1.1 Statistics in Business

Statistics is a mathematical science concerned with the collection, presentation, analysis, interpretation or explanation of data.

Aim- to extract the best possible information from data and use it to make business decisions.

Areas of Application

- Monitoring stock markets, investments and commodities
- Reporting economic data (interest rates, un-employment, consumer confidence, etc).
- Market research, interpreting sales figures, forecasting demand.
- Assessing the effectiveness of advertising
- Detecting credit card, mobile phone and banking fraud.
- Maintaining internet security
- Weather forecast
- Improving production methods and assessing product reliability.

1.2 Basic statistical concepts

Population- A population is a collection of objects (often called units or subjects) of interest. Examples include; ALL small businesses, ALL workers currently employed for Mcdonalds.

Collection of data on a whole population is called a census.

Sample- A sample is a subset of the units in a population.

Some forms of data collection are destructive. For example, crash test statistics for a particular model of car. Sampling is the only option as testing all cars would be impossible as they would all be destroyed.

Two steps in analysing data from a sample; exploratory data analysis and statistical inference. These are related and both should be performed for any given data.

Exploratory data analysis (EDA)- is the first step, in which numerical, tabular and graphical summaries of data are produced to summarise and highlight the key aspects and any special features of the data.

Statistical inference uses sample data to reach conclusions about the population from which the sample was drawn. Usually the main aim of any statistical exercise and involves more formal data analysis techniques. An inference is a conclusion that patterns observed in the data sample are present in the wider population and clearly assumes that the sample is representative of the population.

Parameter- Is a descriptive measure of the population. Usually denoted by Greek letters. Examples of parameters are population mean, population standard deviation and population variance.

A descriptive measure of a sample is called a Statistic. Usually denoted by Roman letters. Examples of statistics are sample mean, sample standard deviation and sample variance.

Distinction between the two terms is important. A business researcher often wants to estimate the value of a parameter or draw inferences about the parameter. However the calculation of parameters is usually impossible or infeasible because of the amount of time and money required. In such cases, the business researcher can take a representative sample of the population and use the corresponding sample statistic to estimate the population parameter.

1.3 Types of Data

Categorical data (qualitative)- is a data type that is simply an identifier or label and has no numerical meaning. For example, the employment of a person (teacher, doctor). This cannot be ranked in any meaningful way and thus is an example of a nominal data type.

As another example, the grade in a test (A,B,C,D,E,F) is again simply a label and is a categorical data type. The test grades have a natural ordering and thus is an ordinal data type.

Numerical data (quantitative)- have a natural order and the numbers represent some quantity.

Two examples are the number of heads in ten tosses of a coin and the weights of rugby players. Note that in the first example we know in advance exactly which values the data may take (0,1…10), whereas in the second example all we can give is a range, say (80-140kg). The first example is that of a discrete data type, where we can list the possible values. The second example is that of a continuous data type, where we can only give a range of possible values for the data. Discrete data often arise from counting processes, while continuous data arise from measurements.

Cross-sectional data- data that are collected at a fixed point in time. Such data give a snapshot of the measured variables at that point in time. For example, monthly surveys of consumer confidence. Provides information on consumer confidence for the given month.

Time series data- data collected over time. Example, data collected of consumer confidence over several months.

1.4 Obtaining data

Secondary data- data that were collected for some other purpose and are already available. Available from external (government departments, industry associations) and internal sources (sales figures, production records).

Primary data- data collected to address a specific need. Might be collected through a survey, experiment or some other study.

1.5 Statistical analysis using Excel

Computers enable fast and accurate statistical analysis of very large amounts of data. In fact, new statistical techniques have become possible due to the increase in computer power.

Business statisticians use software packages including SPSS, MINITAB and SAS. These however are very expensive. Many spreadsheet software packages can also analyse data statistically. Microsoft Excel is the most widely used package for business statistics due to its ease of use, but it is limited in its range of techniques. It cannot perform every type of statistical analysis, and some charts are of poor quality.

Remember that software cannot replace a thorough understanding of correct statistical methods.

1.6 When things go wrong

Incorrect data analysis can have disastrous consequences with large human and business costs.