MU - Jan 21

* Sums on Binning *.

Suppose a group of sales price records has been sorted as follows : 6, 9, 12, 13, 15, 25, 50, 70, 72, 92, 204, 232. Partition them into three bins by equal - frequency (Equi - depth) partitioning method. Perform data smoothing by bin mean.

Solution :- Sort the given data.

6, 9, 12, 13, 15, 25, 50, 70, 72, 92, 204, 232.

2. Smooth the data by equal frequency bins.

Given three bins. There are total 12 observations. Hence, by equi - deth partitioning method, each bin will have 4 observations.

- Bin 1 : 6, 9, 12, 13.
- Bin 2 : 15, 25, 50, 75
- Bin 3 : 72, 92, 204, 232.

2. Smooth the data by bin mean.

→ We take average of each bin and replace each data value by mean value in corresponding bin.

- Bin 1 : 10, 10, 10, 10.
- Bin 2 : 40, 40, 40, 40
- Bin 3 : 150, 150, 150, 150.

* Problems on binning. *

② Partition the given data into 4 bins using Equi-depth binning method and perform smoothing according to the following methods, smoothing by bin mean, smoothing by bin median, smoothing by bin boundaries.

Data : 11, 13, 13, 15, 15, 16, 19, 20, 20, 20, 21, 21, 22, 23, 24, 30, 40, 45, 45, 45, 71, 72, 73, 75

Solution :- Partition into equal depth bins (N=4).

→ (Total 24 Hence in each bin 6 elements in each bin)

{ if dataset is not in ascending order then arrange it in ascending order }.

Bin 1 : 11, 13, 13, 15, 15, 16

Bin 2 : 19, 20, 20, 20, 21, 21.

Bin 3 : 22, 23, 24, 30, 40, 45

Bin 4 : 45, 45, 71, 72, 73, 75.

→ Smoothing by bin means : (Replace each value of bin with mean value)

Bin 1 : 13.83, 13.83, 13.83, 13.83, 13.83, 13.83.

Bin 2 : 20.16, 20.16, 20.16, 20.16, 20.16, 20.16.

Bin 3 : 30.66, 30.66, 30.66, 30.66, 30.66, 30.66

Bin 4 : 63.5, 63.5, 63.5, 63.5, 63.5, 63.5.

→ smoothing by bin boundaries :-

{ In this method each value is replaced with its nearest value either minimum or maximum (boundaries).

Note:-

Bin 1 : 11, 11, 11, 16, 16, 16

Bin 2 : 19, 19, 19, 19, 21, 21

Bin 3 : 22, 22, 22, 22, 45, 45

Bin 4 : 45, 45, 75, 75, 75, 75.

* Smoothing by bin median

{Note – Replace each value with median.

- Bin 1 : 14, 14, 14, 14, 14, 14.
  Bin 2 : 20, 20, 20, 20, 20, 20.
  Bin 3 : 27, 27, 27, 27, 27, 27.
  Bin 4 : 71.5, 71.5, 71.5, 71.5, 71.5, 71.5

③ Divide the given data into bins & size of 3 by bin partition (equal frequency), by bin means, by bin medians and by bin boundaries.

Consider the data : 10, 2, 19, 18, 20, 18, 25, 28, 22.

Solution :- Here data is not in ascending order.

So sort the data first.

Hence, 2, 10, 18, 18, 19, 20, 22, 25, 28.

Now Partition the data into 3 equal bins.

Bin 1 : 2, 10, 18
Bin 2 : 18, 19, 20
Bin 3 : 22, 25, 28.

→ Smoothing by bin mean

Bin 1 : 10, 10, 10
Bin 2 : 19, 19, 19.
Bin 3 : 25, 25, 25

→ Smoothing by bin median

Bin 1 : 10, 10, 10
Bin 2 : 19, 19, 19
Bin 3 : 25, 25, 25

Smoothing by bin boundary

Bin 1 : 2, 2, 18
Bin 2 : 18, 18, 20
Bin 3 : 22, 22, 28.

④ Suppose a group of sales price records has been sorted as follows: 6, 8, 12, 13, 15, 25, 50, 70, 72, 92, 204, 232.

Partition them into

For the given attribute AGE values : 16, 16, 180, 4, 12, 24, 26, 28. Apply the following Binning technique for smoothing the noise.

i) Bin median ii) Bin Boundaries iii) Bin means.

Soln :- Sort the data set in ascending order.

4, 12, 16, 16, 24, 26, 28, 180

→ Now partition into (equal depth) bins = (N=2.)

Bin 1 : 4, 12, 16, 16.

Bin 2 : 24, 26, 28, 180

Smoothing by bin median

Bin 1 : 14, 14, 14, 14.

Bin 2 : 27, 27, 27, 27.

Smoothing by bin mean.

Bin 1 : 12, 12, 12, 12.

Bin 2 : 64.5, 64.5, 64.5, 64.5

→ Smoothing by bin boundaries.

Bin 1 : 4, 16, 16, 16.

Bin 2 : 24, 24, 24, 180.