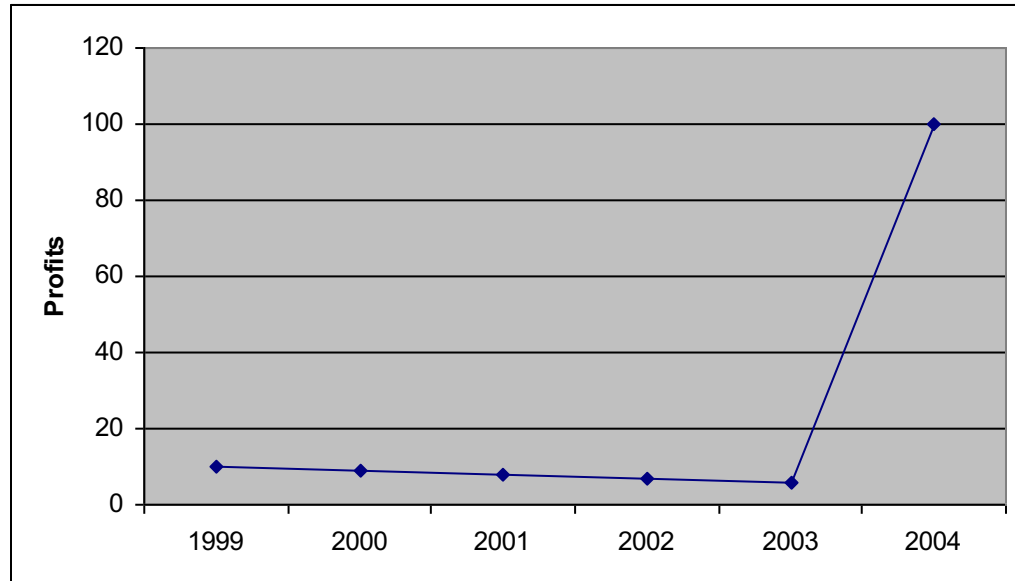


Marketing: Predicting Success



Maikel Groenewoud
1155601
Business Mathematics and Informatics
Paper
May 2004

Preface

In the last year of the study Business Mathematics and Informatics each student writes a paper based on own research. Each student is free in the choice of the subject as long as there is a clear link to the study. Special thanks goes out to dr. ir. G. Jongbloed, my supervisor, who gave me valuable advice while writing this paper.

Maikel Groenewoud

Table of content

Introduction	p. 4
Marketing Basics	p. 5
Consumer Behaviour	p. 7
Marketing Research	p. 10
Data Analysis in Theory	p. 12
Data Analysis in Practice	p. 14
Conclusion/ Summary	p. 18
Literature	p. 19
Appendix: R	p. 20

Introduction

A big threat to businesses is a lack of information or even worse incomplete or incorrect information. Without all the facts decisions are made and risks and opportunities are assessed based on the wrong information. A big problem is that information is often scattered within different archive systems that are not connected to each other. There is also a lack of awareness about statistical and analytical tools and their potential for strategic decision making.

In this paper I describe a way of analyzing consumer characteristics to gain valuable information that can support managers in the decision process concerning the marketing strategy and for the prediction of marketing success. This support is given by determining market segments, finding most likely (new) customers etc. This type of information is known as Market(ing) Intelligence, which is a form of Business Intelligence (BI). The term intelligence is used because collected data is processed and given meaning/ turned into information.

A firm's relationship with its customers can be described by the following phases:

Suspects → Prospects → Customers → Regulars → Ambassadors

- Find possible customers → Suspects
- Select most likely future customers from these suspects → Prospects
- Some prospects have delivered hits → Customers
- (Some) customers have become frequent users → Regulars
- Customers (almost) actively help with Marketing → Ambassadors

In the last phase they love the products/ services so much that they even recommend them to others. That is very important because new products are initially usually only adopted by a few consumers and only gradually spread to the rest of the population. This means that companies that introduce new products must be well financed so that they can stay afloat until their products become a commercial success. So it is important to please initial customers because they in turn might influence many potential new customers.

There is a chapter dedicated to Consumer Behaviour because this is a very important phenomenon in Marketing but in the past there has not been enough attention for it. There is also a chapter dedicated to Marketing Research and two to Data Analysis. Data Analysis is a step in the Marketing Research process and certain techniques used to analyze data are described in the chapter Data Analysis in Theory. The second chapter about Data Analysis shows how it is used in practice by examining some real life data using the software package R.

Marketing Basics

In this chapter some basics from the field of Marketing will be discussed. This chapter is partly based on the book Principles of Marketing by Philip Kotler & Gary Armstrong. Here is an overview:

- Definition of Marketing
- Marketing concept
- Marketing strategy
- Market forces
- Segmentation

Since this chapter and in fact this whole paper deals with Marketing, it is a good idea to get some clarity about what Marketing really is all about. There are many different definitions of Marketing but one that is often used is that of the American Marketing Association (AMA). This definition states the following: “Marketing is the process of all activities necessary for the conception, pricing, promotion and distribution of ideas, goods and services to create exchanges that satisfy individual and organizational objectives.” This definition shows that there are a lot of things that can be associated with Marketing. That is not a surprise because Marketing is a very broad field. The principle however is quite simple: “How to get people to buy/ use your products and services.”

A fundamental part of Marketing theory is the Marketing concept. This is a business philosophy that says that the key to achieving organizational goals consists of the company being more effective than its competitors in creating, delivering and communicating customer value to the chosen target markets. So the whole idea is that to achieve all of your goals, you have to be better in what you do than your competitors. This concept is intuitively clear which makes it quite easy to understand.

A Marketing strategy consists of selecting a segment of the market as the company’s target market and designing the proper mix of the marketing-instruments product/service, price, promotion and distribution system to meet the wants and needs of the consumers within the target market. This mix is known as the marketing-mix. It is vital to establish a Marketing strategy because it gives direction to a firm’s activities.

There are several forces in the market that can be analyzed:

Market forces

- own firm (strengths and weaknesses)
- competing firms (strengths and weaknesses)
- economic conditions (recession, legislation etc.)
- consumers

The focus of this paper is upon the consumers.

The last subject that will be discussed in this chapter is segmentation. This basically involves dividing consumers into groups or segments such that members of each group are as similar as possible to members of that same group but differ as much as possible from members of other groups. This enables a more specific targeting of people in each segment by doing things such as:

- Providing different products (e.g. some consumers like cola while others prefer tea)
- Offering different prices (e.g. some consumers will take the cheapest product available, while others will pay extra money for special desired features)
- Distributing the products where they are most likely to be bought by the targeted segment
- Direct mailing to most likely (new) customers

In an ideal situation a segment would only contain one consumer. This would enable a company to fully customize its products/ services for each consumer individually. In practice a situation like this does not occur because it would be way too expensive to target customers individually. That is why consumers are usually always clustered in groups but the aim is to make those groups or segments as diverse as possible from each other, as long as this process remains cost-effective.

Consumer Behaviour

Markets can be segmented in a number of ways but there are also different levels of segmentation:

Levels of segmentation

- Personal characteristics (Who are they?)
 - Demographics
 - Psychographics
- Desired benefits (What do they want?)
- Behaviour (How do they act?)

This chapter deals with the last level of segmentation → Behaviour.

The information provided in this chapter is partly based on the book Consumer Behaviour by Solomon, Bamossy and Askegaard. Psychology is a field that has for a long time been ignored when considering economics. That is why special attention is paid to it because it is very important to have a good understanding of Consumer Psychology and Behaviour when trying to market products and services. Traditional economics and marketing view people as rational decision makers. From our own experiences however it is obvious that often this is not the case. When deciding to buy a certain product or make use of some service we often make very irrational decisions influenced by many different factors. Something that also should be taken into account is the way we process information and the fact that we are limited in our capacity to process this information. This is especially of importance for the way advertisements are designed.

As already made clear this chapter deals with consumer behaviour. One official definition of consumer behaviour is: "The study of individuals, groups, or organizations and the processes they use to select, secure, use, and dispose of products, services, experiences, or ideas to satisfy needs and the impacts that these processes have on the consumer and society." This definition seems to go a little further than what already has been discussed so far. Not only does it say that individuals need to be studied but you should also look at groups and organizations. But considering the fact that groups and organizations are built up out of individuals, it still comes down to the study of individual behaviour. What this definition emphasizes is that individuals are influenced by other individuals they encounter in groups, organizations etc.

When studying consumer behaviour one would like to know how consumers make purchasing decisions. The rational decision process can be modelled in many different ways but they all come down to the following:

Steps in the rational decision process

- problem recognition
- information search for solutions
- evaluation of alternatives
- purchase

When evaluating alternatives, rational decision makers consider factors directly related to the product or service such as:

Influences on the decision process

- price
- quality
- availability
- (additional) service

This is quite easily comprehended but the simplicity is misleading. Consumers are also influenced by a great deal of other and often non-rational factors that should also be taken into account such as:

Other influences on the decision process

- (sub-)culture
- family & friends
- media
- physical factors (hunger, thirst etc.)
- added value (enhanced self-image, social status etc.)
- experiences/ knowledge
- age or maturity
- reference or admired groups (professional athletes etc.)
- affect or emotional appeal
- mood
- personality
- information processing (capacity, perception, attention etc.)

Most of these factors are not directly related to the quality or functionality of a product or service but they nonetheless have great influence on the decision process and significantly raise the complexity of the model when taken into account. This means more work but in the end it is worth it because this extra work helps develop a better Marketing strategy and that eventually leads to higher profits. Marketers can adapt and

improve their Marketing campaigns to reach the consumers more effectively now. For example, the understanding that consumers are more receptive to food advertising when they are hungry teaches marketers to schedule snack advertisements late in the afternoon.

Marketing Research

The previous chapter discussed consumers, their characteristics and their behaviour. It also discussed the importance of this kind of information for a firm's Marketing strategy. Now that the importance of truly knowing and understanding consumers is clear, the next step is to get the information that will enable you to get that greater insight into those consumers. This can be done by conducting Marketing Research. This type of research involves the process of designing, gathering, analyzing and reporting information that may be used to solve a specific Marketing problem. Another way of viewing Marketing Research is by looking at it as a mechanism to link consumers to the marketer.

The information provided in this chapter is partly based on the book Marketing Research by Burns, Alvin, Bush & Ronald. Good Marketing Research is built up by clearly defined steps to insure that the results will be satisfactory. Here is an example of a good way of breaking this research process down into some smaller steps:

Steps in the Marketing Research process

- 1 Establishing the need for marketing research
- 2 Defining the problem
- 3 Establishing research objectives
- 4 Determining research design
- 5 Identifying information types and sources
- 6 Determining methods of accessing data
- 7 Designing data collection forms
- 8 Determining the sample plan and size
- 9 Collecting data
- 10 Analyzing data
- 11 Preparing and presenting the final research report

Not all of these steps are going to be discussed in detail because that is not the focus of this paper. The steps that will be especially emphasized are those of specific interest from a marketing, statistical or analytical point of view.

The second step of the Marketing Research process is defining the problem. The problem needs to be correctly defined to set the right direction for all subsequent steps of the marketing research process. If this is not done it usually doesn't matter how well the results in those other steps might be, because in general it can be said that the end-result won't be satisfactory.

From a statistical point of view the sample plan and size are of special interest. The sample plan refers to the process used to select units from the population to be included in the sample. This determines how representative the sample is for the population. The sample size refers to how many elements of the population should be included in the sample. Usually it's so that the larger the sample the better, because it will be more

representative then. However, in some cases it is possible to have a sample that is too large. If for example the aim is to predict the behaviour of the entire population based on a small group of people, it wouldn't make any sense to select half of the population. Errors can be made when choosing a sample but there are also so-called non-sampling errors such as selecting the wrong group of people to interview. For example, the objective is to study the behaviour of internet users but the people selected never use the internet.

There are also some alternative ways for getting the data or information as can be obtained through Marketing Research. Direct marketing and customer tracking and personalization tools are two of these. Direct marketing offers exceptional opportunities for getting personal characteristics for segmentation because marketers can buy lists of consumer names, addresses, and phone-numbers that indicate their specific interests. These lists contain a lot of data but have to be further analyzed to get relevant information. Legislation concerning privacy issues can be a problem however because it could place certain restrictions on obtaining consumer data in this way. Using customer tracking and personalization tools on the internet is a truly alternative way of performing marketing research. These tools collect and store data on website visitors. The difference with the lists you buy for direct marketing is, that they also analyze this data and use the obtained information to enable one-to-one marketing. This means that web pages are especially customized for each consumer.

The following two chapters deal with how to analyze the data gathered through Marketing Research.

Data Analysis in Theory

Data collection is very important because no matter what kind of data analysis methods are used, bad data can't be "fixed". Once data is collected, data analysis is used to give meaning to the raw data. However, before this can take place the following needs to be done first:

- entering data into computer files
- inspecting data for errors
- getting data in right format

Data analysis is the tenth step in the Marketing Research process as given in the last chapter, it involves:

- searching for patterns
- performing tabulations
- performing statistical tests.

There are basically two approaches to data analysis:

- Bottom-up analysis (exploratory and data-driven)
- Top-down analysis (confirmatory and model-driven)

Bottom-up analysis is where the user looks for useful information previously unnoticed, by searching through the data for ways of connecting it to create hypotheses. Top-down analysis aims to confirm or reject hypotheses and tries to widen the knowledge of a phenomenon. In reality the two approaches are complementary. The information obtained from a bottom-up analysis, which identifies important relations and tendencies, cannot explain why these discoveries are useful and to what extent they are valid. The tools of top-down analysis can be used to confirm the discoveries and evaluate the quality of decisions based on those discoveries.

Various kinds of analytical techniques can be used to examine Marketing data. These methods include Chi-squared Automatic Interaction Detector (CHAID) trees and logistic regression. CHAID is a multi-variate technique which splits a sample according to the relationship between the response variable in the analysis and a set of predictors or explanatory variables. Splitting is a procedure based on the concept of reducing heterogeneity in the response distribution by separating cases into two disjoint subsets that are more homogeneous when they are separated than when combined. So it gives support when determining the number and composition of market segments. The output comes in the form of a tree diagram.

Logistic regression is a form of generalized linear regression. It describes the relationship between a dichotomous/ binary response variable and a set of explanatory variables. The explanatory variables may be continuous or discrete. In a Marketing context this

technique is of interest because the analysis of consumer data makes it possible to predict the value of the response variable for other consumers. So it may lead to potential new customers. If for example a company is preparing a direct mail campaign, it would need to select people to send the mail to. The logistic regression model can be used to select which customers are most likely to respond. A manager would like to see that the selection based on the logistic regression model leads to a higher response than a random selection. Otherwise the analysis of data and construction of the model would have been a waste of money.

Logistic regression assumes a function of the form:

$$p = \frac{\text{Exp}(\eta)}{(1 + \text{Exp}(\eta))}$$

with η representing the linear predictor:

$$\eta = c_0 + c_1 x_1 + c_2 x_2 + \dots + c_{n-1} x_{n-1} + c_n x_n$$

and $y \sim \text{bern}(p)$

The explanatory variables are represented by $x_1, x_2, \dots, x_{n-1}, x_n$ and the model finds the values of the constant coefficients $c_0, c_1, c_2, \dots, c_{n-1}, c_n$ etc, where c_0 is also called the intercept. The response variable is represented by y and has a so-called Bernoulli distribution with parameter p . A Bernoulli distribution is used because y can only assume two possible values. The parameter p gives information about the probability that y assumes either one of these two values. The next chapter illustrates what data analysis is like in practice.

Data Analysis in Practice

This chapter discusses the results of the analysis of some Marketing data. The aim was not to give a full (statistical) data analysis but to show the value and possibilities for managers. A dataset was examined containing a wide range of consumer characteristics. These characteristics ranged from product usage data about various insurances to things such as age, income, religion, marital status, customer type etc. The analysis took place by using the software package called R. R is a graphical environment that can be used to perform statistical analysis. It's a more powerful version of the package known as Splus. In the appendix the functions used for the data analysis can be found.

Once the data was in the right format, exploratory research could be conducted. Before performing an in depth analysis of the data, it is advisable to conduct some preliminary exploratory research. When performing Exploratory Research it is a good idea to approach the data as if there is not any prior (domain) knowledge available about it. This can reduce possible bias towards the data and may lead to patterns that otherwise might have been ignored. The dimension of a dataset gives the number of rows and columns. In this particular case there were 10000 rows and 86 columns. Each row corresponds to a single customer and the columns to his or her attributes. From now on the attributes will be referred to as variables. All the data was numeric, but some variables contained coded values. The actual values were alphanumeric but they were replaced by numeric values.

Calculating the minimum, maximum, median, mean and variance of all the variables can often quickly give some insight about the data being examined. For example, consider the binary variable that indicates whether people have a caravan policy or not:

Minimum	0
Maximum	1
Mean	0.0594
Variance	0.05587723

This shows that the mean of this variable is 0.0594. Because it is a binary variable this means that $0.0594 * 10000 = 594$ people have a caravan policy. That is only about 6%, so there are much more consumers in the dataset who do not have a caravan policy than there are who do have one. The average customer in this dataset belongs to age group 3, which stands for ages ranging from 40 to 50. It would not make any sense to discuss all the 86 variables here but it is clear how useful this type of information can be. Graphs can be used to compare two variables with each other and they quickly show if there's any direct evidence of a relationship or interdependence between them. The correlation also quickly shows whether there is any direct evidence of a relationship or interdependence between two variables. This was the case with the variables representing a customer's main and subtype. What's special about them is that they are coded variables. When plotting these two variables against each other it looks like there is a direct linear dependence between them. This could suggest that they weren't just randomly coded but also sorted in an ascending order by certain criteria. The correlation between them is

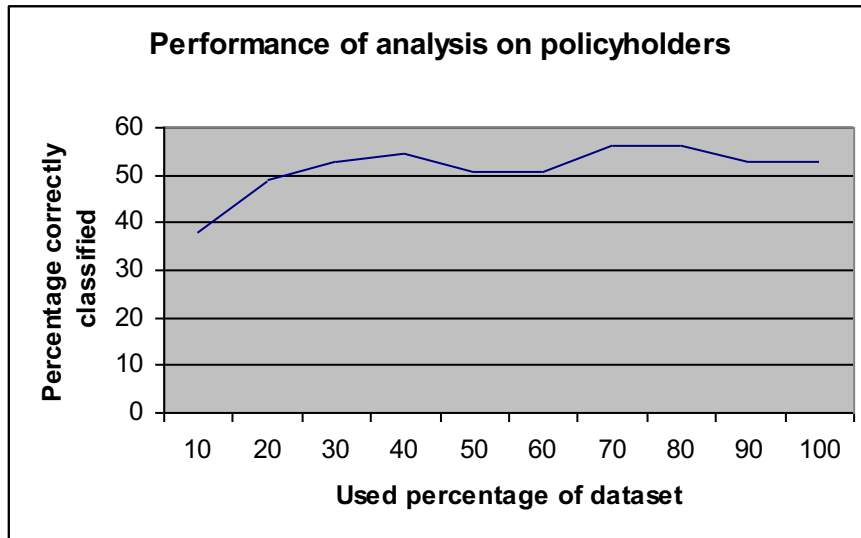
0.9926198. This strongly suggests a dependence between the two variables because the absolute value is very close to one. If the absolute value was approximately equal to zero there probably wouldn't be a dependence between these two variables. Since these variables are so closely related, it probably wouldn't make any sense to have them both in the model. When one of these is already included, the other doesn't really add any predictive value.

After all the exploratory research had taken place, the data was analyzed more extensively by building a logistic regression model. The model was constructed by taking the variable representing whether consumers have a caravan policy or not as response variable and the other variables as explanatory variables. It was just a matter of choice because the other variables could of course also have been taken as response variables. It depends on for what purpose the model will be used. By splitting the data into two smaller datasets, one can be used to build the model and the other to test its predictive capabilities. The original dataset was split into one dataset containing 9000 consumers and another containing 1000 consumers. The first one was used for building the model and the other to test its predictive ability. The logistic regression model determined the coefficients $c_0, c_1, c_2, \dots, c_{86}$ of the explanatory variables x_1, x_2, \dots, x_{86} . In general it cannot be said that the variables with the biggest coefficients have the biggest influence on the response variable. If that was the case the next step might then be to build a new model with only the explanatory variables with relatively big coefficients and determine if the model still fits the data well. Other things have to be considered too however such as the scale and range of the variables.

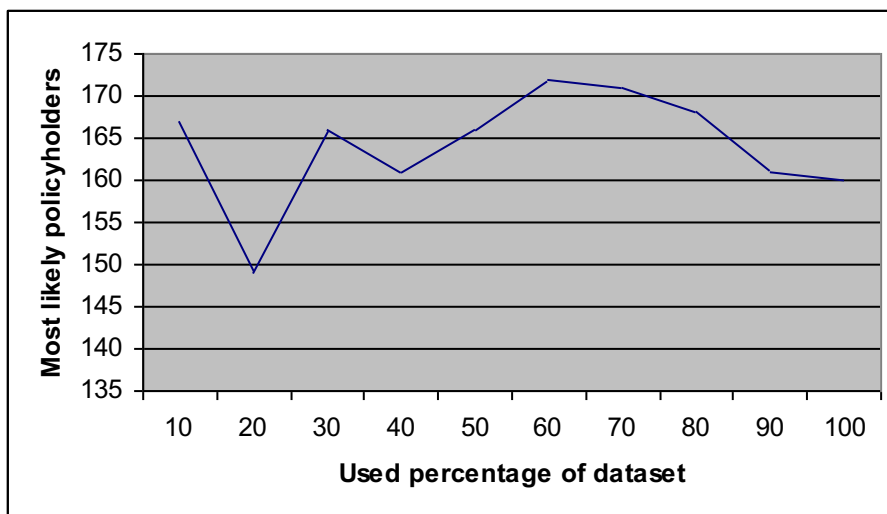
The dataset was also examined after leaving out the coded variables discussed earlier to see if that would make any difference. The conclusion was that omitting these variables is not of significant influence. This is probably so because they were only two variables out of a total 86 and their relative importance might not be so big. Earlier results about the correlation already showed that they are so closely related that their predictive value can almost be viewed as that of only one variable.

There were also some variables containing categorical data, such as the variable discussed earlier indicating to which age group a person belongs. The intervals of these variables were coded by numeric values in an ascending order. So the lowest interval had the smallest value etc. These variables did not influence the model in a negative way.

Graphical representations can be used to evaluate the performance of a model. The dataset used for testing contained 55 consumers with a caravan policy, that's only 5.5 %. Here are the classification results:

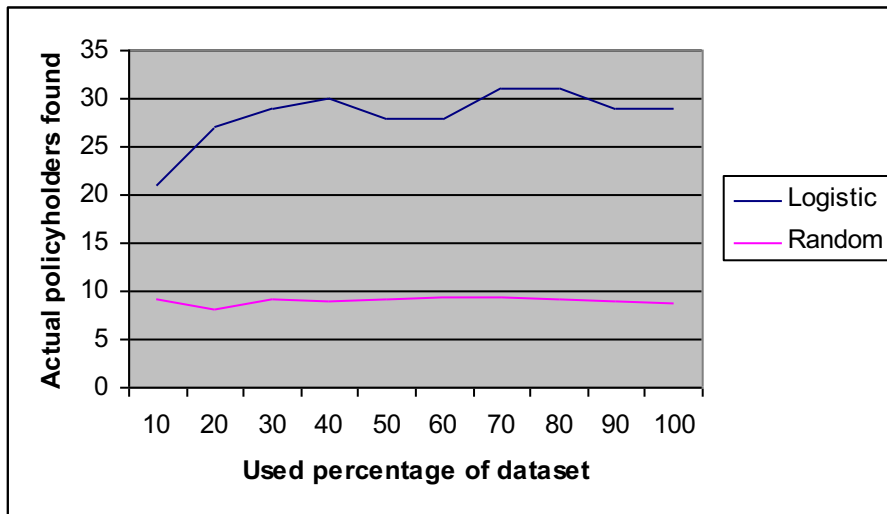


This graph shows that even when only using a small percentage of the dataset for constructing the model, a relatively high percentage of the caravan policyholders was correctly classified. This means that a company can save money by collecting or buying data about a smaller number of customers without that having a negative effect on the predictive strength of the model. On average about 50 percent of the actual policyholders was correctly classified. The following graph shows the numbers of most likely policyholders selected:

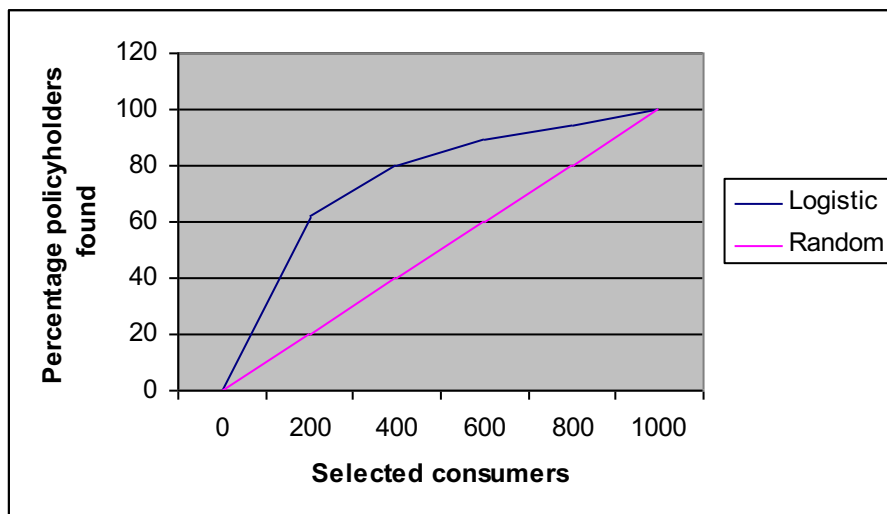


This shows that on average about 160 consumers were selected every time.

The next graph shows the number of policyholders found by using the logistic regression model and by random selection:



This clearly shows that the actual number of policyholders found by using the logistic regression model is much higher than when randomly selecting consumers. On average the results are about three times better. The following graph also clearly shows that the logistic regression gives much better results than a random selection:



These last two chapters have illustrated some of the techniques and approaches that can be used for data analysis and how valuable they can be. More can be used and more statistical research can be done but the primary goal was to make clear how valuable data analysis can be for managers and their companies.

Conclusion/ Summary

The consumer decision process is influenced by many factors, which should be taken into account when determining a Marketing strategy. Analyzing consumer characteristics can give valuable insight when determining market segments and most likely (new) customers. Software such as R can be used to analyze this data and the results can be used to support managers when making strategic Marketing decisions. My overall conclusion is: “A greater understanding allows better segmentation which enables a more specific targeting which leads to higher profits!” This means that data analysis of consumer characteristics such as consumer behaviour, will give managers and their companies a strategic and financial advantage. Data analysis can be used in Marketing as a tool to predict success.

Literature

- Principles of Marketing, Philip Kotler & Gary Armstrong
- Solomon, Bamossy and Askegaard (2002), Consumer Behaviour; A European Perspective. Harlow: Pearson Education (2nd edition)
- Marketing Research - Burns, Alvin C.; Bush, Ronald F.
- Gleitman, H., Fridlund, A.J., & Reisberg, D. (1999). Psychology. Fifth Edition. New York: Norton.
- Dix, J. et al., (1998). Human-Computer Interaction (2nd edition). Prentice Hall.
- Attention' by Harold Pashler, Psychology Press Ltd (1998)
- Ashcraft, M.H. (2002) Cognition. New Jersey, Prentice-Hall Inc
- Data Mining by Ian H. Witten and Eibe Frank, Morgan Kaufmann Publishers
- Applied Data Mining, Paolo Giudici, 2003 John Wiley & Sons, Ltd
- Fitting Models to Biological Data using Linear and Nonlinear Regression, by Harvey Motulsky & Arthur Christopoulos
- Management Information Systems by Laudon and Laudon (7th edition)
- Statistical Computing: An Introduction to Data Analysis using S-Plus, Michael J. Crawley, Wiley (2002)
- The Basics of S and S-PLUS (Third edition), Andreas Krause and Melvin Olson, Springer-Verlag New York (2002)
- An Introduction to R, A Programming Environment for Data Analysis and Graphics, Version 1.8.1 (2003-11-21), W. N. Venables, D. M. Smith and the R Development Core Team

Appendix: R

This Appendix gives an overview of the R functions used for data analysis while writing this paper.

Data preparation

- `read.table()`

(Exploratory) Research

- `dim()`
- `range()`
- `median()`
- `mean()`
- `variance()`
- `summary()`
- `plot()`
- `pairs()`
- `cor()`

Logistic regression

- `glm()`
- `coefficients()` or `coef()`
- `fitted.values()`
- `predict()`

The clear names of these functions make it much easier to understand the data analysis process in software packages such as R.