

## Revision Tutorial Topic 3

### Numerical Descriptive Measures

---

#### Introduction

In this topic we will be looking at Numerical Descriptive Measures or, specifically, Measures of Central Tendency and Measures of Variation.

In the element Measures of Central Tendency one of the items we will be investigating is a pretty common term, an average. An average is a typical value but there are different types of averages. The most common average we use in everyday life is actually called the arithmetic mean. But there are two other types of averages. One is the Median, or the middle value (and this is the average we use when quoting an average house price in Real Estate) and the other is the Mode or the most occurring. And if I'm after a typical value, then I need to decide whether to use the Mean, the Median or, in some instances, the Mode.

Therefore, the aims of this tutorial are to:

- calculate and interpret numerical descriptive measures of central tendency, variation and shape for numerical data
- calculate and interpret descriptive summary measures for a population
- construct and interpret a box-and-whisker plot
- calculate and interpret the covariance and the coefficient of correlation for bivariate data

#### Textbook Questions

- 3.22 Naturally Soap is a small business, based in a coastal town, that makes and sells natural, luxurious, handmade soap bars in a variety of scents. Presently the soap is sold at local markets: Wednesday evening in the coastal town where the business is located, and a scheduled Sunday morning market in a roster of local villages. During the last six months, Naturally Soap has also been available via the Internet.

Naturally Soap is interested in analysing the quantity sold weekly at each market and Internet sales.

While Naturally Soap has complete sales and price data for both markets for the previous year, due to a computer 'problem' there is only a sample of weekly sales and price data for the Internet sales. The data is stored in the **[Dataset: NATURALLY\_SOAP.XLS]** file.

- a. For the Sunday morning market:
  - I. Calculate the mean, variance and standard deviation of the weekly sales for the year.
  - II. What conclusions can you make about the weekly sales for this market?
  - III. Use the empirical rule or the Chebyshev rule, whichever is appropriate, to further explain the variation in the weekly sales.
  - IV. Using the results in (iii), are there any outliers? Explain.

- b. Repeat (a) for the Wednesday evening market.

- 3.34 Many fast-food chains offer salads and low-fat options on their menu as an alternative to their traditional rolls and burgers. Data for a sample of these alternative and traditional menu items are stored in **[Dataset: HEALTHY\_FASTFOOD.XLS]**. For each product category, use the fat in grams per serve data:

- a. List the five-number summary.
- b. Construct the box-and-whisker plot.

- c. What similarities and differences are there in the distributions for the product categories?
- 3.40 The data file **[Dataset: MILK.XLSX]** gives nutrition content (number of calories and total fat, in grams) per 250 mL of a random sample of 20 fresh milks available in Australia.
- Calculate the covariance.
  - Calculate the coefficient of correlation.
  - Which do you think is more valuable in expressing the relationship between calories and fat content – the covariance or the coefficient of correlation? Explain.
  - What conclusions can you reach about the relationship between calories and fat content?

- 3.48 A quality characteristic of interest for a tea-bag-filling process is the weight of the tea in the individual bags. If the bags are under filled, two problems arise. First, customers may not be able to brew the tea as strong as they wish. Second, the company may be in violation of the truth-in-labelling laws. For this product, the label weight on the package indicates that, on average, there are 5.5 grams of tea in a bag. If the average amount of tea in a bag exceeds the label weight, the company is giving away product.

Getting an exact amount of tea into a bag is problematic because of variation in the temperature and humidity inside the factory, differences in the density of the tea, and the extremely fast filling operation of the machine (approximately 170 bags a minute). The table below provides the weight in grams of a sample of 50 tea-bags produced within an hour by a single machine. **[Dataset: TEABAGS.XLS]**

5.65	5.44	5.42	5.40	5.53	5.34	5.54	5.45	5.52	5.41
5.57	5.40	5.53	5.54	5.55	5.62	5.56	5.46	5.44	5.51
5.47	5.40	5.47	5.61	5.53	5.32	5.67	5.29	5.49	5.55
5.77	5.57	5.42	5.58	5.58	5.50	5.32	5.50	5.53	5.58
5.61	5.45	5.44	5.25	5.56	5.63	5.50	5.57	5.67	5.36

- Calculate the mean, median, first quartile and third quartile.
- Calculate the range, interquartile range, variance, standard deviation and coefficient of variation.
- Interpret the measures of central tendency and variation within the context of this problem. Why should the company producing the tea-bags be concerned about the central tendency and variation?
- Construct a box-and-whisker plot. Are the data skewed? If so, how?
- Is the company meeting the requirement set forth on the label that, on average, there are 5.5 grams of tea in a bag? If you were in charge of this process, what changes, if any, would you try to make concerning the distribution of weights in the individual bags?

#### TEXTBOOK REFERENCE:

Basic Business Statistics: Concepts and Applications. *Berenson, M.L. Levine, D.M. Szabat, K.A. O'Brien, M. Jayne, N. Watson, J.* 5th edition. 2019. Pearson Australia Group Pty Ltd. ISBN 9781488617249. Chapter 3, sections 3 to 3.6