

equal depth 4, 8, 9  
21, 21, 29

equal frequency.

Page  
Date

$\begin{bmatrix} 9 \\ 8 \end{bmatrix}$ 
 $\begin{bmatrix} 17 \\ 15 \end{bmatrix}$ 
 $\begin{bmatrix} 21 \\ 20 \end{bmatrix}$

- ① sorted data
- ② make bins according to bin size.
- ③ Arrange and partition data in equal depth bins.

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

⇒ smoothing by bin means:

$$\frac{4+8+9}{3} = \frac{21}{3} = 7.$$

Bin 1: 7, 7, 7.

$$\text{Bin 2: } \frac{21+21+24}{3} = \frac{66}{3} = 22.$$

$$\text{Bin 3: } \frac{26+28+29}{3} = 27.$$

② Bin by boundaries:

Bin ①: 4, 9, 9.

↓

↓

↓

same. 8 is closest value 9.

Bin ②

21, 21, 24

Bin ③

26, 29, 29



$$\frac{n+1}{2} \Rightarrow \text{odd.}$$

$$\frac{n}{2} + \left(\frac{n}{2} + 1\right)$$

Page  
Date

Smoothing by bin median.

Bin 1: 8, 8, 8

Bin 2: 21, 21, 21

Bin 3: 28, 28, 28

24, 26, 28, 29

$$\frac{4}{2} + \left(\frac{4}{2} + 1\right)$$

10, 19, 18, 20, 18, 25, 28, 22  
Bin size: 3

Sorted: 2, 10, 18, 18, 19, 20, 22, 25, 28

Bin 1: 2, 10, 18

Bin 2: 18, 19, 20

Bin 3: 22, 25, 28

Smoothing by bin means.

Bin 1: 10, 10, 10

Bin 2: 19, 19, 19

Bin 3: 25, 25, 25

① Bin by boundaries:

Bin 1: 2, 18, 18

Bin 2: 18, 20, 20

Bin 3: 22, 28, 28

② Smoothing by bin median.

Bin 1: 10, 10, 10

Bin 2: 19, 19, 19

Bin 3: 25, 25, 25



Bin = 4

②

6, 9, 12, 13, 15, 25, 50, 70, 72, 92, 204,  
232, 250, 257, 87

sort

1 2 3 4 5 6 7 8 9 10 11  
6, 9, 12, 13, 15, 25, 50, 70, 72, 87, 92,  
204, 232, 250, 257  
12 13 14 15

Bin 1 6, 9, 12, 13

Bin 2 15, 25, 50, 70

Bin 3 72, 87, 92, 204

Bin 4 232, 250, 257

Smoothing by bin means

Bin: 1 10, 10, 10, 10.  $\frac{6+9+12+13}{4} = \frac{40}{4} = 10$

Bin: 2 40, 40, 40, 40

Bin: 3 ~~113, 113, 113, 113~~ 114, 114, 114, 114

Bin: 4 246, 246, 246. ~~184, 184, 184~~  
185, 185, 185, 185

Bin by boundaries

Bin 1: 6 ~~40~~ 13 } 6, 6, 13, 13  
Bin 2: 15 ~~38~~ 70 } 15, 15, 70, 70  
Bin 3: 72 ~~90~~ 204 } 72, 72, 72, 204  
Bin 4: 232 ~~257~~ 257 } 232, 257, 257

~~Bin by smoothing by median~~

Bin 1: 10, 10, 10, 10

Bin 2: 37, 37, 37, 37

Bin 3: 89, 89, 89, 89

Bin 4: 250, 250, 250, 250



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

Transform Point =

$$V' = \frac{(V - \min)}{(\max - \min)} * (\text{newmax} - \text{newmin}) + \text{newmin}$$

$$V' = \frac{(V - \min)}{(\max - \min)} * (\text{newmax} - \text{newmin}) + \text{newmin}$$

13, 15, 16, 16, 19, 20, 23, 29, 35, 41, 44, 53, 62, 69, 72.

Given range (0, 1).

$V =$  data to be transform: 45

$\min$  data  $\min$ : 13

$\Rightarrow$  range: 0

new max: 1

$$V' = \frac{(V - \min)}{(\max - \min)} * (\text{newmax} - \text{newmin}) + \text{newmin}$$

$$= \frac{(45 - 13)}{(72 - 13)} * (1 - 0) + 0$$

$$= \frac{32}{59} * 1 + 0$$

$$= 32 + 91 * 1 + 0$$

$$= 92.91$$

$$0.5423 * 1 + 0$$

$$= 0.5423$$



$$\frac{(v - \min) * (\text{new max} - \text{new min})}{(\text{max} - \min)} + \text{new min}$$

200, 300, 400, 600, 1000

min = 0, max = 1.

given range 0 - 1.

data transform  $v \rightarrow 300$

new min  $\rightarrow 0$ .

min = 200

new max  $\rightarrow 1$ .

max = 1000

$$\frac{(300 - 200) * (1 - 0) + 0}{(1000 - 200)} \quad \bigg| \quad \frac{(400 - 200) * (1 - 0) + 0}{(1000 - 200)}$$

$$\frac{100 * 1}{800}$$

$$0.125$$

$$0.25$$

### \* Z score Normalisation

$$Z = (x - \mu) / \sigma$$

x = value for which z score is calculated.

$\mu$  = mean

$\sigma$  = standard deviation.

mean salary = 54,000

standard deviation = 16,000

x = 73,600.

$$Z = (x - \mu) / \sigma$$

$$= \frac{73,600 - 54,000}{16,000}$$

$$16,000$$

$$\frac{19,600}{16,000}$$

$$= 1.225$$



Error of the mean.

$$z = \frac{(x - \mu)}{(\sigma / \sqrt{n})}$$

Mean salary 65,000

$$x = 85,800$$

$$\sigma = 18000$$

$$z = \frac{(x - \mu)}{\sigma}$$

$$= \frac{85,800 - 65,000}{18000}$$

$$z = 1.155$$

$$\mu = \frac{8 + 10 + 15 + 20}{4}$$

$$= \frac{53}{4}$$

$$\mu = 13.25$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

$$= \sqrt{\frac{1}{4} \sum_{i=1}^4 (8 - 13.25)^2}$$

$$= 2.5$$

8, 10, 15, 20

$$z = \frac{(x - \mu)}{(\sigma / \sqrt{n})}$$

$$= \frac{(20 - 8)}{6}$$

$$= \frac{12}{6}$$

$$z = 2$$

$$(x - \mu)^2$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

$$\sigma = \sqrt{\frac{1}{4} \sum_{i=1}^4 (20 - 8)^2}$$

$$\sigma = \sqrt{\frac{1}{4} \sum_{i=1}^4 144}$$

$$\sigma = \sqrt{36}$$

$$\sigma = 6$$



$$\mu = \frac{8 + 10 + 15 + 20}{4}$$

$$\mu = \frac{53}{4}$$

$$\mu = 13.25$$

$$\sigma = 4.65$$

$$z = \frac{(x - \mu)}{\sigma}$$

$$= \frac{(10 - 13.25)}{4.65}$$

$$= -0.698$$

$$\frac{15 - 13.25}{4.65}$$

$$= 0.376$$

$$\frac{20 - 13.25}{4.65}$$

$$= 1.45$$

$$\frac{(v - \min)}{(\max - \min)} * (\text{newmax} - \text{newmin}) + \text{newmin}$$

$$\frac{(57 - 13)}{(72 - 13)} * (1)$$

$$\frac{44}{59}$$

$$0.745$$

$$0.745$$

Page \_\_\_\_\_  
Date \_\_\_\_\_

13, 15, 16, 16, 19, 20, 23, 29, 35, 41, 44, 53, 62, 69, 72

Bin size = 5

Bin 1: 13, 15, 16

Bin 2: 16, 19, 20

Bin 3: 23, 29, 35

Bin 4: 41, 44, 53

Bin 5: 62, 69, 72

\* Bin by mean:

Bin 1: 14.67 = 15

Bin 2: 18.33 = 19

Bin 3: 29 = 29

Bin 4: 46 = 46

Bin 5: 67.67 = 68

\* bin by median:

Bin 1: 15, 15, 15

Bin 2: 19, 19, 19

Bin 3: 29, 29, 29

Bin 4: 44, 44, 44

Bin 5: 69, 69, 69

\* Bin by boundaries.

Bin 1: 13, 16, 16

Bin 2: 16, 20, 20

Bin 3: 23, 35, 35

Bin 4: 41, 41, 53

Bin 5: 62, 72, 72



10, 12, 11, 25, 25, 27, 35, 34, 34, 31, 43,  
36, 50, 59, 33.

mean, median, mode,  
standard deviation, variance

10, 11, 12, 25, 25, 27, 31, 33, 34, 34, 35, 36, 43  
50, 59

mean.

$$\text{mean} = \frac{440}{15} = 29.33 \Rightarrow 31$$

$$\text{Median} = 33$$

variance

mode, 25, 34

standard deviation.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - \mu)^2}$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{15}$$

$$\sigma^2 = 52.26$$

$$\frac{(10-31)^2}{15}$$

$$\cdot \frac{\sum (36-31)^2 (43-31)^2}{151.15}$$

$$1.66 + 9.6 + 24.06$$

$$+ 52.26$$

$$= 87.58$$



2.7, 3.1, 3.4, 4.1, 4.3, 4.7, 4.7, 40.8, 2.9, 3.7

⇒ sorting

1 2 3 4 5 6 7 8 9 10  
2.7, 2.9, 3.1, 3.4, 3.7, 4.1, 4.3, 4.7, 4.7, 40.8

$$\Rightarrow \text{mean} = \frac{74.4}{10}$$

$$\text{mean} = 7.44 \\ \approx 7.4$$

$$\Rightarrow \text{median} = \frac{3.7 + 4.1}{2} \\ = 3.9 \\ \approx 4$$

$$\Rightarrow \text{mode} = 4.7$$

⇒ standard deviation-

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - \mu)^2}$$

$$= \sqrt{\frac{1}{10}}$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n}$$

$$= (2.7 - 7)^2 + (2.9 - 7)^2 + (3.1 - 7)^2 + (3.4 - 7)^2 \\ + (3.7 - 7)^2 + (4.1 - 7)^2 + (4.3 - 7)^2 + (4.7 - 7)^2$$



3-7

$$(4.7 - 7)^2 + (10.8 - 7)^2$$

$$= 18.49 + 16.81 + 15.21 + 12.96 +$$

$$22.09 + 20.61 + 18.83 + 16.32 + 13.98$$

$$8.41 + 11.15 + 9.85 + 7.5076$$

$$\sigma = 11.14 \quad \frac{+ 7.5076 + 1112.88}{10}$$

$$\sigma^2 = 124.1144$$

$$\begin{array}{cccccccc} 44 & 50 & 38 & 96 & 42 & 47 & 40 & 39 & 46 & 50 \\ 5 & 9 & 1 & & 4 & 7 & 3 & 2 & 6 & 8 \end{array}$$

$$\begin{array}{cccccccc} 38 & 39 & 40 & 42 & 44 & 46 & 50 & 50 & 96 \\ 1 & 2 & 3 & 4 & & 4 & 3 & 2 & 1 \end{array}$$

$$\text{mean} :- \frac{492}{10} = 49.2$$

$$\text{median} :- \frac{44 + 46}{2} = 45$$

$$\text{mode} :- 50$$

$$\text{Standard deviation} :-$$

$$\begin{array}{r} 125.44 + 104.04 \\ + 84.64 + 51.84 \\ 27.09 + 10.24 \\ 4.84 + 0.64 + \\ 0.64 + 2190.24 \\ \hline 10 \end{array}$$

$$2599.6$$

$$\sigma = \sqrt{\frac{2599.6}{10}} = 16.12$$

$$\begin{aligned} & \frac{(-44 - 49.2)^2 + (50 - 49.2)^2 + (38 - 49.2)^2 + (39 - 49.2)^2 + (40 - 49.2)^2 + (42 - 49.2)^2}{10} \\ & + (44 - 49.2)^2 + (46 - 49.2)^2 + (47 - 49.2)^2 \\ & + (50 - 49.2)^2 + (50 - 49.2)^2 + (96 - 49.2)^2 \end{aligned}$$



44, 50, 38, 96, 42, 47, 40, 39, 46, 50

bin size = 3

38, 39, 40, 42, 44, 46, 47, 50, 50, 96

Bin 1: 4

Bin 1: 38, 39, 40

Bin 2: 42, 44, 46

Bin 3: 50, ~~50~~, ~~96~~, 47, 50, 50, 96

Bin by mean.

Bin 1: 39

Bin 2: 44

Bin 3: ~~189~~  $\frac{243}{3} = 81$   $\frac{243}{4} = 60.75$

Bin by boundaries.

Bin 1: 38, 40, 40

Bin 2: 42, 46, 46

Bin 3: 47, 47, 47, 96

Bin by median.

Bin 1: 39, 39, 39

Bin 2: 44, 44, 44

Bin 3: 50, 50, 50, 50



Bin 1: 38, 39, 40, 42

Bin 2: 44, 46, 47, 50

Bin 3: 50, 96

Bin by mean.

Bin 1: 39.75  $\approx$  40

Bin 2: 46.75  $\approx$  47

Bin 3: ~~36.5~~  $\approx$  37

Bin by boundaries.

Bin 1: 38, 38, 38, 42

Bin 2: 44, 44, ~~44~~, 50

Bin 3: 50, 96, 50

Bin by median:

Bin 1: 39.5

Bin 2: 46.5

Bin 3: 73, 73