Quantitative Methods

Year 2

Semester 1

Lecturer: Patricia Gunning

E_mail: pgunning@computing.dcu.ie

Quantitative Methods

Reading Material

- Any introduction to *Probability & Statistics* text in the library, e.g.
  - *Elementary Statistics: A Step by Step Approach*
    By Alan Blumen. McGraw Hill, 2003.
  - *Basic Business Statistics*
    By Berenson, M.L. and Levine, D.M. Prentice Hall 1996.
  - *Tables for Statisticians*
    By White, J., Yeats, A. and Skipworth, G.
    Stanley Thornes Ltd.

Quantitative Methods

What is Statistics?

The subject of STATISTICS deals with data:

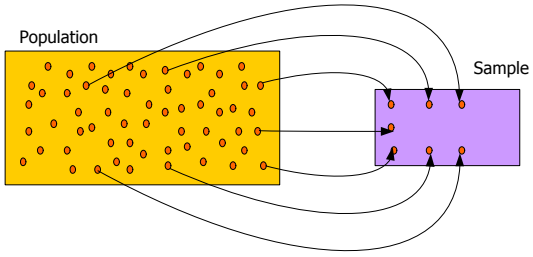
- COLLECTION
- DESCRIPTION
- PRESENTATION
- INTERPRETATION

2 TYPES OF STATISTICS

Descriptive Statistics consists of methods for organizing, displaying and describing data by using tables, graphs, and summary measures.

Inferential Statistics is a process of describing the population based on the sample results.

Population:	The complete set of elements to be studied.
Sample:	A subset of the population.
Parameter:	A characteristic of the entire population.
Statistic:	A characteristic of the sample.



Statistical Inference consists of:

1. Estimation: Estimating the unknown parameters of the population based on the sample results.
2. Hypothesis Testing: A hypothesis is a conjecture about the population, e.g. the response time of the new system is less than the old.

   A test of hypothesis uses the sample data to reject or accept the conjecture.

Statistical Inference

- No sample can fully reflect precisely the qualities of the population.
- We can never be 100% certain when inferring from sample to population.
- Examining a sample rather than the whole population involves a degree of uncertainty called the sampling error.

Example

- In a population of 1,000 accounts, there are 50 accounts in error (i.e. 5%). Suppose a sample of 100 are taken. We can hardly hope to draw exactly that same percentage of accounts in error.
- Results depend on which selection of items we have.
- It is "likely" that "close to" this percentage will turn up in the sample.

- Need to define "likely" and "close to" more precisely.
- Require knowledge of the laws of probability in the face of uncertainty.
- Need the necessary deduction

Probability theory

so that (induction) statistical inference can be developed.

Probability – mathematical tools that enable us to answer the basic deductive question in statistics:

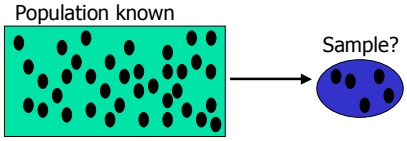
What can we expect of a random sample drawn from a known population?

Use probability theory to ascertain how much sampling uncertainty samples have.

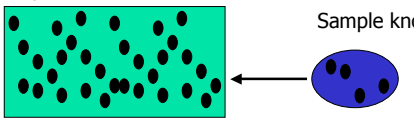
Then addressed the problem of statistical inference of:

How can the population be estimated by the sample?

Deduction (probability)
Population known



Induction (statistical inference)
Population?



Brief Outline of Course

Descriptive statistics
- Numerical summaries.
- Graphical illustrations
  - Box plots
  - Stem and leaf diagrams.

Probability
- Basic concepts and rules of probability.
- Counting Techniques
  - Find number of ways *r* objects can be selected from *n* objects.
  - Probability of an event using counting rules.
- Conditional probability – set of relevant outcomes are restricted by a condition.

Brief Outline of Course

Probability

- Random variables
- Standard Probability Distributions.
  - Geometric
  - Hypergeometric
  - Binomial
  - Poisson
  - Exponential
  - Normal
- Expected value and variance.
  - Decision making using expectation
    - Example: Show how an investor with a certain amount of money and a selection of portfolios can determine the best portfolio.

Brief Outline of Course

Probability

- Randomised response
  - How to use randomisation to guarantee anonymity in a sensitive area of survey sampling.
- Probability of winning the lotto.
- Chances that 2 people in the class share the same birthday.

Brief Outline of Course

Inference

- Basic ideas of inferential statistics.
- Estimation.
  - What conclusions can be drawn from sample data?
  - What is the degree of certainty of these conclusions?
  - What actions should be taken on the basis of the conclusions drawn from these data?
  - Point and interval estimation.
  - Properties of a good estimator.
- Samples sizes.
  - Eg. Determine the sample size necessary to estimate the mean to within 10 euro of the true value with 95% confidence.

APPLICATIONS OF STATISTICS

Actuarial Science: Determining premium rates; designing pension plans. measuring the effectiveness of loss presentation and loss control programs.

Finance: Pricing of options and derivatives; modelling and management of financial risk.

Business, Accounting and Industry: Estimating the volume of retail sales; producing auditing and accounting procedures to provide background for management decisions; designing inventory control systems.

Economics: Measuring indicators such as trade, size of labour force, and standard of living; long and short range forecasts of economic indicators.

Law: Weighing evidence and assessing legal doubt; auditing risk and expert witness.

Opinion Polls: Who will win the next general election?

Basic Terminology

Census: a survey that includes every member of the population.

Sample survey: is the technique of collecting information from a portion of the population.

Representative sample: is a sample that represents the characteristics of the population as closely as possible.

Random sample: is a sample drawn in such a way that each element of the population has a chance of being selected.

If the chance of being selected is the same for each element of the population, then it is called a simple random sample.

Basic Terminology

An Element of a sample or population is a specific subject or object (for example, a person, firm, item, state, or country) about which the information is collected.

A Variable is a characteristic under study that assumes different values for different elements.

An Observation is the value of a variable for an element.

Data set is a collection of observations on one or more variables.

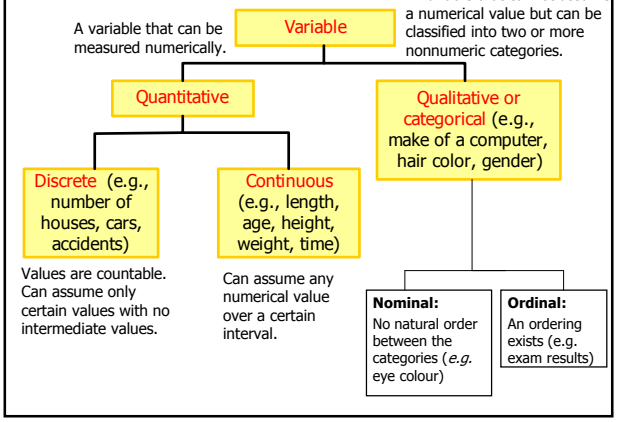
BASIC TERMS

2005 Sales of Five Supermarkets

Company	2005 Sales (millions of euro) ← Variable
Londis	217,799
Centra	85,866
Dunnes	177,260 ← An observation or measurement
Spar	31,168
Supervalue	32,004

An element or a member → Dunnes

Types of Variables



DATA PRESENTATION AND DESCRIPTION

Summary Statistics

1. **Measures of Central Tendency**
   These measure typical or central points in the data. They include:

- **Mean:** Sum of all values divided by the number of cases.
- **Median:** The value of the middle term in a data set that has been ranked in increasing order. 50% of the data lies below this value and 50% above.
- **Mode:** is the value that occurs with the highest frequency in a data set.

Outliers

Values that are very small or very large relative to the majority of the values in a data set are called ***outliers*** or ***extreme values***.

Outliers impact on the mean.

The advantage of using the median as a measure of central tendency is that it is not influenced by outliers. Consequently, the median is preferred over the mean as a measure of central tendency for data sets that contain outliers.

MEASURES OF DISPERSION

Standard Deviation
The square root of the average squared deviations from the mean. This measure how the data values differ from the mean. A small standard deviation implies most values are near the average. The large standard deviation indicates that values are widely spread above and below the average.

Range = Largest value – Smallest Value
The range, like the mean has the disadvantage of being influenced by outliers.

Percentiles
Values that divide cases below which certain percentages of values fall. The 50th percentile is the median.

Interquartile Range
Quartiles are three summary measures that divide a ranked data set into four equal parts.

- Second quartile is the same as the median of a data set.
- First quartile is the value of the middle term among the observations that are less than the median.
- Third quartile is the value of the middle term among the observations that are greater than the median.

Interquartile range:
The difference between the third and first quartiles or equivalently the middle 50% of the data.

Standard Error (SE): Standard deviation of the mean.

Graphical Illustrations

1. Box and Whisker Plot

This is a graphical summary based on the median, quartiles and extreme values. Often called the Box and Whiskers Plot, the box represents the interquartile range which contains 50% of the cases.

The whiskers are lines that extend from the box to the highest and lowest values.

A line across the box indicates the median.

Extreme values are cases more than 1.5 box lengths from the upper or lower end of the box. The extreme cases are listed on the plot.

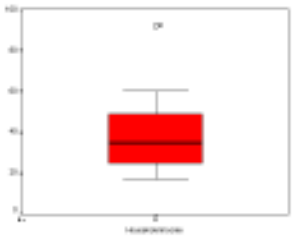
Example of a Box and Whisker Plot

The following data are the incomes (in thousands of euros) for a sample of 12 households.

23 17 32 60 22 52 29 38 42 92 27 46

Construct a Box and Whisker Plot for these data.

Box and Whisker Plot for Household Incomes



2. Stem and Leaf Displays

This is a depiction of the shape of the data based on the actual numbers observed. The stem usually depicts the 10s and the leaves depict the units.

Example of a Stem and Leaf Display

The following are the scores of 30 college students on a statistics test:

75	52	80	96	65	79	71	87	93	95
69	72	81	61	76	86	79	68	50	92
83	84	77	64	71	87	72	92	57	98

Construct a stem-and-leaf display.

Scores Stem-and-Leaf Plot

Frequency	Stem &	Leaf
2.00	5 .	02
1.00	5 .	7
2.00	6 .	14
3.00	6 .	589
4.00	7 .	1122
5.00	7 .	56799
4.00	8 .	0134
3.00	8 .	677
3.00	9 .	223
3.00	9 .	568

Stem width: 10
Each leaf: 1 case(s)