Introduction to Statistics for Business

**Lecture:**

Welcome to the world of Statistics. You are probably wondering “why should I learn statistics?”

Statistics helps us make better sense of the world and make better business decisions.

* For example, by understanding statistics we are able to understand internet article and reports, magazine articles, newspaper articles.
* Also, we are able to use the data provided to us in business memos, business research, technical journals, and reports to make better business decisions.

In business, statistics has several critical uses. We use statistics to:

* summarize business data
* draw conclusions from business data
* make reliable forecasts about activities
* improve business processes

Today's good decisions are driven by data. In all aspects of our lives, and importantly in the business context, an amazing diversity of data is available for inspection and analytical insight. Business managers and professionals are increasingly required to justify decisions on the basis of data. They need statistical model-based decision support systems.

Statistical skills enable them to intelligently collect, analyze and interpret data relevant to their decision-making. Statistical concepts and statistical thinking enable them to:

* solve problems in a diversity of contexts
* add substance to decisions
* reduce guesswork

In competitive environment, business managers must design quality into products, and into the processes of making the products. They must facilitate a process of never-ending improvement at all stages of manufacturing and service. This is a strategy that employs statistical methods, particularly statistically designed experiments, and produces processes that provide high yield and products that seldom fail.

Moreover, it facilitates development of robust products that are insensitive to changes in the environment and internal component variation. Carefully planned statistical studies remove hindrances to high quality and productivity at every stage of production. This saves time and money. It is well recognized that quality must be engineered into products as early as possible in the design process. One must know how to use carefully planned, cost-effective statistical experiments to improve, optimize and make robust products and processes.

**Business Statistics** is a science assisting you to make business decisions under uncertainties based on some numerical and measurable scales.

* Decision making processes must be based on data, not on personal opinion nor on belief.
* A course in appreciation of statistical thinking gives business professionals an edge.
* Professionals with strong quantitative skills are in demand.

To get started in this course, it is important to learn the basic terminology. These include:

**Statistics**

* is the branch of mathematics that transforms data into useful information for decision makers.

**Descriptive Statistics**

* is the process ofcollecting, summarizing, presenting and analyzing data. Inferential statistics uses data collected from a small group to draw conclusions about a larger group.

**Descriptive Methods**

* are used to create charts and tables, to draw conclusions about business data.
* We collect data (like in a survey), present data (in tables and charts), and characterize data (by giving the sample mean).

**Inferential Methods**

* are used to make reliable forecasts about business activities, to develop, quantify, and improve accuracy of predictive models.
* For example, we estimate the population mean weight using the same mean weight. Or, we may test the claim that the population mean weight is 180 pounds.

**Variables**

* are the characteristics of an item or individual and are what you analyze when you use a statistical method.
* For example, sales, expenses, and net profile.
* When used in every day speech, variable suggests that something changes or varies, and you would expect sales, expenses, and net profit to have different value from year to year.

**Data**

* are the different values associated with a variable.
* Data values are meaningless unless their variables have **operational definitions,** universally accepted meanings that are clear to all associated with an analysis.

**Population**

* consists of all the items or individuals about which you want to draw a conclusion.
* The population is the “large group”.
* All registered voters in Ohio is an example of a population.

**Sample**

* is the portion of a population selected for analysis.
* The sample is the “small group”.
* Using the population of all registered voters in Ohio, you could create a sample of 500 registered voters to survey.

**Parameter**

* is a numerical measure that describes a characteristic of a population.
* The average amount of money spent by all customers at a store this weekend is an example because this amount refers to the amount spent in the entire population.

**Statistic**

* is a numerical measure that describes a characteristic of a sample.
* The average amount spent by 30 customers completing the customer satisfaction survey is an example of a statistic.

Before we start looking at graphing, let’s review some basic terminology in organizing data.

**Categorical (*qualitative*) variables**

* have values that can only be placed into categories, such as “yes” and “no”
* Examples include marital status, political party, eye color (defined categories)

**Numerical** (***quantitative***) **variables**

* have values that represent quantities.
* **Discrete** variables
  + arise from a ***counting process***
  + Examples: number of children or defects per hour
* **Continuous** variables
  + arise from a ***measuring process***
  + Examples: weight, voltage

**Nominal Scale**

* Classifies data into distinct categories in which no ranking is implied.
* For example, the response to who is your internet provider results in answers that are not ranked in value. They are all equal.

**Ordinal Scale**

* Classifies data into distinct categories in which ranking is implied.
* For example:
  + the results to what is your faculty rank could be lecturer, instructor, assistant professor, associate professor, and professor
  + Or, what is your grade?
* Each answer has a value that is ranked higher or lower on a scale.

**Surveying and Sampling**

**Establishing A Business Objective Focuses Data Collection**

**Examples of Business Objectives:**

* A marketing research analyst needs to assess the effectiveness of a new television advertisement.
* A pharmaceutical manufacturer needs to determine whether a new drug is more effective than those currently in use.
* An operations manager wants to monitor a manufacturing process to find out whether the quality of the product being manufactured is conforming to company standards.
* An auditor wants to review the financial transactions of a company in order to determine whether the company is in compliance with generally accepted accounting principles.

**Sources of Data**

**Primary Sources**

* The data collector is the one using the data for analysis
  + Data from a political survey
  + Data collected from an experiment
  + Observed data

**Secondary Sources**

* The person performing data analysis is not the data collector
  + Analyzing census data
  + Examining data from print journals or data published on the internet.

**Five Categories of Sources of Data**

* Data distributed by an organization or an individual
* A designed experiment
* A survey
* An observational study
* Data collected by ongoing business activities

**Examples of Data Distributed By Organizations or Individuals**

* Financial data on a company provided by investment services.
* Industry or market data from market research firms and trade associations.
* Stock prices, weather conditions, and sports statistics in daily newspapers.

**Examples of Data from a Designed Experiment**

* Consumer testing of different versions of a product to help determine which product should be pursued further.
* Material testing to determine which supplier’s material should be used in a product.
* Market testing on alternative product promotions to determine which promotion to use more broadly.

**Examples of Survey Data**

* Political polls of registered voters during political campaigns.
* People being surveyed to determine their satisfaction with a recent product or service experience.

**Examples of Data Collected from Observational Studies**

* Market researchers utilizing focus groups to elicit unstructured responses to open-ended questions.
* Measuring the time it takes for customers to be served in a fast food establishment.
* Measuring the volume of traffic through an intersection to determine if some form of advertising at the intersection is justified.

**Examples of Data Collected from Ongoing Business Activities**

* A bank studies years of financial transactions to help them identify patterns of fraud.
* Economists utilize data on searches done via Google to help forecast future economic conditions.
* Marketing companies use tracking data to evaluate the effectiveness of a web site.

**Data Cleaning is Often a Necessary Activity When Collecting Data**

* Often find “irregularities” in the data
  + Typographical or data entry errors
  + Values that are impossible or undefined
  + Missing values
  + Outliers
* When found these irregularities should be reviewed
* Many statistical software packages will handle irregularities in an automated fashion (Excel does not)

**Why Sample?**

* Selecting a sample is less time-consuming than selecting every item in the population (census).
* An analysis of a sample is less cumbersome and more practical than an analysis of the entire population.

**A Sampling Process Begins with a Sampling Frame**

* The sampling frame is a listing of items that make up the population
* Frames are data sources such as population lists, directories, or maps
* Inaccurate or biased results can result if a frame excludes certain portions of the population
* Using different frames to generate data can lead to dissimilar conclusions

**Types of Samples**

* In **convenience sampling**, items are selected based only on the fact that they are easy, inexpensive, or convenient to sample.
* In a **judgment sample,** you get the opinions of pre-selected experts in the subject matter.
* In a **simple random,** every individual or item from the frame has an equal chance of being selected
* **Systematic Sample:** Decide on sample size, divide into a set number of groups, and randomly select one individual from each group.
* **Stratified Sample**: Divide population into two or more subgroups (called *strata*) according to some common characteristic.
* **Cluster Sample:** Population is divided into several “clusters,” each representative of the population (exit polls after elections).

**Types of Survey Error**

**1) Coverage error or selection bias**

* Exists if some groups are excluded from the frame and have no chance of being selected.
* If the frame is inadequate because certain groups of items in the population were not properly included, any random sample selected will provide only an estimate of the characteristics of the frame, not the actual population.

**2) Non response error or bias**

* People who do not respond may be different from those who do respond
* You should make several attempts to convince individuals that may typically not complete the survey to complete it. Mode of response should be considered.

**3) Sampling error**

* Variation from sample to sample will always exist. Chance indicates some group may always be left out.
* Example when you read polls you are told the margin of error is +/- 4 percentage points of the actual value.
* You can reduce sampling error by using larger sample sizes, although doing so increases the cost of doing the survey.

**4) Measurement error**

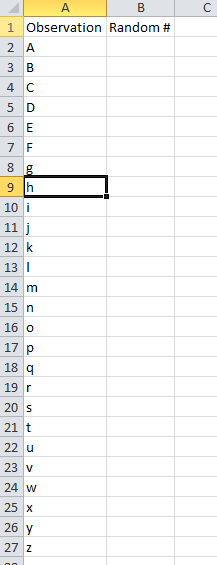
* Due to weaknesses in question design, respondent error, and interviewer’s effects on the respondent (“Hawthorne effect”). Occurs when the interviewee feels compelled to please the interviewer.

**Ethical Issues**

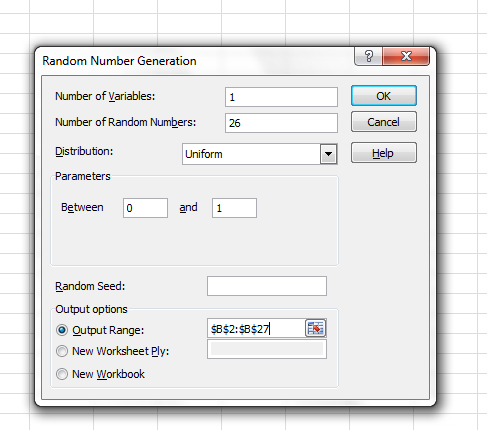
* Coverage error can result in selection bias and becomes an ethical issue if particular groups are purposely excluded from the frame so that the survey results are more favorable to the survey’s sponsor.
* Non-response error can lead to nonresponse bias and becomes an ethical issue if the sponsor knowingly designs the survey so that particular groups or individuals are less likely than others to respond.
* Sampling error becomes an ethical issue if the findings are purposely presented without reference to the sample size and margin of error so that the sponsor can promote a viewpoint that might otherwise be truly insignificant.
* Measurement error becomes an ethical issue if
  + the survey sponsors chooses leading questions that guide the responses in a particular direction,
  + an interviewer, through mannerisms and tone, purposely creates the Hawthorne effect or
  + a respondent willfully provides false information.

**Illustration of Random Sampling**

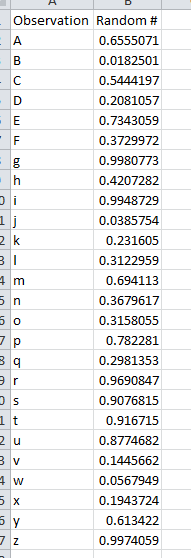
* In this exercise we will use random number generation to take a random sample from population. We begin by entering the data into the first column.
* Suppose this data represents the clients in your company. We have a population of size 26, and we are going to take an individual sample of size 8. Next, we assign random numbers to each individual.

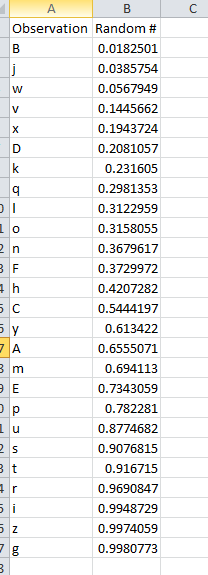


* We want 26 random numbers (1 for each unit in our frame), uniform distribution, and we want the output range to be the cells immediately to the right of our data “$B2:$B27”. You should put a label in the cell above this range.



* Then click OK.
* You’ll see the random #’s fill in.

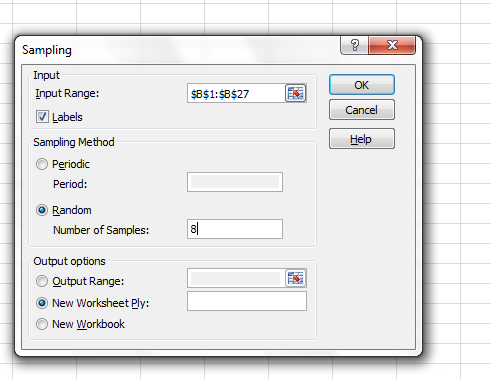




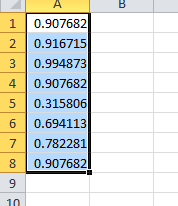
Once we have our random numbers, how does this help us take a random sample?

* If the numbers are random, then if we sort the individuals by these numbers, we should get a random ordering of the individuals.
* You can now pick the first 8 to get your random sample of 8, which would be B, J, W, V, X, D, K, Q. Those would be the people that you survey.
* Determine the appropriate sample size was/will be discussed in other lecture notes. In this problem, you will give the sample size.

Let’s say you wanted to use this same data, but use Excel to randomly pull 8 values from the survey set.



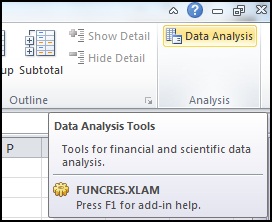
You are then provided with a random sample of 8 numbers from your list of 26

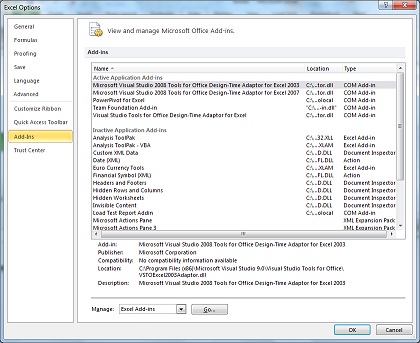


You can then identify the corresponding client letter these numbers are associated with, and that is your survey sample base.

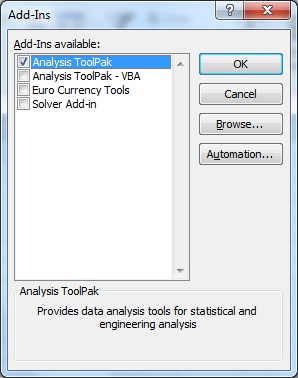
**Getting Started with Excel Data Analysis Toolpak**

* Excel 2010 provides some more advanced statistical and engineering analysis tools that are not installed by default.
* If you’re interested in using Excel for statistical analysis, you should install the Analysis Toolpak for Excel 2010.  Follow the instructions below, or view this video on Youtube: <http://www.youtube.com/watch?v=pTW2wt0jFe8>
* To confirm whether you already have the Analysis Toolpak installed, open the Data tab on the Excel ribbon. If the Analysis Toolpak is installed, you should see a Data Analysis button on the Ribbon, like the one shown here.





* The Add-Ins dialog will open. Here, you can select the check box next to Analysis Toolpak (and any other add-ins you want to install). Click OK.



* The Data Analysis button (as shown in the first screenshot) should now be available on the Excel Ribbon, under the Data tab. This button will open the Data Analysis dialog, which offers access to a variety of analysis tools.