Semester: V Subject: Statistics for AIDS Academic Year: 2023-2024

Estimates on Percentile:

What is percentile?

The value such that P percent of the values take on this value or less and (100-P) percent take on this value or more.

Difference between percentage and percentile.

Consider, Total No. of students = 50

Your marks = 480/750.

To find percentage, we calculate

$$= \frac{480}{750} \times 100$$

$$= 64\%$$

In this case there is no comparison with the class.

If suppose your rank = 9th.

To calculate percentile,

$$\text{Percentile} = \frac{\text{Total no. of students} < \text{than you} \times 100}{\text{Total students}}$$

$$= \frac{41}{50} \times 100$$

$$= 82\%\text{ile}$$

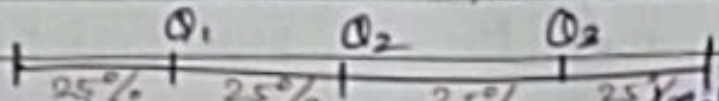
This means 82%ile of students have scored less marks than you. In this case, we can actually compare with the entire class.

Semester: V Subject: Statistics for AIDS Academic Year: 2023-2024

Quartiles:-

→ Quartiles divide a data set into 4 equal groups.

→ They are marked as Q_1, Q_2 and Q_3



——— 25th percentile

——— 50th percentile.

——— 75th percentile.

Example:-

Finding data values that correspond Q_1, Q_2, Q_3.

dataset:

14 18 15 6 21 9 16 12 23 7 11 19 24 5 31 10 27 17 25 8.

Arrange the data set smallest to largest

5 6 7 8 9 10 11 12 14 15 16 17 18 19 21 23 24 25 27 31

Q_1 (25%) $\frac{9+10}{2} = 9.5$

median $= \frac{15+16}{2} = 15.5$ Q_2.

50% percentile.

Q_2 (50%

Q_3 (75%)

$\frac{21+23}{2} = \frac{44}{2} = 22$

$Q_1 = 9.5$, $Q_2 = 15.5$, $Q_3 = 22$.

* 25%tile of data is less than 9.5
* 50%tile of data is less that 15.5 and more than 15.5.

Semester: IV Subject: Statistics for AIDS Academic Year: 20 - 20

*75%-tile of data less than 22.

Interquartile Range (IQR).

The difference between 3rd and 1st quartiles.

$IQR = Q_3 - Q_1$.

$IQR = 22 - 9.5 = 12.5$

Why we need IQR?

IQR is used to identify the outliers.

example:-

Using the same dataset of earlier, just consider that we add 170.

5 6 7 8 9 10 11 14 15 16 18 19 21 23 24 25 31 170 — 18

$Q_1 = 9.5$ $Q_2 = 15.5$ $Q_3 = 23.5$

$IQR = Q_3 - Q_1$.

$= 23.5 - 9.5 = 14$

multiply IQR by 1.5 = $14 \times 1.5 = 21$.

$Q_1 - 21 = -11.5$

$Q_3 + 21 = 23.5 + 21$

$= 44.5$

The Range is between (-11.5 to 44.5)

Any number outside this range can be considered as outlier.

In the above example, 170 is the outlier.

Note:- Outlier is any data value smaller than $Q_1 - 1.5(IQR)$ any data value larger than $Q_3 + 1.5(IQR)$.

Semester : IV Subject : Statistics for AIDS Academic Year: 2023-2024

Example:-

Find Q_1, Q_2 and Q_3 for the following data set. Identify the outliers.

{5, 40, 42, 46, 48, 49, 50, 50, 52, 53, 55, 56, 58, 75, 102}

Q_2

Solution:-

$Q_2 = 50$.

$Q_1 = 46$.

$Q_3 = 56$.

Interquartile Range (IQR) $= Q_3 - Q_2$.

$= 56 - 50$

$\boxed{IQR = 6}$.

$Q_1 - 1.5(IQR) = 46 - 1.5(6)$.

$\boxed{= 31}$

$Q_3 + 1.5(IQR) = 56 + 1.5(6)$

$= 71$.

Range $= (31 - 71)$.

Since 5 is less than 31 it is an outlier.

75, 102 is greater than 71, it is also an outlier.