TUGAS 2 DATA MINING

Seuby Ayu Saputry
18105010001
Universitas Tidar

2.2 Age : 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) $\bar{x} = \frac{809}{27} = 29,96296$ Me $\Rightarrow \frac{27+1}{2} = 14 \Rightarrow$ me = 25

(b) mode = 25 & 35 (bimodal)

(c) midrange = $\frac{70-13}{2} = 28,5$

(d) $Q_1 \Rightarrow (27+1) \times \frac{1}{4} = 7 \Rightarrow 20$

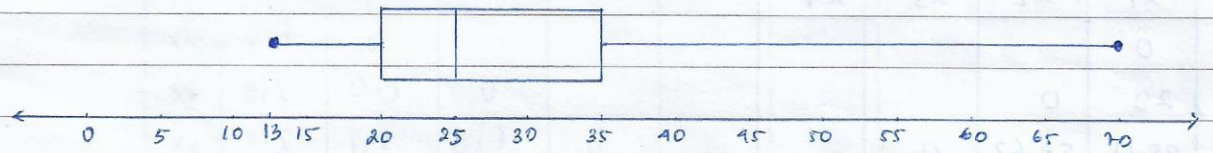
$Q_3 \Rightarrow (27+1) \times \frac{3}{4} = 21 \Rightarrow 35$

(e) five - number summary

min, Q_1, me, Q_3, max

13, 20, 25, 35, 70

(f) Boxplot



2.4

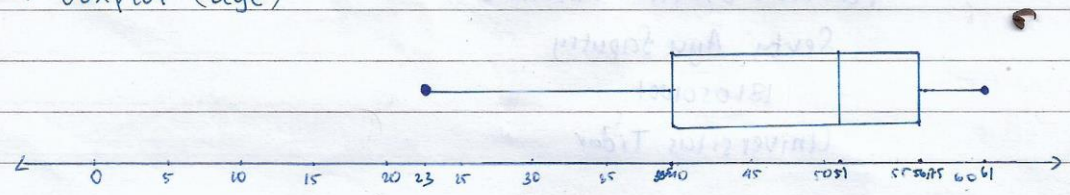
age	23	23	27	27	39	41	47	49	50	52	54	54	56	57	58	58	60	61
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

(a)

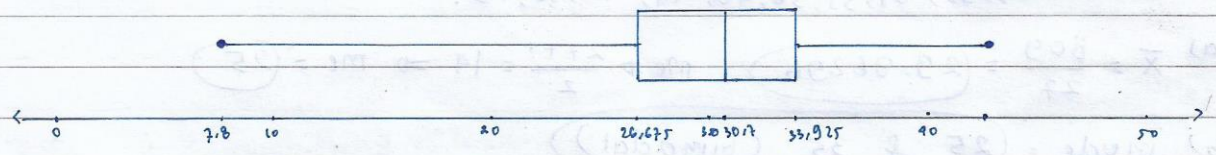
age	% fat
◦) $\bar{x} = \frac{836}{18} = 46,4$	◦) $\bar{x} = \frac{518,1}{18} = 28,78$
◦) me $\Rightarrow \frac{18+1}{2} = \frac{no.18 + no.19}{2} \Rightarrow \frac{50+52}{2} = 51$	◦) me $\Rightarrow \frac{18+1}{2} \Rightarrow \frac{30,2+31,2}{2} = 30,7$
◦) $s = \sqrt{\frac{2970,4}{17}}$	◦) $s = \sqrt{\frac{1459,325}{17}}$
$= 13,22$	$= 9,265$
(b) $Q_1 \Rightarrow 39,5$	$Q_1 \Rightarrow 26,675$
$Q_3 \Rightarrow 56,75$	$Q_3 \Rightarrow 33,925$

GELATIK

•) boxplot (age)



•) boxplot (%fat)



2.6 (22, 1, 42, 10) and (20, 0, 36, 8)

point	a1	a2
x_1	22	20
x_2	1	0
x_3	42	36
x_4	10	8

(a) Euclidean distance

L	x_1	x_2	x_3	x_4
x_1	0			
x_2	29	0		
x_3	25,61	54,62	0	
x_4	16,97	12,09	42,52	0

(b) Manhattan distance

L	x_1	x_2	x_3	x_4
x_1	0			
x_2	41	0		
x_3	36	77	0	
x_4	24	17	40	0

2.8

	A1	A2
x_1	1.5	1.7
x_2	2	1.9
x_3	1.6	1.8
x_4	1.2	1.5
x_5	1.5	1.0

(a) Query d (1,4, 1,6)

Euclidean

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0					
x_2	0,54	0				
x_3	0,14	0,41	0			
x_4	0,36	0,89	0,5	0		
x_5	0,7	1,03	0,81	0,58	0	
x_6	0,14	0,67	0,21	0,17	0,61	0

Manhattan

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0					
x_2	0,7	0				
x_3	0,2	0,5	0			
x_4	0,5	1,2	0,7	0		
x_5	0,7	1,4	0,9	0,8	0	
x_6	0,2	0,9	0,4	0,3	0,7	0

Supremum

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	0					
x_2	0,5	0				
x_3	0,1	0,4	0			
x_4	0,3	0,8	0,4	0		
x_5	0,7	0,9	0,8	0,5	0	
x_6	0,1	0,6	0,2	0,2	0,6	0

d_1 (1,5 ; 2 ; 1,6 ; 1,2 ; 1,5 ; 1,4)

d_2 (1,7 ; 1,9 ; 1,8 ; 1,5 ; 1,0 ; 1,6)

$d_1 \cdot d_2 = 2,55 + 3,8 + 2,88 + 1,8 + 1,5 + 2,20 = 14,77$

$\|d_1\| = \sqrt{2,25 + 4 + 2,56 + 1,44 + 2,25 + 1,96} = 3,803$

$\|d_2\| = \sqrt{2,89 + 3,61 + 3,24 + 2,25 + 1 + 2,56} = 3,943$

$sim(d_1, d_2) = \frac{14,77}{3,803 \cdot 3,943}$

$= \frac{14,77}{14,997}$

$= 0,99$

GELATIK

(b) $d(3,1) = 0,14$ dan $d(6,1) = 0,14$