

# **ECON10005 - QUANTITATIVE METHODS 1**

**Study Notes**

# Contents

|                                               |    |
|-----------------------------------------------|----|
| 1 Data and Statistical Questions .....        | 2  |
| 2 Data Visualisation .....                    | 5  |
| 3 Numerical Descriptive Statistics .....      | 9  |
| 4 Probability .....                           | 15 |
| 5 Joint and Conditional Distributions .....   | 19 |
| 6 Discrete Random Variables .....             | 23 |
| 7 Continuous Random Variables .....           | 27 |
| 8 Sampling Distributions .....                | 30 |
| 9 Confidence Intervals .....                  | 34 |
| 10 Hypothesis Testing .....                   | 38 |
| 11 Comparing Two Populations .....            | 43 |
| 12 Simple Linear Regression .....             | 47 |
| 13 Regression Inference and Diagnostics ..... | 52 |
| 14 Communicating Quantitative Evidence .....  | 56 |
| A Formula Reference .....                     | 59 |
| B Excel Methods Reference .....               | 62 |
| C Definition Reference .....                  | 65 |

# 1. Data and Statistical Questions

Statistics begins with a question about a population and a set of observations that can answer it. The arithmetic comes later. A clean analysis identifies the cases, variables, population and sampling process before choosing a graph or formula.

## Definition – Population, sample, parameter and statistic

A **population** is the complete group of cases of interest. A **sample** is the subset actually observed. A **parameter** is a numerical feature of the population, such as the population mean  $\mu$ . A **statistic** is calculated from sample data, such as the sample mean  $\bar{X}$ .

Suppose a restaurant chain wants to know the mean annual sales per square foot across all its locations. Every location in the chain forms the population, the population mean is  $\mu$ , and a random set of observed locations forms the sample. The sample mean  $\bar{X}$  describes the observed locations and can also be used to estimate  $\mu$ . Those are different jobs: description concerns the data in hand; inference reaches beyond them.

## 1.1 Cases, variables and observations

A **case** is the unit described by one row of a dataset. A **variable** is a recorded characteristic, usually stored in one column. An **observation** is the value of a variable for a particular case.

**Table 1.1.** A rectangular dataset stores one case per row and one variable per column.

| Restaurant | Sales per sq ft | Floor area | Inside seats | Income group |
|------------|-----------------|------------|--------------|--------------|
| A          | \$412           | 2,300      | 84           | High         |
| B          | \$365           | 1,850      | 62           | Low          |
| C          | \$478           | 2,720      | 96           | High         |

The case is a restaurant. Sales per square foot, floor area and inside seats are quantitative variables; income group is categorical. The label in the first column identifies cases but does not measure a characteristic.

## 1.2 Descriptive and inferential statistics

### Key concept – Description and inference answer different questions

Descriptive statistics organise and summarise observed data. Inferential statistics use sample information, together with a probability model, to make a statement about an unobserved population quantity.

A histogram of sampled sales values is descriptive. A confidence interval for the population mean is inferential. The same sample can support both, but the language must remain distinct.

**Table 1.2.** The wording of a question indicates whether the target is descriptive or inferential.

|                | Descriptive                                          | Inferential                                                        |
|----------------|------------------------------------------------------|--------------------------------------------------------------------|
| Target         | Observed cases                                       | A wider population or process                                      |
| Typical verbs  | describe, compare, display, summarise                | estimate, test, infer, assess evidence                             |
| Uncertainty    | Variation among observations                         | Sampling variation of a statistic                                  |
| Typical output | graph, mean, median, standard deviation, correlation | confidence interval, test statistic, p-value, regression inference |

## 1.3 Types of variables

The variable type controls which summaries and displays are meaningful.

### Definition – Categorical and quantitative variables

A **categorical variable** places cases into groups. A **quantitative variable** records a numerical amount for which arithmetic has a meaningful interpretation.