2016-05-10)


Crook, T., Frasca, B., Kohavi, R. and Longbotham, R. (2009), ‘Seven pitfalls to avoid when running controlled experiments on the web’, *Proceedings of the 15th* ...

**URL:** http://dx.doi.org/10.1145/1557019.1557139


**URL:** http://portal.acm.org/citation.cfm?doid=1721654.1721667


Kamallakharan, B. (n.d.), ‘Intrepreneurship in iceland-the story of tempo’.


URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-69649095419&partnerID=tZOtx3y1


URL: http://dl.acm.org/citation.cfm?doid=2339530.2339653


   URL: http://www.nber.org/papers/w16852


Optimizely (n.d.a), ‘How optimizely works: Snippet order of execution, javascript evaluation timing, and cookies’.

Optimizely (n.d.b), ‘Traffic allocation: Limit experiment traffic or launch winning variations’.
   URL: https://help.optimizely.com/TargetYourVisitors/TrafficAllocation%3Alimit,experiment,traffic,or,launch,winning,or,win,variants (Accessed: 2016 – 05 – 10)


URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.29.2323 &rep=rep1&type=pdf