1. **Solution**: (time spent 1 hour**)**

Partition into (equal-freq) Mean

Bin 1: 11, 13, 15, 17, 19 75/5= 15

Bin 2: 21, 21, 23, 23, 23 111/5=22.2

Bin 3: 23, 25, 27, 30, 33 138/5=27.6

Bin 4: 33, 33, 33, 36, 36 171/5= 34.2

Bin 5: 38, 40, 46, 48. 54 226/5= 45.2

Smoothing by bin means:

Bin 1: 15, 15, 15, 15, 15

Bin 2: 22.2, 22.2, 22.2, 22.2, 22.2

Bin 3: 27.6, 27.6, 27.6, 27.6, 27.6

Bin 4: 34.2, 34.2, 34.2, 34.2, 34.2

Bin 5: 45.2, 45.2, 45.2, 45.2, 45.2

**Illustration:** Interval range of values in each bin is constant which is 5, the depth (frequency) of each bin.

1. **Solution:** (time spent 2 hour**)**

Group of data (marks): 100, 200, 4000, 700, 1100

1. **Min-max normalization by setting min = 0 and max = 1:**

(new\_maxA – new\_minA) + minA.

Mean of the data is 500.

**Data after normalization**

**For 100** =>(1 – 0) + 0 = **0**

**For 200** =>(1 – 0) + 0 =  **= 0.1**

**For 400** =>(1 – 0) + 0 =  **= 0.3**

**For 700** =>(1 – 0) + 0 =  **= 0.6**

**For 1100** =>(1 – 0) + 0 =  **= 1**

1. **Z-score normalization:**

Standard deviation of the data, = 406.20

**Data after normalization**

**For 100, =**> = **- 1.101**

**For 200, =>**  = **- 0.826**

**For 400, =>**  = - **0.275**

**For 700, =>**  = **0.550**

**For 1100, =>**  = **1.651**

**(c) Z- score normalization with mean absolute deviation:**

Mean absolute deviation for data are, :

= = 320

**After Data normalization**

**For 100, =**> = **- 1.25**

**For 200, =>**  = **- 0.9375**

**For 400, =>**  = - **0.3125**

**For 700, =>**  = **0.625**

**For 1100, =>**  = **1.875**

1. **Normalization by decimal scaling:**

The number of decimal points moved depends on the maximum absolute value.

**In this case,**

= 0.01

= 0.02

= 0.04

= 0.07

= 0.11

**3. Solution:** (time spent 1.5 hour**)**

(a) min-max normalization:

(new\_maxA – new\_minA) + minA.

**Data after normalization**

(1.0 – 0.0) + 0.0 = **0.326**

(b )

Mean for the data is 28.84 & Standard deviation is 10.8653. So,

**=**> = **- 0.3534**

(c )

The number of decimal points moved depends on the maximum absolute value.

**In this case,**

25/100 = 0.25

(d ) **Comment**:

I prefer to use normalization by decimal scaling because the transformation maintains the data distribution, making interpretation easy to communicate.

Min-Max normalization has the unexpected impact of not allowing any future values to fall outside of the current minimum and maximum values without encountering an “out of bounds error”. As such values may be present in future data, so this method is a little inappropriate.

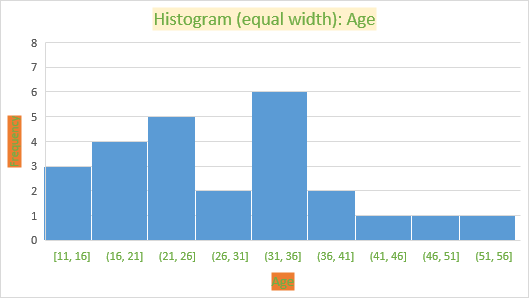
Z-score normalization may not be robust enough against outliers as the deviations are squared and may still be influenced by the outliers. Both min-max and Z-score normalization change the original data quite a bit compared to Normalization by decimal scaling.

1. Solution:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| age | %fat |  | age | |  | %fat |  |  |
| 20 | 8.4 |  |  | |  |  |  |  |
| 22 | 25.3 |  |  | |  |  |  |  |
| 25 | 7.6 |  |  |
| 25 | 18.8 |  |  |
| 36 | 27.5 |  | 13.97837837 | | Standard Deviation | 8.775486691 |  |  |
| 40 | 24.6 |  | 0.855386657 | | Correlation Coefficient |  |  |  |
| 45 | 28.1 |  | 104.9277778 | | Covariance |  |  |  |
| 48 | 28.8 |  | 20 | | Minium | 7.6 |  |  |
| 49 | 30.2 |  | 37 | | Quartile Q1 | 25.85 |  |  |
| 51 | 32.7 |  | 50 | | Quartile Q2 | 30 |  |  |
| 53 | 40.2 |  | 57.75 | | Quartile Q3 | 33 |  |  |
| 53 | 29.8 |  | 62 | | Maximum | 40.2 |  |  |
| 57 | 32.3 |  |  | |  |  |  |  |
| 58 | 30.7 |  |  | |  |  |  |  |
| 59 | 33.9 |  | 195.3950617 | | Variance | 77.00916667 | 15047.21087 | Cov(x,y) |
| 60 | 40.1 |  | 13.97837837 | | Standard Deviation | 8.775486691 | 122.6670733 | SxSy |
| 61 | 33.1 |  |  | |  |  |  |  |
| 62 | 36.4 |  |  | | COV(x,y) |  |  |  |
|  |  |  | r= | | SxSy |  |  |  |
|  |  |  |  | |  |  |  |  |
|  |  |  |  | | 122.6670733 |  |  |  |
|  |  | **Correlation coefficient =** | | | **0.855386657** |  |  |  |

**5. Solution:** (time spent 1 hour**)**

**Histogram Chart 1:**



**Histogram Chart 2:**

