Sheet – 05: Boxplot

- The **pth percentile** of a dataset is the value below which a certain percent (p) of the observations fall.
- To find the location of the pth percentile in a dataset, you can follow these steps:
- **Step 1**: Arrange the data in ascending order.
- **Step 2**: Calculate the index for the pth percentile using the formula:
- **Index** = $\boldsymbol{p/100 * (n + 1)}$

where:

$\boldsymbol{p}$ is the desired percentile (between 0 and 100),

$\boldsymbol{n}$ is the number of data points in the dataset.

- **Step 3**: If the index obtained in Step 2 is an **integer**, **that is the position** of the pth percentile in the ordered dataset.
- **Step 4**: If the **index is not an integer**, round it down to the nearest whole number (j) and take the next number (k). The pth percentile is then estimated by **linear interpolation** between the jth and kth values in the ordered dataset.

- **Example**

For example, let's say we have a dataset with 7 values ($n = 7$), and we want to find the 30th percentile ($p = 30$).

Step 1: Arrange the data in ascending order. Let's say our ordered data is: **2, 4, 6, 8, 10, 12, 14**.

Step 2: Calculate the index: Index = $30/100 * (7 + 1) = 2.4$

Step 3: Since the index is not an integer, we round it down to the nearest whole number, which is 2 (j), and take the next number, which is 3 (k).

Step 4: The 30th percentile is then estimated by linear interpolation between the 2nd and 3rd values in the ordered dataset. So, the 30th percentile is $4 + 0.4 * (6 - 4) = 4.8$.

So, the 30th percentile of this dataset is 4.8.

- **Example**

Suppose we have a dataset with 10 values ($n = 10$), and we want to find the 75th percentile ($p = 75$).

Step 1: Arrange the data in ascending order. Let's say our ordered data is: **1, 2, 3, 4, 5, 6, 7, 8, 9, 10**.

Step 2: Calculate the index: Index = 75/100 * (10 + 1) = 8.25

Step 3: Since the index is not an integer, we round it down to the nearest whole number, which is 8 (j), and take the next number, which is 9 (k).

Step 4: The 75th percentile is then estimated by linear interpolation between the 8th and 9th values in the ordered dataset. So, the 75th percentile is $8 + 0.25 * (9 - 8) = 8.25$.

So, the 75th percentile of this dataset is 8.25. This method is consistent with the previous example and confirms the process for finding the pth percentile in a dataset.

---

1) What is **Q1** in the following data set: **1, 2, 3, 4, 5, 6, 7, 8, 9, 10**?
   (a) 2.75
   (b) 3.5
   (c) 4.5
   (d) 5.5
2) What is **Q2** in the following data set: **1, 2, 3, 4, 5, 6, 7, 8, 9, 10**?
   (a) 4.5
   (b) 5.5
   (c) 6.5
   (d) 7.5
3) What is **Q3** in the following data set: **1, 2, 3, 4, 5, 6, 7, 8, 9, 10**?
   (a) 6.5
   (b) 7.5
   (c) 8.25
   (d) 9.5
4) What is the **20th percentile** of the following data set: **1, 2, 3, 4, 5, 6, 7, 8, 9, 10**?
   (a) 2
   (b) 2.2
   (c) 3
   (d) 3.5
5) What is the **90th percentile** of the following data set: **1, 2, 3, 4, 5, 6, 7, 8, 9, 10**?
   (a) 8.5
   (b) 9
   (c) 9.5
   (d) 9.9

6) A computer science class has 10 students who scored the following marks in a test: **50, 60, 70, 80, 90, 100, 110, 120, 130, 140**. What is the first quartile (**Q1**) of the scores?
(a) 67.5
(b) 70
(c) 75
(d) 80

7) In a computer networking course, the time taken (in milliseconds) by 9 packets to travel from a server to a client was recorded as: 10, 20, 30, 40, 50, 60, 70, 80, 90. What is the median (Q2) time?
(a) 40
(b) 50
(c) 60
(d) 70

8) An electrical engineering student measures the resistance (in ohms) of 7 resistors: 10, 20, 30, 40, 50, 60, 70. What is the third quartile (Q3) of the resistances?
(a) 50
(b) 60
(c) 70
(d) 80

9) A computer scientist records the time (in seconds) it takes for an algorithm to sort 8 different arrays: 1, 2, 3, 4, 5, 6, 7, 8. What is the interquartile range (IQR) of the times?
(a) 3.5
(b) 4
(c) 4.5
(d) 5

10) In a digital electronics course, the voltage (in volts) of 5 different signals was measured as: 1, 2, 3, 4, 5. What is the 60th percentile of the voltages?
(a) 3
(b) 3.2
(c) 3.4
(d) 3.6

11) A computer science student runs a program 6 times, recording the following runtimes (in seconds): 10, 20, 30, 40, 50, 60. What is Q1 of the runtimes?
(a) 15
(b) 17.5
(c) 25
(d) 30

12) An electrical engineer measures the current (in amperes) in a circuit 6 times, getting the following measurements: 1, 2, 3, 4, 5, 6. What is Q2 of the measurements?
(a) 3
(b) 3.5
(c) 4
(d) 4.5

13) A computer scientist measures the time (in milliseconds) it takes for a function to execute 7 times: 10, 20, 30, 40, 50, 60, 70. What is Q3 of the times?
(a) 50
(b) 55
(c) 60
(d) 65

14) An electrical engineer measures the resistance (in ohms) of 5 resistors: 10, 20, 30, 40, 50. What is the IQR of the resistances?
(a) 20
(b) 25
(c) 30
(d) 35

15) In a computer networking course, the time taken (in milliseconds) by 5 packets to travel from a server to a client was recorded as: 10, 20, 30, 40, 50. What is the 80th percentile of the times?
(a) 40
(b) 42
(c) 44
(d) 48

16) A computer science student records the time (in seconds) it takes for an algorithm to sort 6 different arrays: 1, 2, 3, 4, 5, 6. What is Q1 of the times?
(a) 1.75
(b) 2
(c) 2.5
(d) 3

17) An electrical engineering student measures the current (in amperes) in a circuit 7 times, getting the following measurements: 1, 2, 3, 4, 5, 6, 7. What is Q2 of the measurements?
(a) 3
(b) 4
(c) 5
(d) 6

18) A computer scientist measures the time (in milliseconds) it takes for a function to execute 8 times: 10, 20, 30, 40, 50, 60, 70, 80. What is Q3 of the times?
(a) 62.5
(b) 65

(c) 67.5
(d) 70

19) An electrical engineer measures the resistance (in ohms) of 6 resistors: 10, 20, 30, 40, 50, 60. What is the IQR of the resistances?
(a) 25
(b) 30
(c) 35
(d) 40

20) In a digital electronics course, the voltage (in volts) of 6 different signals was measured as: 1, 2, 3, 4, 5, 6. What is the 70th percentile of the voltages?
(a) 4.1
(b) 4.2
(c) 4.3
(d) 4.9