MA208 Quantitative Techniques for Business

Lecture 3 Statistics ctd.

Dr Kirsten Pfeiffer

School of Mathematics, Applied Mathematics and Statistics
NUI Galway

Lectures 3/4 - Outline

- More measures of centre:
  - (Arithmetic) Mean
  - Harmonic mean
  - Geometric mean
  - Weighted mean
  - Mode
- Standard deviation
- Chebyshev's Inequality
- Coefficient of variation

Mean

Medians are a good choice to identify the centre of a skewed distribution. When we have a symmetric dustribution, the arithmetic mean (or just mean) is a good measure of the centre. We find the mean by adding up all of the data values and then dividing them by the number of data values.

The Mean

For a data set $\{x_1, x_2, \ldots, x_n\}$ the mean of the data set is

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Mean

Example

Calculate the mean of each of the exam score data set

$$\{77, 88, 92, 57, 55, 73, 64, 64, 89, 44, 76\}.$$

$$\bar{x} = \frac{77 + 88 + \ldots + 76}{11} = \frac{779}{11} \approx 70.82$$

The arithmetic mean is not always ideal. Consider the following:

Example

Alma drives 10 km at 60 km/h, and 30 km/h on the way back. What is her average speed?

10 km at 60 km/h → 10 min
10 km at 30 km/h → 20 min

20 km in 30 min
⇒ average 40 $\frac{km}{h}$

Harmonic Mean

We define a new mean, the harmonic mean as follows:

Harmonic Mean

$$\bar{x} = \left(\frac{\sum_{i=1}^{n} \frac{1}{x_i}}{n}\right)^{-1} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}$$

Example (ctd.)

Use formula. $x_1 = 60, \quad x_2 = 30$

$$\bar{x} = \left(\frac{\frac{1}{60} + \frac{1}{30}}{2}\right)^{-1} = \left(\frac{\frac{1}{20}}{2}\right)^{-1} = \left(\frac{1}{40}\right)^{-1} = \underline{\underline{40}}$$

Note: The harmonic mean is appropriate for situations when the average of **rates** is required.

Geometric Mean

Another type of mean is the geometric mean, which is the n^{th} root of the product of n numbers, i.e. for a set of numbers $\{x_1, x_2, ..., x_n\}$ the geometric mean is

Geometric Mean

$$\overline{x} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Example

The geometric mean of $\{3, 12\}$ is

$$\overline{x} = \sqrt{3 \cdot 12} = 6.$$

Note: A geometric mean is often used when comparing different items.

In this course we will generally only use the arithmetic mean, denoted by $\bar{x}$ or μ. This mean

- is familiar to most people,
- always exists,
- is unique,
- uses all the data.

However, in some cases the mean does not seem to give us an accurate value. Consider the following:

Which Mean?

Example

Consider two tutorial groups for the same maths course, one tutorial is held on Tuesdays, the other one on Fridays. We have the size and average test scores for both groups:

Tuesday tutorial:	20 students,	$\overline{x} = 80$
Friday tutorial:	30 students,	$\overline{x} = 90$

What is the average mark?

We notice that the arithmetic mean $\frac{80+90}{2} = 85$ does not account for the difference in number of students in each class.

Weighted Mean

We could add all individual results and divide by $20 + 30 = 50$ **or** use a weighted mean of all the classes as follows:

Example

$$\overline{x} = \frac{(20 \times 80)(30 \times 90)}{20 + 30} = 86$$

The weighted mean of a data set $\{x_1, x_2, ..., x_n\}$ with non-negative weights $\{w_1, w_2, ..., w_n\}$ is

Weighted Mean

$$\overline{x} = \frac{w_1 x_1 + w_2 x_2 + \cdot ... \cdot + w_n x_n}{w_1 + w_2 + \cdot ... \cdot + w_n} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Mode

The mode of a sample is the element that occurs most often in the collection.

Example

Determine the mode of the following sample set.

$$\{1, 3, 5, 5, 5, 5, 6, 6, 7, 7, 11, 11, 11, 19\}$$

Sln: mode = 5

- The mode is **not unique**.
- Unlike mean and median, the concept of mode also makes sense for **non-numerical** data. For example, taking a sample of Irish family names, one might find that "Murphy" occurs more often than any other name. Then "Murphy" would be the mode of the sample.

Example

Consider two patients A and B and their pulse.

Patient	Pulse	Mean
A	72, 76, 74	$\overline{x} = 74$
B	72, 91, 58	$\overline{x} = 74$

We note that the mean is the same, but we know that this is not the whole story, the difference of the variance matters.
We have introduced the midrange and the interquartile range (IQR). However, these measures for spread are not always ideal.

Example

Calculate the midrange and IQR for the following data sets.

1. $\{5, 18, 18, 18, 18, 18\}$
2. $\{5, 6, 7, 9, 10, 18\}$

(1) midrange $= \frac{18-5}{2} = \frac{13}{2} = 6.5$

IQR $= 18 - 18 = 0$

(2) midrange $= \frac{13}{2} = 6.5$

IQR $= 10 - 6 = 4$

Standard Deviation

A more powerful measure of spread is the standard deviation which takes into account how far each data is from the mean.

Given a set of data values $x_1, x_2, x_3, ..., x_n$, the standard deviation is

Standard Deviation

$$\sigma = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n-1}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

Standard Deviation

Exercise

A hospital records the following data of blood samples:

$$\{8, 11, 7, 13, 10, 11, 7, 9\}$$

Find the mean $\overline{x}$ and the standard deviation σ.

$$\bar{x} = \frac{8 + 11 + \ldots + 9}{8} \approx 9.5$$

σ:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
8	-1.5	2.25
11	1.5	2.25
7	-2.5	6.25
(...)	(...)	(...)
7	-2.5	6.25
9	-0.5	0.25
Total		32

$$\sigma = \sqrt{\frac{32}{8-1}} = \sqrt{\frac{32}{7}} \approx 2.14$$

Using Standard Deviation

One crude but simple tool for understanding standard deviation is the range rule of thumb, which is based on the principle that for many data sets, the vast majority (such as 95%) of sample values lie within two standard variations of the mean.

Range Rule of Thumb

We informally define *usual* values in a data set to be those that are typical and not too extreme. If the standard deviation of a data set is known, use it to find rough estimates of the minimum and maximum *usual* sample values as follows:

$$\begin{aligned}\text{minimum “usual” value} &= (\text{mean}) - 2 \times (\text{standard deviation}) \\ \text{maximum “usual” value} &= (\text{mean}) + 2 \times (\text{standard deviation})\end{aligned}$$

Example

The Wechsler Adult Intelligence Scale involves an IQ test designed so that the mean score is 100 and the standard deviation is 15. Use the range rule of thumb to find the minimum and maximum "usual" IQ scores. Then determine whether an IQ score of 135 would be considered "unusual".

Exercise for you

We'll discuss this on Friday. ☺