

UNIT 6: BUSINESS DECISION MAKING

1

WEEK SEVEN
LECTURER: N. QUARRIE

Learning Outcome Two (2)

2

- LO2 Understand a range of techniques to analyze data effectively for business purposes

Objective

3

- By the end of this lesson you should be able to:
- 2.4 explain how quartiles, percentiles and the correlation coefficient are used to draw useful conclusions in a business context



Overview

4

- Last week's session focused on how to calculate and interpret measures of dispersion, specifically, standard deviation. This week we want to take a look at how to calculate quartiles, percentiles and the correlation coefficient, as well as, draw conclusions from the results.
- It is important to note that in order to calculate the above, the data collected from the questionnaire must be summarized.

Quartiles: Definitions and General interpretation

- According (Web.mnstate.edu, 2016):
- “The **lower half** of a data set is the set of all values that are to the left of the median value when the data has been put into increasing order.
- The **upper half** of a data set is the set of all values that are to the right of the median value when the data has been put into increasing order.
- The **first quartile**, denoted by Q_1 , is the median of the *lower half* of the data set. This means that about 25% of the numbers in the data set lie below Q_1 and about 75% lie above Q_1 .
- The **third quartile**, denoted by Q_3 , is the median of the *upper half* of the data set. This means that about 75% of the numbers in the data set lie below Q_3 and about 25% lie above Q_3 .”

Quartiles: Example #1

Source: (Web.mnstate.edu, 2016)

6

- Find the first and third quartiles of the data set {3, 7, 8, 5, 12, 14, 21, 13, 18}.
- First, we write data in increasing order: 3, 5, 7, 8, 12, 13, 14, 18, 21.



Identify the mean, lower half and upper half as can be seen above.

We previously said that Q_1 is the median of the lower half and Q_3 .

$$Q_1 = \frac{5+7}{2} = \frac{12}{2} = 6$$

$$Q_3 = \frac{14+18}{2} = \frac{32}{2} = 16$$

Quartiles: Example #2

Source: (Web.mnstate.edu, 2016)

7

- Find the first and third quartiles of the set $\{3, 7, 8, 5, 12, 14, 21, 15, 18, 14\}$.
- First, we write the data in increasing order: 3, 5, 7, 8, 12, 14, 14, 15, 18, 21.
- As before, the median is 13 (it is the mean of 12 and 14 — the pair of middle entries).
- Therefore, the lower half of the data is: $\{3, 5, 7, 8, 12\}$.
- Notice that 12 is included in the lower half since it is below the median value.
- Then $Q_1 = 7$ (there are five values in the lower half, so the middle value is the median). Similarly, the upper half of the data is: $\{14, 14, 15, 18, 21\}$, so $Q_3 = 15$.

Percentiles: Example

Source: (Statistics, 2016)

8

- For example, suppose you have 25 test scores, and in order from lowest to highest they look like this: 43, 54, 56, 61, 62, 66, 68, 69, 69, 70, 71, 72, 77, 78, 79, 85, 87, 88, 89, 93, 95, 96, 98, 99, 99.
- To find the 90th percentile for these (ordered) scores, start by multiplying 90% times the total number of scores, which gives $90\% * 25 = 0.90 * 25 = 22.5$ (the index). Rounding up to the nearest whole number, you get 23.

Percentiles: Example

Source: (Statistics, 2016)

9

- Counting from left to right (from the smallest to the largest value in the data set), you go until you find the 23rd value in the data set. That value is 98, and it's the 90th percentile for this data set.
- Now say you want to find the 20th percentile. Start by taking $0.20 \times 25 = 5$ (the index); this is a whole number, so proceed from Step 3 to Step 4b, which tells you the 20th percentile is the average of the 5th and 6th values in the ordered data set (62 and 66). The 20th percentile then comes to $(62 + 66) \div 2 = 64$.
- The median (the 50th percentile) for the test scores is the 13th score: 77.

Correlation: Definition

10

- According to (Stattrek.com, 2016) “Correlation coefficients measure the strength of association between two variables. The most common correlation coefficient, called the Pearson product-moment correlation coefficient, measures the strength of the *linear association* between variables.”

Correlation: Formula

Source: (Psc.dss.ucdavis.edu, 2016)

11

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

Correlation: Example

Source: (Psc.dss.ucdavis.edu, 2016)

12

<u>X</u>	<u>Y</u>	<u>XY</u>	<u>X²</u>	<u>Y²</u>
5	5	25	25	25
10	20	200	100	400
6	4	24	36	16
8	15	120	64	225
4	11	44	16	121
4	9	36	16	81
3	12	36	9	144
10	18	180	100	324
2	7	14	4	49
6	2	12	36	4
7	14	98	49	196
9	17	153	81	289
<u>ΣX = 74</u>	<u>ΣY = 134</u>	<u>ΣXY = 942</u>	<u>ΣX² = 536</u>	<u>ΣY² = 1874</u>

Correlation: Example

Source: (Psc.dss.ucdavis.edu, 2016)

13

$$\begin{aligned} &= \frac{12(942) - (74)(134)}{\sqrt{[12(536) - (74)^2][12(1874) - (134)^2]}} \\ &= \frac{11304 - 9916}{\sqrt{(6432 - 5476)(22488 - 17956)}} \\ &= \frac{1388}{\sqrt{956(4532)}} = \frac{1388}{\sqrt{4,332,592}} = \frac{1388}{2081.49} = .667 \end{aligned}$$

Correlation: Interpretation

Source: (Kean.edu, 2016)

14

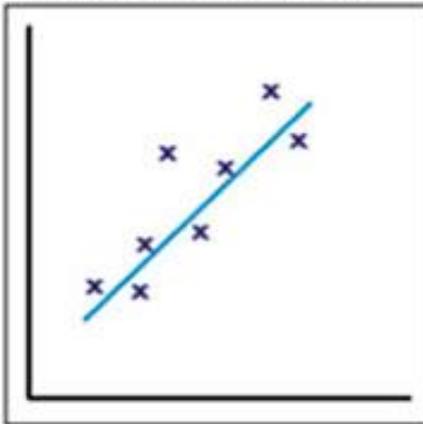
- Values of r range from -1 to $+1$. A correlation coefficient of 0 means that there is no relationship. A value of -1 is a perfect negative coefficient and a correlation value of $+1$ indicates a perfect positive correlation.

Correlation: Interpretation

(McLeod, 2016)

15

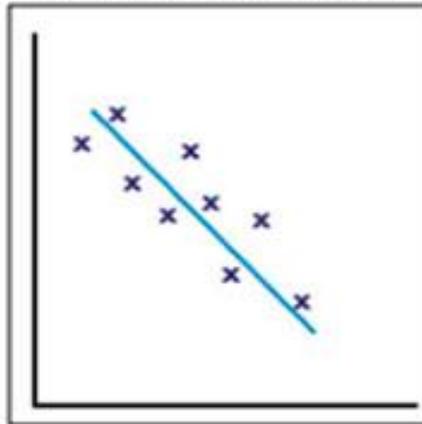
Positive correlation



The points lie close to a straight line, which has a positive gradient.

This shows that as one variable **increases** the other **increases**.

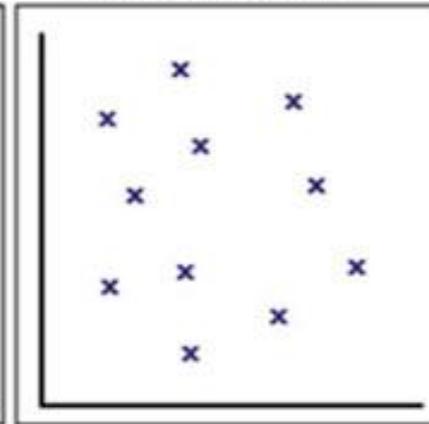
Negative correlation



The points lie close to a straight line, which has a negative gradient.

This shows that as one variable **increases**, the other **decreases**.

No correlation



There is no pattern to the points.

This shows that there is **no connection** between the two variables.

Videos

16

1. Calculating Correlation:

<https://www.youtube.com/watch?v=SC1kvvoH10Y>

2. Calculating Percentiles and Quartiles:

<https://www.youtube.com/watch?v=Snf6Wpn-L4c>

Review Questions

17

1. Explain the steps that should be followed whenever you are percentiles
2. a. Assume that you collected data from students regarding their age and grades received in a Statistics class that it is represented in the table below. Calculate the correlation.

X (age)	Grade
23	50
19	70
30	100
27	20
29	16
50	45

- 2b. Interpret the results from part “a”

References/Additional Reading List

- 1. Kean.edu, (2016). *Regression and Correlation*. [online] Available at: <http://www.kean.edu/~fosborne/bstat/09rc.html> [Accessed 8 Feb. 2016].
- 2. McLeod, S. (2016). *Correlation Method in Psychology | Simply Psychology*. [online] [Simplypsychology.org](http://www.simplypsychology.org). Available at: <http://www.simplypsychology.org/correlation.html> [Accessed 8 Feb. 2016].
- 3. Psc.dss.ucdavis.edu, (2016). *Pearson r: Hand calculation*. [online] Available at: http://psc.dss.ucdavis.edu/sommerb/sommerdemo/correlation/hand/pearson_hand.htm [Accessed 8 Feb. 2016].

References/Additional Reading List

4. Statistics, H. (2016). *How to Calculate Percentiles in Statistics - For Dummies*. [online] Dummies.com. Available at: <http://www.dummies.com/how-to/content/how-to-calculate-percentiles-in-statistics.html> [Accessed 8 Feb. 2016].
5. Stattrek.com, (2016). *Correlation: Definition*. [online] Available at: <http://stattrek.com/statistics/dictionary.aspx?definition=Correlation> [Accessed 8 Feb. 2016].
6. Web.mnstate.edu, (2016). *First Quartile and Third Quartile*. [online] Available at: <http://web.mnstate.edu/peil/MDEV102/U4/S36/S363.html> [Accessed 8 Feb. 2016].