

STUDY GUIDE

ETC1000- Business and economic statistics

1. Data, Information, Decisions

- Statistics is about getting information from data and using the information to make decisions
- intuition or accepted wisdom can often be wrong or misleading, we need hard evidence

Business Intelligence- software applications used to analyse an organisation's raw data

1. Descriptive Analytics - **describe data**
 - describing characteristics of a set of data
2. Diagnostic Analytics - **correlations in data**
 - discovering patterns or correlations in data
3. Predictive Analytics - **building models**
 - building models that allow us to predict outcomes
4. Prescriptive Analytics - **using models**
 - using models to design actions that will influence outcomes

Big Data

- a collection of large and complex data sets
- bringing together data from different sources
- difficult to process using standard data management tools
- requires **computing** (to manage and organise the data), **visualisation** (to identify patterns and relationships), **statistics** (to discover patterns and develop models)

2. Sources of Data

- data comes from many sources: company, databases, government reporting, market research, etc.

Electronic data capture

- electronic data capture has allowed the volume of data to grow enormously and improve quality of data (less measurement error)
- difficulty to link data about one subject from different sources

Surveys

- source of data
- extracting targeted information
- allows us to relate information together (eg. person characteristics and preferences)
- learning directly from the people we are most interested in
- Challenges:
 1. Questions that extract the right information and are clearly worded
 2. Samples that are representative, not biased, sample from the wider population

Bias

1. Selection Bias: the sample is chosen in a way that makes it not representative (solution: take random sample)
 2. Non-Response Bias: not all who are selected for a sample will respond
 3. Response Bias: responding with the incorrect information
- it is more important to have a **random & representative sample** than a large sample!

Notes:

What kinds of things are we trying to achieve in an initial data exploration?

- Discovering what data is available