IBEXI

REFLECTIONS

Solving Insurance Business Problems Using Statistical Methods

Anup Cheriyan

Table of Contents

Executive Summary

This paper deals with the use of statistical methods for solving insurance business problems. It explains why statistical methods are well suited for insurance, and identifies some business problems that can be solved using these approaches.

It introduces common statistical methods used in solving business problems. Statistical methods like descriptive statistics, inferential statistics, hypothesis testing, chi square analysis of frequencies, strength of association and scatter plots are described.

Finally, some specific business problems are discussed. These include insurance agent attrition, productivity of insurance agents and high lapse ratio. For each problem, the impact and cost to the business are explained. Then some approaches to solve the problem, the data required and the statistical method that can be employed are described.

About the Author

Anup Cheriyan is Associate consultant with Ibexi Solutions. He has experience in implementation of Business Intelligence solutions for insurance companies in India.

Introduction

Decision making processes must be based on data, not on personal opinion or on belief. In the real world decisions need to be made even though there are uncertainties. Business statistics is a science assisting managers to make business decisions under uncertainties based on some numerical and measurable scales. It is used in many disciplines, such as financial analysis, production, operations and marketing research.

Business statistics is used to make inferences about certain characteristics of a population based on information contained in a random sample from the entire population. This enables managers to:

- Solve problems in a diversity of contexts.
- Add substance to decisions.
- Minimize guesswork.

The insurance business is rich in data, and in data complexity. Retail insurance business (e.g. life and motor) typically has a large number of clients and policies. Mature companies in many countries have millions of policies. In emerging countries such as India, even new companies less than 5 years old have a million clients; older large companies e.g. LIC have over 130 million policies. Insurance policies have a large amount of data, and they are complex in structure, with variations such as benefits, face amounts, schemes, pricing, claims, multiple client relationships, medical history and family history and underwriting. This combination of volume and complexity is unusual; this makes it difficult to manually understand the data, and its trends. Thus, the insurance business is ideally suited for the application of statistical methods.

Currently, use of statistics is largely limited to actuaries for determination of the insurance premium rates. Statistics can have wide applications in other departments of an insurance company. For instance, the agency department can use statistical methods for combating high agent attrition rates and hiring productive agents. Also, the marketing department can use statistics to identify target customers for cross selling a new insurance policy.

Common statistical methods

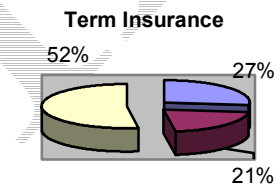
Before we venture into the applications, let us review common statistical methods like descriptive statistics, inferential statistics, hypothesis testing, chi-square analysis of frequencies, scatter plots, simple linear regression, multiple regression and non-linear regression, and their direct uses. Descriptive statistics are used to summarize or describe the population sample, whereas inferential statistics are used to test theories about the population and for generalizing the behaviour of a population sample.

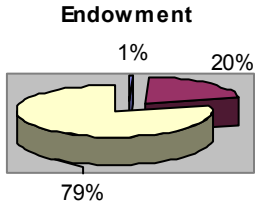
1. Descriptive statistics

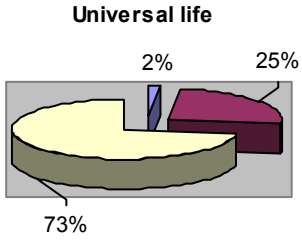
Descriptive statistics are used to describe the shape of a sample, where the data tend to cluster, and how the data are dispersed. This type of analysis is supported by most of the OLAP tools available in the market, and can be used to get insights into the business.

<table>
<tr><th>Type of analysis</th><th>Example</th></tr>
<tr><td>Central Tendency, using

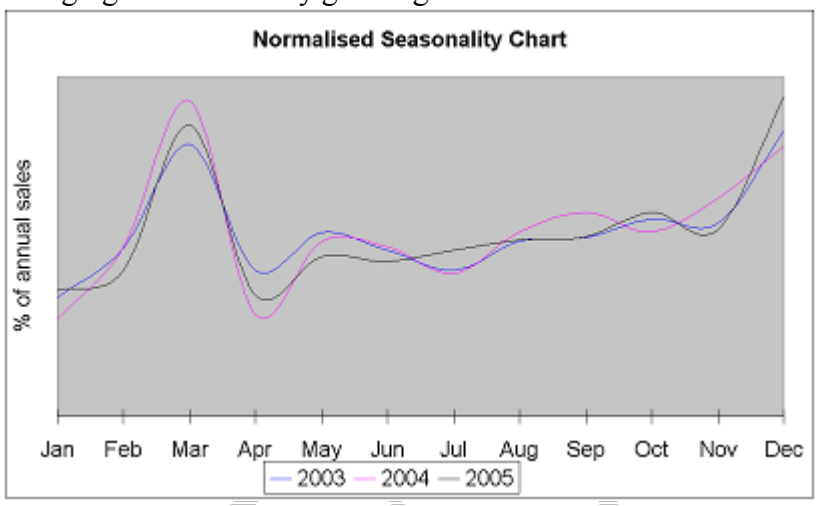
❑ Mode
❑ Median
❑ Mean</td><td>What is the number of policies per month sold?
<table>
<tr><th>Month</th><th>Number of policies sold</th></tr>
<tr><td>January</td><td>295</td></tr>
<tr><td>February</td><td>599</td></tr>
<tr><td>March</td><td>630</td></tr>
<tr><td>April</td><td>509</td></tr>
<tr><td>May</td><td>465</td></tr>
<tr><td>June</td><td>600</td></tr>
<tr><td>July</td><td>309</td></tr>
<tr><td>August</td><td>630</td></tr>
<tr><td>September</td><td>645</td></tr>
<tr><td>October</td><td>500</td></tr>
<tr><td>November</td><td>307</td></tr>
<tr><td>December</td><td>780</td></tr>
</table>
Mode=630;
Median=554;
Average=522
This reveals that we sell 522 policies per month, on an annual average; however, as the median is 554, which is greater than the average, it indicates that the lower-volume months are pulling down the performance more sharply than can be made up by the higher-selling months.</td></tr>
<tr><td>Distribution of Data, using

❑ Frequency Distribution
❑ Histogram</td><td>What is the distribution of policies by product class, assured gender?
Illustration:
<table>
<tr><th>Product / Gender</th><th>Company</th><th>Female</th><th>Male</th></tr>
<tr><td>Level Term Insurance</td><td>32,525</td><td>25,296</td><td>63,295</td></tr>
<tr><td>Pure Endowment</td><td>2,090</td><td>56,092</td><td>217,092</td></tr>
<tr><td>Universal Life</td><td>490</td><td>7,267</td><td>20,880</td></tr>
</table>






The distribution data indicate a difference in the purchase pattern for pure term insurance among genders, as opposed to investment products like endowment and universal life.</td></tr>
</table>

Variability, using ❑ **Variance** ❑ **Standard Deviation** ❑ **Coefficient of variation**	What is the seasonal variance across months in a year? Is this trend changing? Is seasonality growing with time?  In this example, the standard deviation for monthly sales for each year can indicate whether the seasonality is increasing or decreasing.

2. Inferential statistics

Inferential statistics involves a decision-making process that allows us to objectively quantify if treatment effects are significant or not. The success of this process requires that we make certain assumptions about how well the sample represents the larger population. These assumptions are based on two important concepts of statistical reasoning: probability and sampling error. A statistical comparison of sample parameters is required to determine if the difference is "true" or "false": such a comparison is called hypothesis testing. The chi-square statistic is a nonparametric statistical technique used to determine if a distribution of observed frequencies differs from theoretical expected frequencies. Various strengths of association are also discussed in this section.

Analysis	Statistical method	Example
Hypothesis testing	❑ Hypothesis testing ❑ Chi Square test ❑ Simulation	Is the increase in sales an outcome of ad campaign?
Strength of association	❑ Correlation coefficient ❑ Linear regression ❑ Multiple regression ❑ Non-linear regression	Is there a correlation between ad spending and sales volume?
Segmentation	❑ Cluster analysis ❑ Scatter plots	Which customer segment buys maximum number of whole life policies?

Data Sources

Collection of relevant data is prerequisite for employing statistical methods for the same. Data required can be categorized into primary and secondary data depending upon the data source.

Primary data – fresh data collected by the researcher himself
Secondary data – data already collected by others or available internally within the company, which will be "re-used"
Secondary data may be internal to the firm, such as sales invoices and warranty cards, or may be external to the firm such as published data or commercially available data. The government census is a valuable source of secondary data.

It is possible to categorise secondary data based on source and degree structure. This typically indicates the ease of using the data. Internal data is more likely to be complete and available easily; external data like those from research bureaus and government are often not complete for the analysis. Also, structured data is more amenable to analysis than unstructured data.

Examples of secondary data in different categories

DATA SOURCE	DATA STRUCTURE: *Structured*	DATA STRUCTURE: *Unstructured*
External	IMRB, Census	Interest rates, Tax changes, Regulatory changes, Product launched by competition
Internal	Typically from databases and transactional systems	Launch of new products by the company, new partnership agreements

Secondary data has the advantage of saving time and reducing data gathering costs. However, there are limitations of secondary data. These include:

- It is collected for a different purpose
- There might be problem of definitions
- Secondary data will have problem of comparability over time
- Lack of awareness of sources of error/bias in data is a limitation.

Therefore secondary data may not fit the problem perfectly. Secondary data alone is not sufficient to analyse business problems. Secondary data must be supplemented by primary data originated specifically for the study at hand. Primary data can be obtained by initiating a market research.

Some common types of primary data are:

- Demographic and socio-economic characteristics
- Psychological and lifestyle characteristics
- Attitudes and opinions
- Awareness and knowledge - for example, brand awareness, insurance awareness
- Intentions - for example, purchase intentions. While useful, intentions are not a reliable indication of actual future behaviour.
- Motivation - a person's motives are more stable than his/her behaviour, so motive is a better predictor of future behaviour than is past behaviour.

Use Of Statistics For Specific Business Problems

In insurance, there are many business problems in different areas that can be tackled using statistical approaches. These include:

Agency department:

- Agency force attrition
- Insurance agent productivity and agent success factors

Renewals department:

- High lapse in the initial years of the policy

Marketing & sales department:

- Identification of customer segment for cross selling
- Features to be added to a new product and understanding customer needs
- Identifying gaps in product mix
- Identifying gaps in product mix
- Customer segmentation

Actuarial department:

- Enhancing product profitability

Underwriting department:

- Fine tuning of Auto underwriting methods

Operations department:

- Reducing turnaround times of New business and policy owner servicing processes
- Fraud detection patterns
- Enhancing product profitability

To solve a real-world business problem, it is necessary to use a combination of standard statistical methods, a combination of tools and data sources.

1. Insurance Agent Attrition

High attrition rate of insurance agents is one of the biggest challenges for an insurance company. Conservative estimates put the attrition rates at 35-40 per cent.

For new insurance companies still struggling to break even, the rising attrition rate is yet another challenge that they have to battle. For mature companies too, the attrition rate especially in the face of rising competition is a growing threat.
Agents leave company because of various reasons like:

- Competition offers more remuneration
- Company does not having good products to offer
- Attrition of sales managers
- Lack of proper training resulting in unsuccessful sales career
- Unsuccessful in adapting to sales
- Unsuccessful in adapting to insurance
- High work pressure and targets.

Solving this problem involves finding reasons resulting in high attrition rate, and identifying actions, which can address these reasons. Business questions that would answer this include

- Are most of those who drop out non-performers?
- Can adequate training increase performance of agents? Will this help in reducing attrition?
- Will introducing new products which appeal to the customers better, solve the problem of high attrition?
- Has booming economy caused the rampant poaching of insurance agents? Will it settle down once industry has their expanding networks in place?

To answer the business questions, some of the data that is required include:

- Historical data of existing agents and terminated agents: Agent profile, number of policies sold, first premium collected, renewal premium collected
- Training period of agents in the company
- Remuneration and reward models for agency force within the company
- Agents monthly and annual targets within the company
- Exit interviews of agent who leave the company

- Average training period of agents in the industry
- Remuneration and reward models for agency force of competitors
- Agents monthly and annual targets for competitors

Note that the last three are external data, and may be incomplete.

Recommended actions that can be taken to solve this problem include:
- Adjust the hiring pattern to match demographic and psychographic patterns
- Training effectiveness and duration can be changed if required
- Build a predictive agent attrition model to determine agents likelihood of attrition

Illustration: Predictive agent attrition model

Predictive model is an equation used to predict a variable. Statistical methods like regression analysis can be used to develop predictive models.

The solution involves the use of past data of company's agents as well as data about terminated agents. This data can used to develop a predictive model that would assign a score to each and every agent. This score would indicate the agent's propensity to leave in a defined timeline. Input data for modelling included key demographic and behavioural information.

Agents could now be scored for their likelihood of attrition. The attrition scores, used in conjunction with the expected lifetime value of the agent can business managers with a continuous trade-off framework will help decide on offers to negotiate with agents in a bid to retain them.

2. Productivity Of Insurance Agents

New people are ultimately joining the insurance industry; these days an image makeover as "life insurance advisors" is in fashion. But little has changed in the basic nature of the business - insurance still needs to be sold to a reluctant populace. Sustenance requires constant networking and acquiring new relationships for business.

In a business such as insurance one has to accept the fact that 20 per cent of the work force will bring in 80 per cent of their business.

Agent productivity depends on factor like:
- Agents' age, gender
- Educational qualification
- Work experience
- Product knowledge (training received)
- Selling skills
- Motivation

Solving this problem will involve identifying factors that results in high productivity of agents.

- Can adequate training increase performance of agents?
- Are older agents showing better performance than younger ones?
- Is higher educational qualification positively or negatively correlated with performance?
- Are female agents more productive than male agents or wise versa?
- Does an agent with work experience perform better than a new agent? Is it wise to poach insurance agents from competitors?

To answer the business questions, some of the data that is required include:
Internal data:

- Training period of agents in the company
- Remuneration and reward models for agency force within the company
- Work experience of agents
- Number of policies sold, first premium collected, renewal premium collected for agents.
- Educational qualification of agents
- Agent details (age, gender, etc)

External data:

- Average training period of agents in the industry
- Remuneration and reward models for agency force of competitors

Illustration: Insurance agent performance and age

Did you ever wonder whether the insurance agents performance depends on his age? Strength of such a relationship between performance and age can be quantified by using correlation coefficient. The correlation coefficient summarizes the relationship between two variables

It might be that the agents who are aged have more experience, and they sell more policies. This would be an example of a positive correlation, because high values of one variable (e.g., agents age) are associated with high values on the other variable (e.g., better performance on sales). Or it might be the other way around: age of person is associated with poorer sales performance. The latter is an example of a negative correlation, because high values on one variable are associated with low values on another variable.

Similarly strength of association of other factors like agents training duration, educational qualifications, etc on agent performance can be determined by using correlation coefficient.

3. Lapse Ratio & NTU Of Policies

Lapse ratio is calculated by dividing number of policies lapsed by total number of policies. It is a ratio used to measure the effectiveness of an insurer's renewal strategy. A lower lapse

ratio is better, particularly because insurance companies pay high commissions to brokers and agents bringing new clients.

Higher lapse ratio and NTU can be due to various reasons like:

- Not providing timely and accurate (e.g. notice to wrong address) reminders to customer for renewal, results in higher Lapse ratio.
- Churning. This is a major contributor to the high lapse and surrender rates. Churning is the practice of encouraging customer to cancel an existing policy to take out a new policy or another investment. What many policyholders do not realize is that this results in double cost: penalties apply in most cases where a policy is cancelled and then a whole new round of costs are incurred on the new investment. One of the major reasons for churning is the commission structure of the life insurance industry. Currently, in the case of most life assurance products, commissions are paid upfront, with 35 to 40% of the commission paid in the first year of the policy. As a result financial advisor will encourage customer to switch investments to earn a living.
- Pressure sales tactics by agents, resulting in purchase of policies for higher premiums than can be continually paid by the client during renewal.

Business questions that may help address this problem are:

- Is there a correlation between lapse ratio and customers annual income, especially if the premium is worked at as a percentage of their income?
- Does certain product characteristic of policy result in higher lapse and NTU?
- Is there a specific demographic profile of customers who NTU & lapse their policies?
- Do some channels contribute significantly more to high lapsed and NTU?
- Are there group of agents who contribute to higher lapse ratios?
- Does policies issued at end of a quarter have more probability of lapsing than policies issued at other times?

To answer the business questions, some of the data that is required include:

Internal data:

- Remuneration and reward models for agency force within the company
- Agent targets and evaluation parameters
- Agent wise data of number of policies issued, Inforced, lapsed and NTU
- Agent details (age, gender, etc)

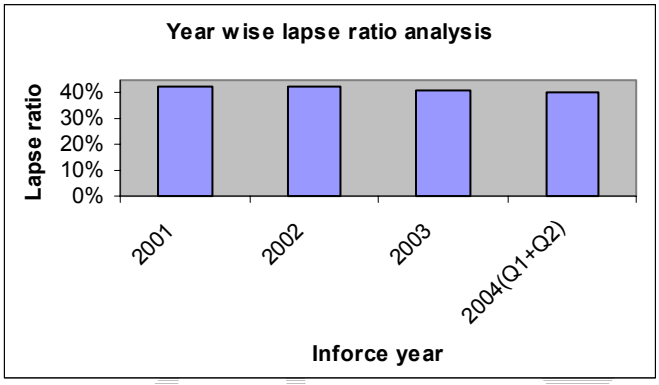
External data:

- Remuneration and reward models for agency force of competitors
-

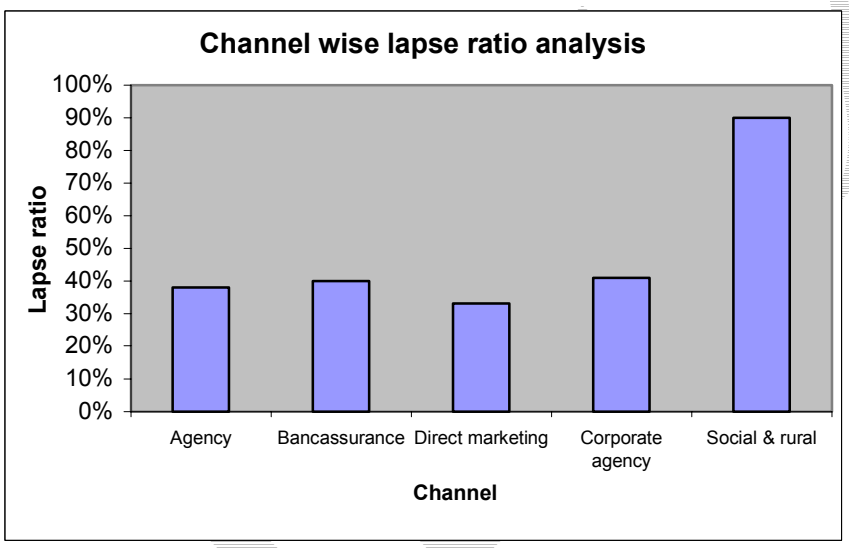
Illustration: In 3rd Quarter of 2005 an insurance company decides to do Lapse ratio analysis

Histogram can be used to describe the lapse ratios of policies with respect to policy Inforced year.

It can be seen that company's lapse ratio is as high as 40%. Also, this is consistent across policies issued in different years. Changes in renewal management and lapse reduction schemes have made no impact on the lapse ratio.



Now it will be useful if one know which are the channels leading to high lapses. Therefore a histogram showing lapse ratio across channels for period from 2001 to Q3 2005 will serve our purpose.



The histogram shows that 90% of the policies have lapsed in "social and rural" channel.

This indicates the need to focus on this channel for improving the lapse ratio.

Now the company would like to know if owner income and annual premium have any impact on lapses. This can be found out by calculating the coefficient of correlation.

It may also be instructive to correlate lapse ratio with the %(premium/ owner's income). This may lead to changes in financial underwriting. Further one can formulate a regression equation. A regression equation allows us to express the relationship between two or more variables algebraically. It indicates the nature of the relationship between two or more variables. In particular, it indicates the extent to which you can predict some variables by knowing others.

4. New Product & Market Development

Company should choose the expansion method that best fits company's product or service, strengths and weaknesses, and the limitations of cash, credit and existing resources.

- New product development can be one of the methods of expansion.
- New market development can help you reach new customers

While developing a new product business users would need to address the following questions:

- Which product features are important to customer?
- What are the gaps in product portfolio that need to be filled?
- What are the gaps in company's product mix? For example if the company has very low sales of policies in insured age band between 25 to 30 years then it can introduce new product to cater to this specific age band and fill in the gap in the product mix.

Market segmentation analysis is a key tool for understanding what drives consumers' purchase behaviour and how their values and beliefs affect their behaviours. By identifying distinctive groups of consumers—their characteristics, needs, purchase patterns, attitudes and/or values—segmentation research can help firms develop offerings targeted to the most profitable consumers.

While taking products to a new market following questions need to be addressed by the management:

- What is the demographic profile of people who buy the product?
- Does the demographic profile vary across products?
- Has it varies across years?

To answer the business questions, some of the data that is required include:

Internal data:

- Using Existing customer base to vet a new product's potential market value.
- Customer profile for each policy (Demographic and socio-economic)

External Data:

- Psychological and lifestyle characteristics
- Attitudes and opinions
- Awareness and knowledge - for example, brand awareness
- Intentions - for example, purchase intentions. While useful, intentions are not a reliable indication of actual future behaviour.

Illustration: Determine features of insurance policy important to a people in a particular market segment

Products promise a set of benefits like adjustable face amount, flexible premiums, option to add riders, Ease of making claims, premium reminders, tacking policy details online, Tax advantage, security, retirement planning, unit linked, etc.

These product features can be grouped into 3 factors:

Convenience (Factor X):
Variable 1: Ease of making claims
Variable 2: Premium reminders
Variable 3: Tacking policy details online

Flexibility (Factor Y):
Variable 1: Adjustable face amount
Variable 2: Flexible premiums
Variable 3: Option to add riders

Investment returns (Factor Z):
Variable 1: Tax advantage
Variable 2: Retirement planning
Variable 3: Unit linked

The factor loading can be defined as the correlations between the factors and their underlying variables. A factor-loading matrix is a key output of the factor analysis.

	Factor X	Factor Y	Factor Z
Variable 1	0.7	0.6	0.6
Variable 2	0.6	0.6	0.8
Variable 3	0.9	0.7	0.8
Column's Sum of Squares	1.66	1.21	1.64

Each cell in the matrix represents correlation between the variable and the factor associated with that cell. The square of this correlation represents the proportion of the variation in the variable explained by the factor. The sum of the squares of the factor loadings in each column is called an eigenvalue. An eigenvalue represents the amount of variance in the original variables that is associated with that factor. The communality is the amount of the variable variance explained by common factors.

5. Cross Selling

It costs 5 times more to acquire a new customer than retaining an existing customer. Encouraging existing customers to spend more can have a dramatic effect on company's

profit margins. For example, Telecom company entering Insurance business can sell insurance to their existing customers; an Insurance company with composite license can sell life insurance to there general insurance customers**;** an Insurance company can aim at repeat business from existing satisfied customers by offering them riders and new policies based on the customers needs. Cross selling if done successfully enhances loyalty and reduces the cost of sales.

While encouraging existing customers to spend more management should find answers to the following questions:

- Who are the target customers for cross selling a particular product?
- Will investing heavily in call centres and technology creates more satisfied customers for cross selling?
- Will training to customer-service representatives that increases the quality and consistency of service, help cross selling?
- For any given customer household, what product or products should be marketed next?

To answer the business questions, some of the data that is required include:

Internal data:

- Details of policies sold to existing customers
- Customer profile (age, gender, annual income) that can be used to predict financial needs.
- Training details of customer-service representatives that increases the quality and consistency of service
- Investments in call centres and technology

External data:

- Life style data (wants, needs and desires)

Illustration: Identifying the next insurance product to sell to a particular client

The approach is to develop a targeting model for each possible product or service type. This predictive model would use recent product or service acquisition as the dependent variable. It would use other product usage and demographic and lifestyle data as predictors.

By using pre-existing product ownership, demographics and lifestyle variables predictors, we can generate not only a powerful scoring model for cross selling, but also a richer understanding of why a household's characteristics indicate that the household is a good candidate for targeting.

Each client is scored on each model separately, using a separate model for each product or service. Then the model scores are ranked from high to low for each client. And the next product to sell to a particular client is determined by the highest scoring product from among the array of candidate products that the client has not yet acquired.

Conclusion

Today's competitive insurance market has made the critical link between good decision-making and success more important than ever. Increasingly organizations are turning to statistical analysis to guide decision-making processes.

Business data exist in many different forms, including time series and longitudinal data, and may contain various types of variables, such as time to event, occurrences, categorical and continuous measurements. The challenge in using statistics for business is that one has to estimate the parameters without comprehensive information. There are many statistical analysis tools on the market today that provides a complete, comprehensive and integrated platform for data analysis. Using optimal statistical techniques can provide new information that improves process turnaround time, reduces policy lapse ratios, and helps retain valued and satisfied agents; hence driving development and revenues.

Annexure 1- Important Statistical Definitions

Hypothesis Testing

Estimation of population parameters is also useful to answer questions about comparisons or relationships: "Is one treatment more effective than the other?" or "Is there a relationship between length of treatment and treatment outcome?" These types of questions usually involve comparison of means, proportions, standard deviations, or some other statistic.

According to the concept of sampling error, sample parameters (e.g. means) of different samples will be expected to differ based on the nature of sampling. A statistical comparison of those parameters is required to determine if the difference is "true" or "false". Such a comparison is called hypothesis testing.

When evaluating observed differences between different samples, we must consider two different outcomes: either the two samples are truly different, or the differences are due to chance occurrence. The second of these two outcomes is called the null hypothesis, H_0. The first of these two outcomes is called the alternate hypothesis, H_1.

Chi-Square Analysis of Frequencies

The chi-square statistic is a nonparametric statistical technique used to determine if a distribution of observed frequencies differs from theoretical expected frequencies.

Goodness of Fit: tests if proportions are the same as theoretical expected proportions,

Independence Tests: tests for relationships among categorical data based on proportions

McNemur Test: used for correlated samples (e.g. comparing proportions of correct diagnosis using two different modalities, such as MRI versus CT for detecting cartilage defects)

Coefficients of Association: tests if two categorical variables are related significantly

Any appropriately performed test determines the degree of confidence one can have in accepting or rejecting a hypothesis. Typically, the hypothesis tested with Chi Square is whether or not two different samples are different enough in some characteristic or aspect of their behaviour that we can generalize from our samples that the populations from which our samples are drawn are also different in the behaviour or characteristic. Chi Square is used most frequently to test the statistical significance of results reported in bivariate tables

Correlation

The correlation coefficient summarizes the relationship between two variables. Correlation is a measure of association that tests whether a relationship exists between two variables. It

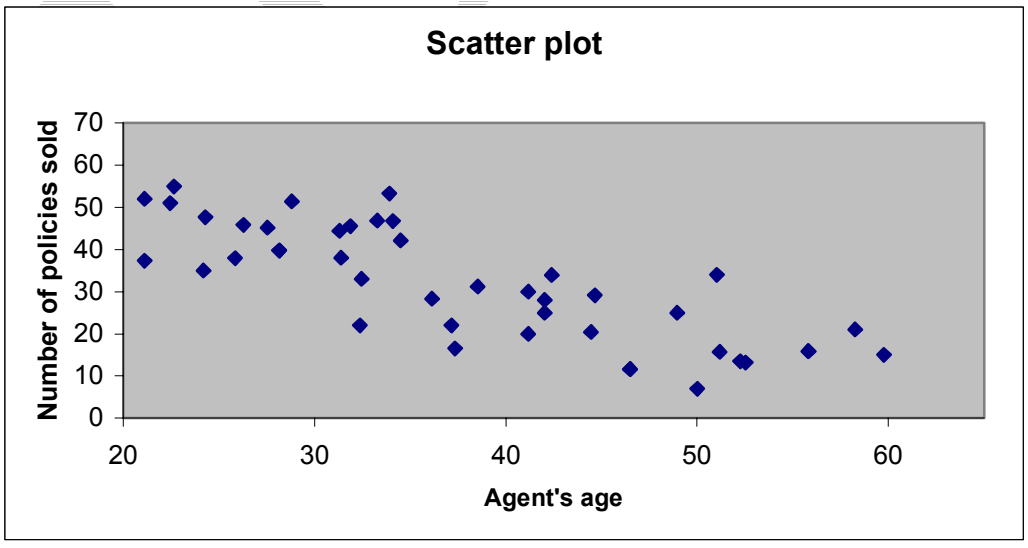
indicates both the strength of the association and its direction. A correlation coefficient is a number between -1 and 1, which measures the degree to which two variables are linearly related.

Let's take an example. Did you ever wonder whether the insurance agents performance depends on his age? It might be that the agents who are aged have more experience, and they sell more policies. This would be an example of a positive correlation, because high values of one variable (e.g., agents age) are associated with high values on the other variable (e.g., better performance on sales). Or it might be the other way around: age of person is associated with poorer sales performance. The latter is an example of a negative correlation, because high values on one variable are associated with low values on another variable.

Scatter plots

It is useful to obtain a plot of the joint distribution of the values of the two variables, "age of the agent"(X) and "sales"(Y). These are called scatter plots If small values of X are associated with small values for Y, and large values of X are associated with large values of Y, then the data will stretch from the lower left hand corner of the plot to the upper right hand corner of the plot. This indicates a positive relationship.

If small values of X are associated with large values for Y, and large values of X are associated with small values of Y, then the data will stretch from the upper left hand corner of the plot to the lower right hand corner of the plot. This indicates an inverse relationship. If there is no discernible pattern to the distribution, then the two variables probably are not related in a linear fashion.



The above scatter plot depicts a negative correlation between agents age and sales performance.

Simple Linear Regression

Simple linear regression aims to find a linear relationship between a response variable and a possible predictor variable by the method of least squares.

Multiple Regression

Multiple linear regression aims is to find a linear relationship between a response variable and several possible predictor variables.

Non-linear Regression

Non-linear regression aims to describe the relationship between a response variable and one or more explanatory variables in a non-linear fashion.

Regression Equation

A regression equation allows us to express the relationship between two (or more) variables algebraically. It indicates the nature of the relationship between two (or more) variables. In particular, it indicates the extent to which you can predict some variables by knowing others, or the extent to which some are associated with others.

A linear regression equation is usually written as:

$Y = a + bX + e$

Where,
Y is the dependent variable
a is the intercept
b is the slope or regression coefficient
X is the independent variable (or covariate)
e is the error term

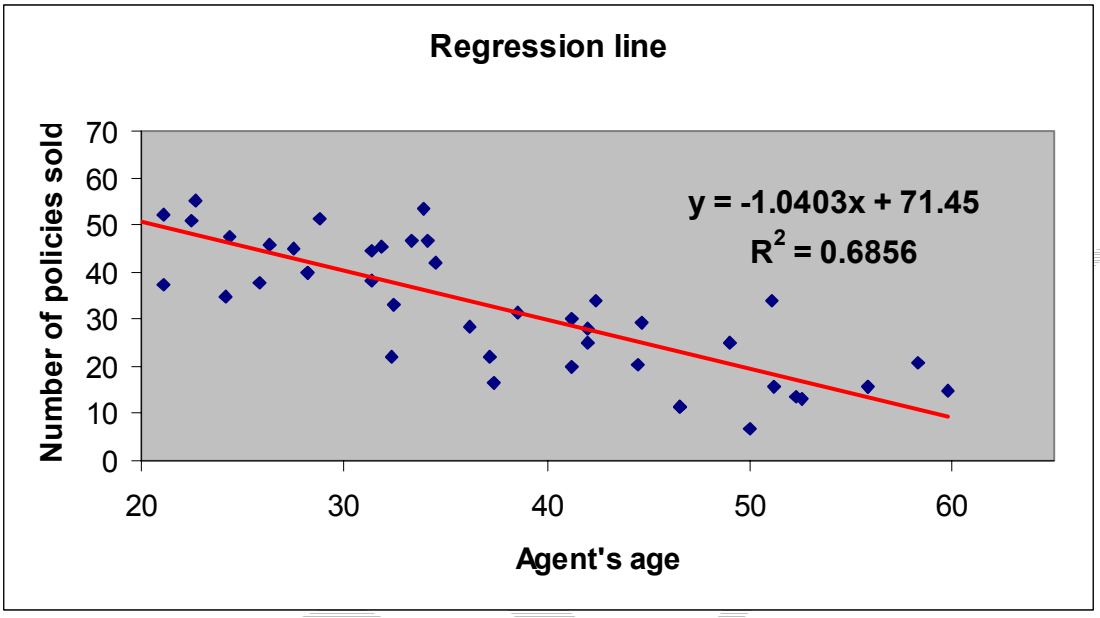
The equation will specify the average magnitude of the expected change in Y given a change in X.

Regression line

The regression equation is often represented on a scatter plot by a regression line.

Regression lines can be used to visually depicting the relationship between the independent (x) and dependent (y) variables in the graph. A straight line depicts a linear trend in the data (i.e., the equation describing the line is of first order. For example, $y = 2x + 7$). A curved line represents a trend described by a higher order equation (e.g., $y = 3x^2 + 7x - 9$).

How well this equation describes the data (the 'fit'), is expressed as a correlation coefficient (R^2). The closer R^2 is to 1.00, the better the fit.



Annexure 2- References

www.georgetown.edu
www.entrepreneur.com
www.clearlybusiness.com
www.clomedia.com
www.thehindubusinessline.com
www.stats-consult.com
www.home.ubalt.edu
www.smartdrill.com
www.nowsell.com
www.financetech.com
www.persfin.co.za
www.stats.govt.nz
www.sas.com