



ECOM5005

Tutorial 3

Numerical Descriptive Measures

Preparation expectation. You are expected to prepare for the tutorial prior to your tutorial. If you cannot obtain an answer, demonstrate that you attempted the problem and bring specific questions to class.

Learning Outcomes

1. Describe, calculate and interpret numerical descriptive measures of central tendency, variation, and shape for numerical data.
2. Describe, calculate and interpret descriptive summary measures for a population.
3. Construct and interpret a box-and-whisker plot.
4. Calculate and interpret the covariance and the coefficient of correlation for bivariate data.

PART 1 - Short answer questions

Question 1.

Given below is the frequency distribution of the number of defective tomatoes in 50 different shipments.

Defective Items	Frequency
0 but less than 5	35
5 but less than 10	7
10 but less than 15	5
15 but less than 20	2
20 but less than 25	1
25 but less than 30	0

Referring to the table above, what is the approximate mean number of defective tomatoes?

Question 2.

For certain data sets, such as home prices and household or individual income, the median is often used instead of the mean. The questions below explore the mean and median in different situations to help you understand the information that they communicate.

(a) The brightness of celestial bodies depends on many factors, two of the most important being the distance from Earth and size. The eight brightest objects in the night sky are listed below with their approximate distance from Earth (in light years).

Object	Distance in light years
Moon	0.000000038
Venus	0.000048
Jupiter	0.000067
Mars	0.000076
Mercury	0.000095
Syrius	8.6
Canopus	310
Saturn	0.0014

Calculate the mean and median for these distances. Would the typical distance of these celestial bodies best be communicated using the mean or the median? Why?

(b) Suppose the mean annual income for a sample of one hundred Perth residents was \$50,000. Do you think the median income for this sample would have been greater than, equal to, or less than \$50,000? Explain.

Question 3.

A bank branch located in a commercial district of a city has developed an improved process for serving customers during the noon to 1pm lunch period. The waiting time in minutes (defined as the time the customer enters the line to when they reach the teller window) of all customers during this hour is recorded over a period of one week. A random sample of 15 customers is selected, and the results are as follows:

4.21, 5.55, 3.02, 5.13, 4.77, 2.34, 3.54, 3.20, 4.50, 6.10, 0.38, 5.12, 6.46, 6.19, 3.79

- Calculate the mean, median, first quartile and third quartile.
- Calculate the variance, standard deviation, range, interquartile range, coefficient of variation and Z scores. Are there any outliers? Explain.
- Are the data skewed? If so, how?
- As a customer walks into the branch office during the lunch hour, she asks the branch manager how long she can expect to wait. The branch manager replies, 'Almost certainly less than five minutes'. On the basis of the results of (a) and (b), evaluate the accuracy of this statement.

Question 4. Investment Performance and Risk - two managed funds

A financial adviser is comparing two diversified managed funds. The annual returns over the most recent six years are shown below. Returns are expressed as percentages.

Year	Fund A	Fund B
1	12%	8%
2	-8%	7%
3	18%	9%
4	5%	6%
5	10%	8%
6	3%	7%

- Calculate the arithmetic mean annual return for each fund.
- Calculate the geometric mean annual rate of return for each fund. Explain why the geometric mean is especially relevant for investment returns over time.
- Calculate the sample standard deviation and coefficient of variation for each fund.
- If \$10,000 had been invested at the beginning of Year 1 and all returns were reinvested, calculate the value at the end of Year 6 for each fund.
- Using return and risk together, write a short recommendation to a risk-conscious investor. Your explanation must refer to both the average return and variability.

Question 5. Digital Marketing Analytics - advertising spend and online sales

An online retailer wants to know whether weekly digital advertising expenditure is associated with online sales. The following eight weeks are observed. Both variables are measured in thousands of dollars.

Week	Advertising spend (\$000)	Online sales (\$000)
1	5	45
2	7	47
3	8	58
4	10	49
5	12	63
6	13	55
7	15	72
8	18	68

- Construct a scatter diagram with advertising spend on the horizontal axis and online sales on the vertical axis.
- Calculate the sample covariance between advertising spend and online sales.
- Calculate the sample coefficient of correlation, r .
- Interpret the direction and strength of the relationship in business language.

(e) Explain why a strong correlation does not prove that higher advertising expenditure causes higher sales. Identify at least two possible confounding factors.

PART 2 - Data exercise

Question 6. Beer calorie analysis in Microsoft Excel

The data in the file Beer.xlsx contain the calorie content per 100 ml for a sample of 95 beers together with beer type. Complete the analysis entirely in Microsoft Excel. Your workbook should be organised so that another person can reproduce every result.

Required Excel workflow

1. Inspect the Beer worksheet. Identify the variables, classify each variable, confirm the number of observations, check for missing values, and report the number of Regular, Light and Non-alcoholic beers.
2. For Calories/100ml, use Excel Data Analysis ToolPak -> Descriptive Statistics to obtain the standard summary output. Add the first quartile, third quartile, interquartile range and coefficient of variation using Excel formulas.
3. Calculate a Z-score for every beer. Use the Lecture 3 rule $|Z| > 3$ to flag any extreme outlier. Report the observation(s) and interpret what the Z-score means.
4. Construct a frequency distribution using the common class intervals 10 to <20, 20 to <30, ..., 80 to <90. Include frequency and percentage. Use this table to create a histogram with clearly labelled axes and an informative title.
5. Compare calorie content by beer type. Produce a table showing count, mean, median, sample standard deviation, coefficient of variation, minimum and maximum for Regular, Light and Non-alcoholic beers. Construct a box-and-whisker plot comparing the three types.
6. Write a short interpretation of the centre, spread, shape and unusual observations. Compare the three beer types and explain which type has the greatest relative variability.
7. Check that every table and figure follows good presentation practice: meaningful title, labelled axes, sensible number formats, consistent decimal places and no unnecessary chartjunk.

PART 3 - Pearson MyLab

1. Please ensure that you have finished the MyLab Quiz 1 which is due yesterday!

Note: If you are having issues obtaining access to MyLab, re-read the instructions available in the Welcome topic. If you are still unsuccessful, email the teaching team explaining the issue. Homework and Quiz completion are important because the final e-test uses MyLab.