Measures of central tendency (Measures of location)

Grouped data.

Mean of grouped data

Classes	Frequency	Midpoint (x_i)
A – B	f_1	A+B/2
B – C	f_2	B+C/2
C – D	f_3	C+D/2
⋮	⋮	⋮
K – H	f_K	K+H/2

Mean $$\boxed{\bar{X} = \frac{\sum_{i=1}^{K} x_i f_i}{\sum_{i=1}^{K} f_i}}$$

Ex. Calculate the mean of the given data.

Actual Classes	Frequency f_i	Midpoint x_i	$(x_i)(f_i)$
1) 9.5 – 19.5	4	14.5	58
2) 19.5 – 29.5	7	24.5	171.5
3) 29.5 – 39.5	12	34.5	414
4) 39.5 – 49.5	8	44.5	356
5) 49.5 – 59.5	3	54.5	163.5

$$\text{Sample Mean} = \frac{\sum_{i=1}^{5} x_i f_i}{\sum_{i=1}^{5} f_i} = \frac{1163}{34} \cong 34.21$$

Mode of grouped data.

Modal class is the class with largest frequency.
The approximate mode is the modal class center.

Ex. Find the mode of the following data.

Actual Classes	Frequency f_i
9.5 – 19.5	4
19.5 – 29.5	7
29.5 – 39.5	12
39.5 – 49.5	8
49.5 – 59.5	3

Soln. For this data, the only modal class is the third class, and the approximate mode is

$$M = \frac{29.5 + 39.5}{2} = 34.5 .$$

Exc. For the following data, construct a frequency table of 5 classes, and the approximate its mode.

12, 82, 43, 54, 65, 67, 87, 13, 24, 53, 54, 64, 76, 87, 90, 34, 54, 75, 88, 99, 66, 55, 44, 33, 22, 11, 90, 99. ← This is Ex. 1.7.9 in the text.

Median of grouped data.

$$\boxed{\text{Median} = 50^{th}\text{ lower limit} + \frac{\text{smallest change}}{\text{largest change}} \times \text{class width.}}$$

Ex. Calculate the median of the given data.

Actual Upper Class limit	Actual Class	Frequency	Cumulative Frequency
9.5	0 – 9.5	0	0
19.5	9.5 – 19.5	4	4
29.5	19.5 – 29.5	7	11
39.5	29.5 – 39.5	12	23
49.5	39.5 – 49.5	8	31
59.5	49.5 – 59.5	3	34

$$\underbrace{\overbrace{11 \longrightarrow}^{\text{smallest change in freq.}} \text{order} \longrightarrow 23}_{\text{largest change in freq.}}$$

Soln. Sample size is $n = 34$.

Order of 50^{th} percentile is $n \times \frac{50}{100} = \frac{n}{2} = \frac{34}{2} = 17$.

The order 17 lies between 11 and 23, and hence the median lies between the corresponding actual upper class limits 29.5 and 39.5.

50^{th} lower limit $= 29.5$

Largest change in frequency $= 23 - 11 = 12$.

smallest change in freq. $= 17 - 11 = 6$.

Corresponding class width $= 39.5 - 29.5 = 10$.

$$\text{Median} = 50^{th} \text{ lower limit} + \frac{\text{smallest change}}{\text{largest change}} * \text{class width}.$$

$$= 29.5 + \frac{6}{12} * 10$$

$$= 34.5.$$

Percentiles of grouped data.

$$\boxed{p^{th} \text{ percentile} = p^{th} \text{ lower limit} + \frac{\text{smallest change}}{\text{largest change}} * \text{class width}.}$$

First quartile $= 25^{th}$ percentile.

Second quartile $= 50^{th}$ percentile.

Third quartile $= 75^{th}$ percentile.

Ex. Given the following data

Actual Upper Class limit	Actual Class	Frequency	Cumulative Frequency
9.5	0 – 9.5	0	0
19.5	9.5 – 19.5	4	4
29.5	19.5 – 29.5	7	11
39.5	29.5 – 39.5	12	23
49.5	39.5 – 49.5	8	31
59.5	49.5 – 59.5	3	34

(1) Find the 70^{th} percentile.
(2) Find the 1^{st} quartile.
(3) Find the 2^{nd} quartile. Exc.
(4) Find the 3^{rd} quartile. Exc.

Soln. Sample size is $n = 34$.

(1) Order of 70^{th} percentile $= 34 * \frac{70}{100} = 23.8$.

70^{th} lower limit $= 39.5$

largest change in freq. $= 31 - 23 = 8$.

smallest change in freq. $= 23.8 - 23 = 0.8$

Class width $= 49.5 - 39.5 = 10$.

$$70^{th} \text{ percentile} = 70^{th} \text{ lower limit} + \frac{\text{smallest change}}{\text{largest change}} * \text{class width}.$$

$$= 39.5 + \left(\frac{0.8}{8}\right)(10) = 40.5.$$

---

(2) First quartile is the 25^{th} percentile.

Order of 25^{th} percentile $= 34 * \frac{25}{100} = 8.5$

25^{th} lower limit $= 19.5$

largest change in freq. $= 11 - 4 = 7$.

smallest change in freq. $= 8.5 - 4 = 4.5$

class width $= 10$

$$1^{st} \text{ quartile} = 19.5 + \frac{4.5}{7}(10) = 25.93.$$

---

(3) and (4) are left as an Exc.

Def. The interquartile range is the difference between the third and first quartiles. That is

$$IQR = Q_3 - Q_1$$

Ex: In the above example, we have

The 1^{st} quartile is $Q_1 = 25.93$

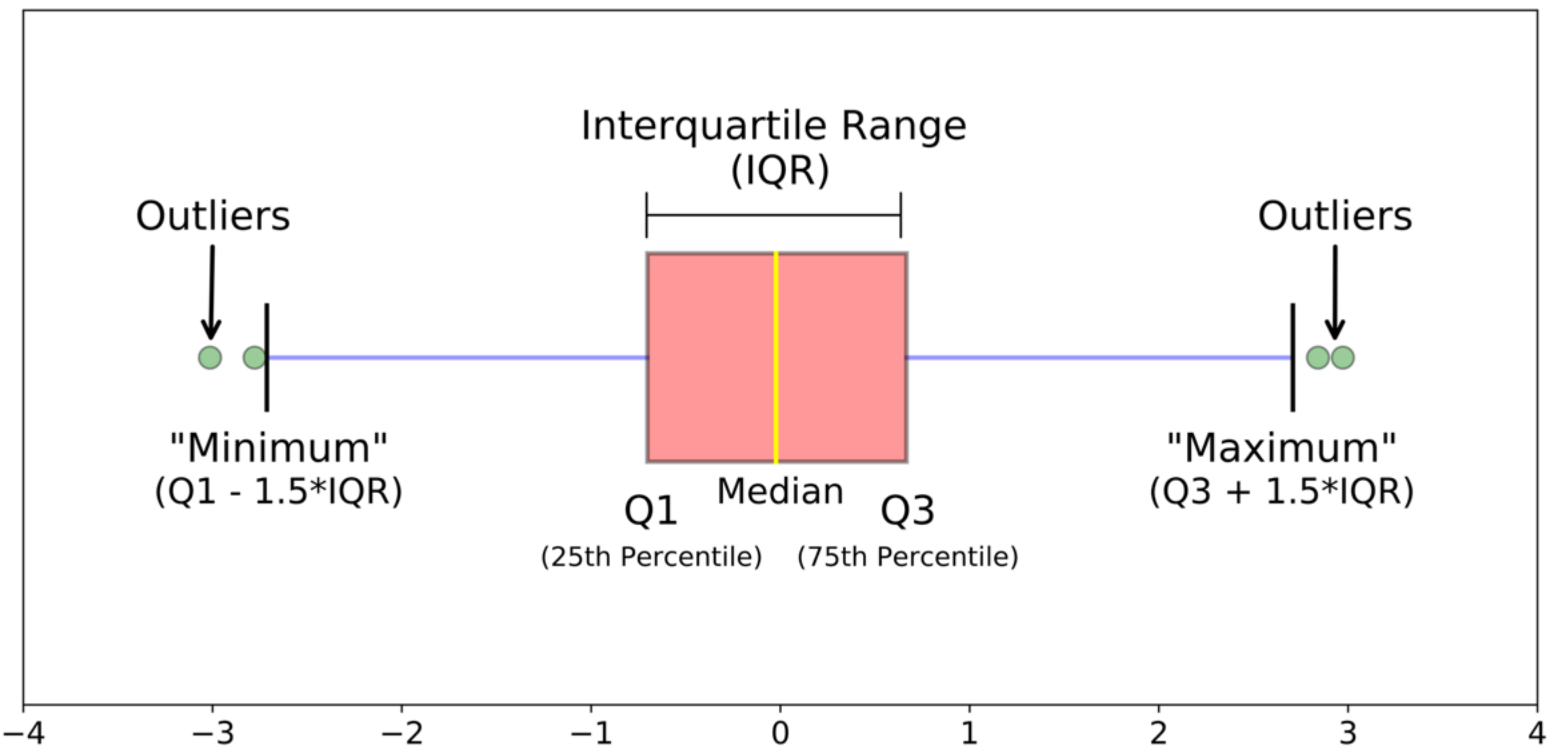
The 3^{rd} quartile is $Q_3 = 42.63$.

Then $IQR = Q_3 - Q_1 = 42.63 - 25.93 = 16.7$.

Def. An outlier is a data point that differs significantly from other observations. بعيدة، شاذة، متطرفة

If $x > Q_3 + 1.5\,IQR$ or $x < Q_1 - 1.5\,IQR$ then x is an outlier.

A Boxplot is a standardized way of displaying the distribution of data.



This picture was taken fromn https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51?gi=80721712d0d7

Searching keywords:

- Measures of location.
- Grouped data.
- Mean, median, mode.
- Percentiles.
- Interquartile range, outlier, boxplots.
- The University of Jordan الجامعة الأردنية
- Principles of Statistics مبادئ الإحصاء
- Baha Alzalg بهاء الزالق

References: See the course website

http://sites.ju.edu.jo/sites/Alzalg/Pages/131.aspx

For any comments or concerns, please use my email to contact me.

د. بهاء محمود الزالق
The University of Jordan
Dr. Baha Alzalg
baha2math@gmail.com