Nama : Ferza Reyaldi

NIM : 09021281924060

Tugas 2 Data Mining

Buku Edisi ke-2 (Halaman 98)

2.2. Suppose that the values for a given set of data are grouped into intervals. The intervals and corresponding frequencies are as follows.

Age	Frequency
1-5	200
5-15	450
15-20	300
20-50	1500
50-80	700
80-110	44

Compute an approximate median value for the data.

Answer:

misalkan n = banyak data

L_1 = batas bawah interval median

width = lebar interval median

$freq_{median}$ = frekuensi dari interval median

$(\sum freq)_1$ = frekuensi kumulatif interval sebelum interval median.

$n = 200 + 450 + 300 + 1500 + 700 + 44$

$= 3194.$

median berada pada suku ke-i, $i = \frac{3194}{2} = 1597$. Sehingga median terletak pada interval 20-50.

$(\sum freq)_1 = 200 + 450 + 300$

$= 950.$

$width = 50 - 20 = 30.$

$freq_{median} = 1800$

$L_1 = 20.$

Sehingga, $median = L_1 + \left(\frac{i - (\sum freq)_1}{freq_{median}} \right) \cdot width$

$= 20 + \left(\frac{1597 - 950}{1500} \right) \cdot 30$

$= 20 + \frac{647.30}{1500} = 20 + 12,94 = 32,94$ //

KIKY

2.4 Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order)

13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

(a) what is the mean of the data? what is the median

(b) what is the mode of the data? comment on the data's modality.

(c) what is the midrange of the data?

(d) Can you find (roughly) the first quantile (Q_1) and the third quantile (Q_3) of the data?

(e) Give the five-number summary of the data.

(f) Show a boxplot of the data.

(g) How is a quantile-quantile plot different from a quantile plot?

Answer:

misalkan n adalah banyak data, $n = 27$.

(a) mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{27}(13+15+16+16+19+20+20+21+22+22+25+25+25+25+30+33+33+35+35+35+35+36+40+45+46+52+70)$$

$$= \frac{809}{27}$$

$$= 29,9696\ldots \approx 30.$$

median: posisi median $= i = \left(\frac{n+1}{2}\right) = \frac{27+1}{2} = 14.$

$$\text{median} = X_{14}$$

$$= 25.$$

(b) Terdapat 2 modus pada data, sehingga kelompok data tersebut adalah bimodal.

Modusnya adalah 25 dan 35.

(c) midrange $= \frac{\text{data terkecil} + \text{data terbesar}}{2}$

$$= \frac{X_1 + X_{14}}{2}$$

$$= \frac{13+70}{2}$$

$$= 41,5$$

(d) Q_1 adalah median dari setengah pertama data

Q_3 adalah median dari setengah akhir data.

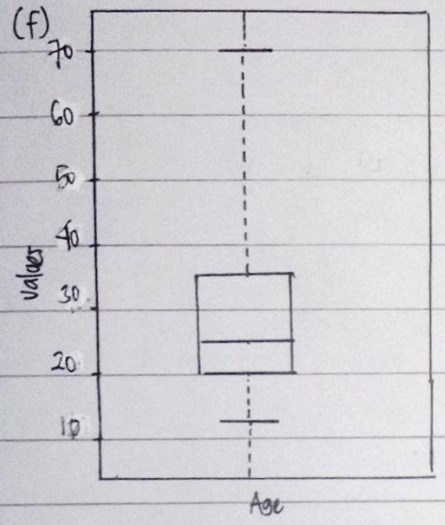
posisi $Q_1 = \frac{27+1}{4} = 7$, $Q_1 = X_7 = 20$

Posisi $Q_2 = \frac{3(27+1)}{4} = 21$. $Q_3 = X_{21} = 35$.

KIKY

(e) five-number summary adalah nilai terkecil, Q_1, median, Q_3, dan nilai terbesar.
Sehingga five-number summary dari data adalah 13, 20, 25, 35, 70.

(f)



(g) - Quantile plot digunakan untuk menunjukkan persentase nilai di bawah atau sama dengan variabel independen dalam distribusi univariat.

- Sedangkan Q-Q plot digunakan untuk membandingkan kuantil dari satu distribusi univariat terhadap kuantil yang bersesuaian dari distribusi univariat lainnya. Garis $y=x$ dapat ditambahkan untuk memperkaya informasi. Jika titik berada diatas garis, maka nilai quantil pada sumbu y lebih tinggi dibandingkan nilai kuantil yang di plot pada sumbu x, dan sebaliknya.