Introduction to Applied statistics.

Reshmi Dey
17/10/2021

Median for continuous frequency distribution.

In case of continuous frequency distribution, the class corresponding to cumulative frequency juust greater than ½ N is called the median class, and the value of median is obtained by the following

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

Cumulative frequency - Cumulative frequency is used to determine the number o**f** observations that lie above (or below) a particular value in a data set. The cumulative frequency is calculated by adding each frequency from a frequency distribution table to the sum of its predecessors.

$$\text{Median} = l + \frac{h}{f}\left(\frac{N}{2} - c\right)$$

Here , l is the lower limit of the median class , f is the frequency of the median class, h is the magnitude of median class, and c is the c.f. of the class preceding the median class. And N is su
mmation of all frequencies.

Q. Find the median wages of the following distributions:

Wages (in Rs)	2,000 - 3,000	3,000-4,000	4,000-5,000	5,000-6,000	6,000-7,000
No. of workers	3	5	20	10	5

Solution:-

Wages (in Rs)	No. of workers	Cumulative frequency (c.f)
2,000 - 3,000	3	3
3,000 - 4,000	5	3+5= 8
4,000 - 5,000	20	8+20=28
5,000 - 6,000	10	28+10=38
6,000 - 7,000	5	38+5=43

Here , N = 43

N/2 = 21.5

Cumulative frequency just greater than 21.5 is 28 and the class corresponding to it is 4,000 - 5,000.

l= 4,000

h= 5,000 - 4,000 = 1,000

f = 20 and c = 8

Median = 4,000 +(1,000/20)(21.5 - 8) = 4,000 + (50*13.5) = 4,000 + 675 = 4,675

Therefore, median wage is Rs. 4,675

Quartiles, Deciles and Percentiles.

Q. Eight coins were tossed together and the number of heads resulting was noted. The operation was repeated 256 times and the frequencies (f) that were obtained for different values of x, the number of heads, are shown in the following table. Calculate the median quartiles, 4th deciles and 27th percentile.

x	0	1	2	3	4	5	6	7	8
f	1	9	26	59	72	52	29	7	1

Solution:-

x:	0	1	2	3	4	5	6	7	8
f:	1	9	26	59	72	52	29	7	1
c.f.:	1	10	36	95	167	219	248	255	256

For Median, N=256, N/2 = 128

Cumulative frequency just greater than 128 is 167 and the class corresponding is 4. Therefore Median = 4.

For Q1, N/4 = 256/4 = 64 and the c.f. Just greater than 64 is 95. Therefore Q1=3.

For Q3, (¾)*N= 192 and c.f. Just greater than 192 is 219. Thus Q3=5.

Now for 4th decile, D4 (4/10)*N = 102.4 and the c.f. Just greater than 102.4 is 167. Hence, D4 =4.

Lastly for 27th percentile, p27

(27/100)*N = 69.12

So, the c.f. Just greater than 69.12 is 95, which make p27 = 3.

Measures of Dispersion.

Consider the series (i) 2,3,4,5,6 (ii)8,2,5,1,4 (iii) 7,4,4,2,3. In all these cases we have mean 4 and the number of observation is 5. We see that we cannot form an idea whether the mean is of 1st series or 2nd or 3rd or any other series with number of observation as 5 and mean as 4. Thus we see that the measures of Central Tendency are inadequate to give us a complete idea of the distribution.

In such cases the central tendency must be supported by some other measures.

One such measures is Dispersion.

Dispersion basically mean scatteredness.

Definition:-

“ The measure of scatteredness of the mass of figures in a series about an average is called the measures of dispersion.” -Simpson and kafka.

“ The degree to which numerical data tend to spread about an average value is called dispersion of data.”

- -Spiegel.

Various measures of dispersion can be classified into two broad categories :-

(a) The measures which express the spread of observations in terms of distance, also termed as distance measures, e.g. range and quartile deviation.
(b) The measures which express the spread of observations in terms of the average of deviation of observations from some central value, e.g mean absolute deviation and standard deviation

Range

The difference between two extreme observations of the distribution.

Lets say A and B are the greatest and smallest value in a distribution, then range is given by

Range = A - B

Quartile Deviation

Quartile deviation or semi-interquartile range Q is given by:

$$Q = \frac{1}{2}(Q3 - Q1)$$

Quartile deviation uses 50% of the data so it is definitely a better measure than the range. But since it ignores the other 50% of the data, it cannot be regarded as a reliable measure.

Mean Absolute Deviation.

If xi | fi is the frequency distribution, where i = 1,2,3,...,n then mean deviation about M (M here can be mean,median or mode) is given by

$$\text{MAD about M} = 1/N \sum fi\ |Xi - M|\ ,$$

Here, i = 1,2,3,...,n.

| xi - M | represents absolute value of deviation (xi - M), where negative sign is ignored.

Since it uses all the data, it is a better measure than range or quartile deviation. But steps of ignoring the negative sign makes it inappropriate for further mathematical treatment.

Q. Calculate (i) Quartile deviation (ii) Mean deviation from mean for the following data

Marks:-	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of students:	6	5	8	15	7	6	3

Solution:-

Marks	Mid value(x)	No. of students(f)	|x - M| |x- 35|	f*|x-M|	c.f.
0-10	5	6	30	180	6
10-20	15	5	20	100	11
20-30	25	8	10	80	19
30-40	35	15	0	0	34
40-50	45	7	10	70	41
50-60	55	6	20	120	47
60-70	65	3	30	90	50

Solution :- (i) Here, N = 50;

For Q1, (¼)*N = 12.75

For Q3 (¾)* N = 37.25

The c.f. Just greater than 12.75 is 19. And the class corresponding is 20-30.

Q1 = 20 + 10/8 (12.75 - 11) = 22.19

The c.f. Just greater than 37.25 is 41. And the class corresponding is 40-50.

Q3 = 40 + .10/7 (37.25 - 34) = 44.64

Hence , Q.D. =(½)*(Q3 - Q1) = (½)(44.64 - 22.19) = 11.23.

(ii) Mean = 245/7 = 35, $\sum f|x - 35| = 640$

M.D. about mean = (1/N)640 = 640/50 = 12.8

Standard deviation

Standard deviation usually denoted by Greek letter small sigma(σ) is the positive square root of the arithmetic mean of the squares of the deviation of the given values from their arithmetic mean.

$$\sigma = \sqrt{1/N \sum fi(xi - mean)^2}$$

The steps of squaring overcomes the drawback of ignoring the negative sign. Hence standard deviation is suitable for further mathematical operation.

Variance

The square root of Standard deviation is called variance and is given by

$$\sigma^2 = 1/N \sum fi(xi - mean)^2$$

Standard deviation can also be written as

$$SD = \sqrt{1/N \sum fi(xi - mean)^2}$$

Also knowm as Root mean square deviation.

Q. Calculate the mean and standard deviation for the following table giving the age distribution of 542 members

Age (in years):-	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of members	3	61	132	153	140	51	2

Solution:-

Age group	No. of students (f)	Xi - mean	$(xi - mean)^2$	Mid point (x)
20-30	3	43.57	1898.34	25
30-40	61	33.57	1126.94	35
40-50	132	23.57	555.54	45
50-60	153	13.57	184.14	55
60-70	140	3.57	12.74	65
70-80	51	6.43	41.34	75
80-90	2	16.43	269.94	85

N = 542

Mean = 480 / 7 = 68.57

Variance = σ^2 = 1/N $fi(xi - mean)^2$

$fi(xi - mean)^2$ = 180374.9

σ^2 = 1/ 542 *(180374.9) = 332.795

S.D. = $\sqrt{\ }$ σ^2 = 18.24

$fi(xi - mean)^2$
5695.14
68743.34
73331.28
28173.42
1783.6
2108.34
539.88