BUSS1020 Quantitative Business Analysis

Complete Study Notes

Contents

How to use these notes 1

1. Introduction to data and sampling 2

2. Descriptive measures and presenting data 5

3. Basic probability 9

4. Discrete probability distributions 12

5. Continuous probability distributions 15

6. Sampling distributions 18

7. Confidence intervals 20

8. One-sample hypothesis tests 22

9. Two-sample hypothesis tests 26

10. Simple linear regression 28

11. Multiple regression 33

12. Exam toolkit 35

Page numbers refer to the page footers.

BUSS1020 Quantitative Business Analysis – Complete Study Notes

These notes cover all 12 content modules of BUSS1020 Quantitative Business Analysis: data and sampling, descriptive statistics, probability, distributions, sampling distributions, confidence intervals, hypothesis testing and regression. Each topic includes plain-English explanations, the key formulas, Excel functions and original worked examples with full solutions.

How to use these notes

- **Formulas** are written in plain text so they are easy to copy: $\bar{x}$ = sample mean, s = sample standard deviation, μ = population mean, σ = population standard deviation, π = population proportion, p = sample proportion.
- **Worked examples** are original and show every step, so you can check your method.
- **The exam toolkit at the end** has a "which test do I use?" table, a full list of Excel functions and common mistakes.

Module	Topic
1	Introduction to data and sampling
2	Descriptive measures and presenting data
3	Basic probability
4	Discrete probability distributions
5	Continuous probability distributions
6	Sampling distributions
7	Confidence intervals
8	One-sample hypothesis tests
9	Two-sample hypothesis tests
10–11	Simple linear regression
12	Multiple regression

Memory hook: Parameters describe a **P**opulation and use Greek letters (μ, σ, π, ρ, β). **S**tatistics describe a **S**ample and use Latin letters ($\bar{x}$, s, p, r, b).

1. Introduction to data and sampling

Statistics turns data into decisions. Before any calculation, you need to know what kind of data you have, where it came from and whether the sample represents the population.

Key definitions

- **Business analytics:** using statistical methods and technology on historical data to gain insight and improve decisions.
- **DCOVA framework: D**efine the variables → **C**ollect data → **O**rganise → **V**isualise → **A**nalyse.
- **Three branches of statistics:**
  - **Descriptive:** summarise and present data (means, charts).
  - **Inferential:** use a sample to draw conclusions about a population (confidence intervals, tests).
  - **Predictive:** use a model to forecast outcomes (regression).

Term	Meaning	Example
Variable	A characteristic of an item or individual (also called an attribute)	Height, income, gender
Data	The observed values of a variable	172 cm, 165 cm…
Operational definition	A clear, agreed meaning for a variable	"Monthly sales = invoiced revenue, excluding GST"
Population	Every item you want to draw a conclusion about	All customers of a bank
Sample	The part of the population you actually analyse	500 surveyed customers
Parameter	A number describing a population (a fixed constant)	Population mean μ
Statistic	A number describing a sample (changes from sample to sample)	Sample mean $\bar{x}$

A statistic is used to **estimate** a parameter. For example, the mean height of a sample of 25 students ($\bar{x}$) estimates the mean height of all students (μ).

Types of variables

Type	Subtype	Description	Examples
Categorical (qualitative)	Nominal	Labels with **no natural order**; no arithmetic	Gender, eye colour, brand, student ID
Categorical (qualitative)	Ordinal	Categories with a **natural order**, but uneven or unknown gaps	Likert scale, education level, grade (HD/D/CR/P)
Numerical (quantitative)	Discrete	**Counts**: whole numbers	Number of children, website visits, complaints
Numerical (quantitative)	Continuous	**Measurements**: any value in a range	Height, weight, time, temperature

Levels of measurement for numerical data:

- **Interval scale:** differences are meaningful but zero is arbitrary, so values can go below zero and ratios make no sense. Example: temperature in °C (20°C is not "twice as hot" as 10°C).
- **Ratio scale:** has a true zero, so ratios make sense. Examples: height, speed, income.

Data sources and structure

- **Primary data:** collected by the analyst, such as the firm's own sales records, a survey, an experiment or an observational study.
- **Secondary data:** collected by someone else, such as the ABS, market research firms or web analytics providers.
- **Structured data:** fits rows and columns, with one column per variable and one row per item (spreadsheets, databases).
- **Unstructured data:** irregular formats that need interpretation, such as text, images, video, audio, emails and social media posts.

Data cleaning

Cleaning is required before analysis, because errors and extreme values distort results.

- **Look for:** invalid values (text in a numeric field, values out of range), coding errors ("NSW" vs "N.S.W."), and integration errors (duplicate rows, redundant columns, mixed units).

- **Fix by:** finding the correct value, deleting the wrong value, deleting the whole row, or substituting a value.

Why sample?

Studying the whole population is usually impossible or too costly. Sampling is quicker, cheaper and more practical. The **sampling frame** is the list you draw from, such as a customer database or directory. If groups are missing from the frame, results can be biased.

Sampling methods

Method	How it works	Pros and cons
Non-probability (not random; can't generalise reliably)		
Convenience	Pick whoever is easy to reach (includes snowball sampling)	Fast and cheap, but highly biased
Judgement	Experts choose who to include	Uses expertise, but subjective
Self-selection	People choose to take part (online polls)	Easy, but attracts people with strong views
Probability (random; can generalise)		
Simple random	Every item has an equal chance; use a random number generator (Excel: =RANDBETWEEN(1, N))	Unbiased, but needs a full frame
Systematic	k = N / n; pick a random start in the first k, then every kth item	Easy to run; risky if the list has a pattern
Stratified	Split the frame into strata (age, region), then randomly sample within each	Ensures every group is represented; more precise
Cluster	Split into clusters (suburbs, stores), randomly select whole clusters	Cheap for spread-out populations; less precise

Stratified vs cluster: stratified samples from **every** group; cluster samples **some** whole groups.