**Ôn thi Khai phá dữ liệu**

Đề 1:

Ảnh có chứa văn bản

Mô tả được tạo tự động

a. Dữ liệu X:

= = = 6.81

Sắp xếp dãy: 4.5 5.3 5.5 6.0 6.5 6.7 7 7 7 8.5 8.6 9.1

Dãy chẵn => Median = = = 6.85

Mode X = 7.0

Dữ liệu Y:

= = 7.33

Sắp xếp dãy: 5.5 6.0 6.0 6.5 6.8 7.2 7.5 8.0 8.5 8.5 8.5 9.0

Dãy chẵn => Median = = = 7.35

Vì 8.5 xuất hiện 3 lần => Mode X = 8.5

b.Boxplot Y

Số phần tử quan sát: 12

Min = 5.5 Max = 9.0

Q1 = =(6.0 + 6.5)/2 = 6.25

Q2 = Median = 7.35

Q3 = = (8.5 + 8.5)/2 = 8.5

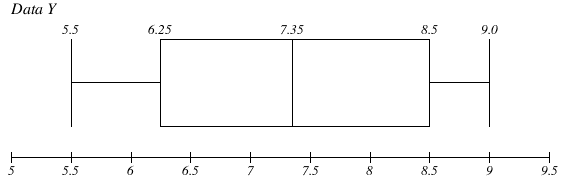
IQR = Q3 – Q1 = 2.25

L = Q1 – 1.5 x IQR = 6.25 – 1.5 x 2.25 = 2.875

U = Q3 + 1.5 x IQR = 8.5 + 1.5 x 2.25 = 11.875

Ngoại lai: không có.

Biểu đồ Boxplot:



c. Chuẩn hoá Y về z-score

=

= 1.15

v’ =

v’1 = = = -0.1

|  |  |  |
| --- | --- | --- |
| v’2 = 0.66 | v’6 = -1.32 | v’10 = -0.82 |
| v’3 = 1.16 | v’7 = 0.17 | v’11 = -0.52 |
| v’4 = 1.16 | v’8 = 1.16 | v’12 = 1.65 |
| v’5 = -1.81 | v’9 = -1.32 |  |

d. Trơn: trung bình(bin means), chia thùng: chiều rộng(Equal-width) với số thùng là 4.

Sắp xếp dãy: 5.5 6.0 6.0 6.5 6.8 7.2 7.5 8.0 8.5 8.5 8.5 9.0

Độ lệch (9.0 – 5.5)/4 = 0.875

Bin 1: [5.5, 6.375]: 5.5 6.0 6.0

Bin 2: [6.375, 7.25]: 6.5 6.8 7.2

Bin 3: [7.25, 8.125]: 7.5 8.0

Bin 4: [8.125, 9]:8.5 8.5 8.5 9.0

Trung bình các bin:

Bin 1 = (5.5 + 6.0 + 6.0)/3 = 5.83

Bin 2 = (6.5 + 6.8 + 7.2)/3 = 6.83

Bin 3 = (7.5 + 8.0)/2 = 7.75

Bin 4 = (8.5 + 8.5 + 8.5 + 9.0)/4 = 8.625

=> Làm trơn dữ liệu theo bin means ta được:

Bin 1: 5.83 5.83 5.83

Bin 2: 6.83 6.83 6.83

Bin 3: 7.75 7.75

Bin 4: 8.625 8.625 8.625 8.625

e. Mối tương quan giữa X và Y

= = = 0.73

=> >0 => x và y là tương quan dương.

Ảnh có chứa bàn

Mô tả được tạo tự động

min\_sup = 60% => min\_sup\_count = 60% x 5 = 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập C1 | | >=3 | Tập L1 | |
| Items | Sub\_count | Items | Sub\_count |
| I1 | 3 | I1 | 3 |
| I2 | 3 | I2 | 3 |
| I3 | 4 | I3 | 4 |
| I4 | 2 | I5 | 3 |
| I5 | 3 |  |  |
| I6 | 2 |  |  |
| I7 | 1 |  |  |

L1 x L1 được tập C2:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập C2 | | >=3 | Tập L2 | |
| Items | Sub\_count | Items | Sub\_count |
| I1, I2 | 2 | I1, I3 | 3 |
| I1, I3 | 3 | I2, I3 | 3 |
| I1, I5 | 2 | I2, I5 | 3 |
| I2, I3 | 3 | I3, I5 | 3 |
| I2, I5 | 3 |  |  |
| I3, I5 | 3 |  |  |

L2 x L2 được tập C3:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập C3 | | >=3 | Tập L3 | |
| Items | Sub\_count | Items | Sub\_count |
| I1, I3, I2 | 2 | I2, I3, I5 | 3 |
| I1, I3, I5 | 1 |  |  |
| I2, I3, I5 | 3 |  |  |

=> Tập phổ biến là: { I2, I3, I5}

min\_conf = 75%

Từ tập phổ biến là: { I2, I3, I5}, ta có các tập: {I2}, {I3}, {I5}, {I2, I3}, {I2, I5}, {I3, I5}

Các luật có dạng X^Y => Z với min\_conf là:

{I2, I3} => {I5} => conf = 3/3 = 100%

{I2, I5} => {I3} => conf = 3/3 = 100%

{I3, I5} => {I2} => conf = 3/3 = 100%

Vậy các luật kết hợp mạnh cần tìm là: {I2, I3} => {I5}, {I2, I5} => {I3}, {I3, I5} => {I2}

Đề 2:

Ảnh có chứa văn bản

Mô tả được tạo tự động

a. Dữ liệu X:

= = 6.81

Sắp xếp dãy: 4.5 5.3 5.5 6.0 6.5 6.7 7 7 7 8.5 8.6 9.1

Dãy chẵn => Median = = 6.85

Mode X = 7.0

Dữ liệu Y:

= = 7.33

Sắp xếp dãy: 5.5 6.0 6.0 6.5 6.8 7.2 7.5 8.0 8.5 8.5 8.5 9.0

Dãy chẵn => Median = = 7.35

Mode X = 8.5

b.Boxplot X

Sắp xếp dãy:

Số phần tử quan sát: 12

Min = 4.5 , Max = 9.1

Q1 = (5.5 + 6)/2 = 5.75

Q2 = Median = 7.35

Q3 = (7 + 8.5)/2 = 7.75

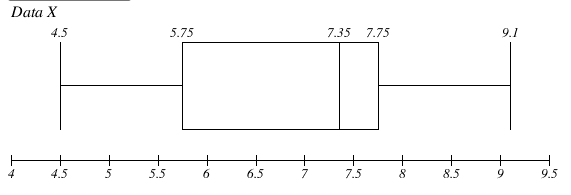
IQR = Q3 – Q1 = 7.75 – 5.75 = 2

L = Q1 – 1.5 x IQR = 5.75 – 1.5 x 2 = 2.75

U = Q3 + 1.5 x IQR = 7.75 + 1.5 x 2 = 10.75

Ngoại lai: không có.

Biểu đồ Boxplot:



c. Chuẩn hoá X về z-score

= = 1.33

v’ =

v’1 = = = 0.14

|  |  |  |
| --- | --- | --- |
| v’2 = 1.27 | v’6 = -1.35 | v’10 = -0.6 |
| v’3 = 1.35 | v’7 = -0.08 | v’11 = -0.98 |
| v’4 = 1.72 | v’8 = -0.23 | v’12 = 0.14 |
| v’5 = -1.74 | v’9 = 0.14 |  |

d. Làm trơn theo biên (bin boundary), chia thùng: chiều sâu (Equal-frequency)

Sắp xếp dãy: 4.5 5.3 5.5 6.0 6.5 6.7 7 7 7 8.5 8.6 9.1

Chia thùng theo chiều sâu

=> Số phần tử mỗi bin là : 12/4 = 3

Bin 1: 4.5, 5.3, 5.5

Bin 2: 6.0, 6.5, 6.7

Bin 3: 7, 7, 7

Bin 4: 8.5, 8.6, 9.1

Làm trơn dữ liệu theo biên ta được:

Bin 1: 4.5, 5.5, 5.5

Bin 2: 6.0, 6.7, 6.7

Bin 3: 7, 7, 7

Bin 4: 8.5, 8.5, 9.1

Ảnh có chứa bàn

Mô tả được tạo tự động

Khởi tạo tâm:

Cụm 1:{A,B,D} => Tâm T1(2+4+6/3, 2+4+6/3) = (4, 4)

Cụm 2: {F,G} => Tâm T2(0+4/2, 0+4/2) = (2, 2)

Cụm 3: {C, E} => Tâm T3(5+9/2, 5+9/2) = (7, 7)

Euclid: d(i, j) =

Lặp 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(4,4) | T2(2,2) | T3(7,7) | Cụm |
| A(2, 2) | 2.83 | 0 | 7.07 | 2 |
| B(4, 4) | 0 | 2.83 | 4.24 | 1 |
| C(5, 5) | 1.41 | 4.24 | 2.83 | 1 |
| D(6, 6) | 2.83 | 5.66 | 1.41 | 3 |
| E(9, 9) | 7.04 | 9.9 | 2.83 | 3 |
| F(4, 0) | 4 | 2.83 | 7.62 | 2 |
| G(0, 4) | 4 | 2.83 | 7.32 | 2 |

Cụm 1: {B, C} => Tâm T1(4+5/2, 4+5/2) = (4.5, 4.5)

Cụm 2: {A,F,G} => Tâm T2(2+4+0/3, 2+0+4/3) = (2, 2)

Cụm 3: {D,E} => Tâm T3(6+9/2, 6+9/2) = (7.5, 7.5)

Lặp 2:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(4.5, 4.5) | T2(2,2) | T3(7.5, 7.5) | Cụm |
| A(2, 2) | 3.53 | 0 | 7.78 | 2 |
| B(4, 4) | 0.71 | 2.83 | 4.95 | 1 |
| C(5, 5) | 0.71 | 4.24 | 3.53 | 1 |
| D(6, 6) | 2.83 | 5.66 | 2.83 | 1 |
| E(9, 9) | 6.36 | 9.9 | 2.83 | 3 |
| F(4, 0) | 4.53 | 2.83 | 8.23 | 2 |
| G(0, 4) | 4.53 | 2.83 | 8.23 | 2 |

Cụm 1: {B, C, D} => Tâm T1(4+5+6/3, 4+5+6/3) = (5, 5)

Cụm 2: {A,F,G} => Tâm T2(2+4+0/3, 2+0+4/3) = (2, 2)

Cụm 3: {E} => Tâm T3(9, 9)

Lặp 3:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(5, 5) | T2(2,2) | T3(9, 9) | Cụm |
| A(2, 2) | 4.24 | 0 | 9.9 | 2 |
| B(4, 4) | 1.41 | 2.83 | 7.07 | 1 |
| C(5, 5) | 0 | 4.24 | 5.66 | 1 |
| D(6, 6) | 4.24 | 5.66 | 4.24 | 1 |
| E(9, 9) | 5.66 | 9.9 | 0 | 3 |
| F(4, 0) | 5.1 | 2.83 | 10.29 | 2 |
| G(0, 4) | 5.1 | 2.83 | 10.29 | 2 |

Ta thấy các cụm của lặp 3 không thay đổi so với lặp 2 => Thuật toán dừng

Vậy: Cụm 1: {B, C, D} có tâm T1(5, 5)

Cụm 2: {A,F,G} có tâm T2(2, 2)

Cụm 3: {E} có tâm T3(9, 9)

Đề 3:

Ảnh có chứa văn bản

Mô tả được tạo tự động

a, b, c, e Làm ở trên

d. Làm trơn theo trung bình (bin means), chia thùng theo chiều sâu (Equal-depth)

Chia thùng theo chiều sâu => Số phần tử mỗi bin là: 12/4 = 3

Bin 1: 4.5, 5.3, 5.5

Bin 2: 6.0, 6.5, 6.7

Bin 3: 7, 7, 7

Bin 4: 8.5, 8.6, 9.1

Giá trị trung binh mỗi bin là:

Bin 1: (4.5+5.3+5.5)/3 = 5.1

Bin 2: (6+6.5+6.7)/3 = 6.4

Bin 3: (7+7+7)/3 = 7

Bin 4: (8.5+8.6+9.1)/3 = 8.73

Làm trơn dữ liệu theo trung bình (bin means), ta được:

Bin 1: 5.1, 5.1, 5.1

Bin 2: 6.4, 6.4, 6.4

Bin 3: 7, 7, 7

Bin 4: 8.73, 8.73, 8.73

Ảnh có chứa văn bản

Mô tả được tạo tự động

|  |  |  |  |
| --- | --- | --- | --- |
| ID | Items | ID | Items |
| T1 | L, J, C | T6 | L, C, P, B |
| T2 | P, B, L | T7 | P, B, C |
| T3 | L, C, S | T8 | J, C |
| T4 | P, B, S | T9 | L, C, B, P |
| T5 | J, S, C | T10 | J, S |

min\_sup = 20% => min\_sup\_count = 20% x 10 = 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập C1 | | >=2 | Tập L1 | |
| Items | Sub\_count | Items | Sub\_count |
| L | 5 | L | 5 |
| J | 4 | J | 4 |
| C | 7 | C | 7 |
| P | 5 | P | 5 |
| B | 5 | B | 5 |
| S | 4 | S | 4 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập C2 | | >=2 | Tập L2 | |
| Items | Sub\_count | Items | Sub\_count |
| L, J | 1 | L, C | 4 |
| L, C | 4 | L, P | 3 |
| L, P | 3 | L, B | 3 |
| L, B | 3 | J, C | 3 |
| L, S | 1 | J, S | 2 |
| J, C | 3 | C, P | 3 |
| J, P | 0 | C, B | 3 |
| J, B | 0 | C, S | 2 |
| J, S | 2 | P, B | 5 |
| C, P | 3 |  |  |
| C, B | 3 |  |  |
| C, S | 2 |  |  |
| P, B | 5 |  |  |
| P, S | 1 |  |  |
| B, S | 1 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập C3 | | >=2 | Tập L3 | |
| Items | Sub\_count | Items | Sub\_count |
| L, C, B | 2 | L, C, B | 2 |
| L, C, P | 2 | L, C, P | 2 |
| J, C, S | 1 | C, P, B | 2 |
| C, P, B | 2 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập C4 | | >=2 | Tập L4 | |
| Items | Sub\_count | Items | Sub\_count |
| L, C, B, P | 2 | L, C, B, P | 2 |

=> Tập phổ biến là {L, C, B, P}

b. min\_conf = 75%

Từ tập phổ biến là {L, C, B, P} ta có các tập con: {L}, {C}, {B}, {P}, {L, C}, {L, B}, {L, P}, {C, B}, {C, P}, {B, P}, {L, C, B}, {L, C, P} {L, B, P}, {C, B, P}.

Ta có các luật kết hợp với conf là:

|  |  |  |
| --- | --- | --- |
| {L} => {C, B} | {L} => {C} | {L} => {C, P, B} |
| {L} => {C, P} | {L} => {B} | {C} => {L, B, P} |
| {L} => {B, P} | {L} => {P} | {B} => {L, C, P} |
| {C} => {L, B} | {C} => {L} | {P} => {L, C, B} |
| {C} => {L, P} | {C} => {B} | Bỏ @@ |
| {C} => {B, P} | {C} => {P} |  |
| {B} => {L, C} | {B} => {L} |  |
| {B} => {L, P} | {B} => {C} |  |
| {B} => {C, B} | {B} => {P} |  |
| {P} => {L, C} | {P} => {L} |  |
| {P} => {C, P} | {P} => {B} |  |
| {P} => {C, B} | {P} => {C} |  |

Đề 4:

Ảnh có chứa văn bản

Mô tả được tạo tự động

Câu 1:

a. Dữ liệu X:

= = 67.925

Sắp xếp dãy: 60.7 62.9 63.1 65 68 68 68 70.2 70.5 72.6 72.6 73.5

Dãy chẵn => Median = = 68

Mode X = 68

Dữ liệu Y:

= = 69.375

Sắp xếp dãy: 61.6 63.5 65 67.7 68.8 70.2 71.3 71.3 72.4 72.9 73.6 74.2

Dãy chẵn => Median = = 70.75

Mode Y = 71.3

b. Boxplot Y

Số phần tử quan sát: 12

Min = 61.6 Max = 74.2

Q1 = (trung vị nửa trái) = (65 + 67.7)/2 = 66.35

Q2 = Median = 70.75

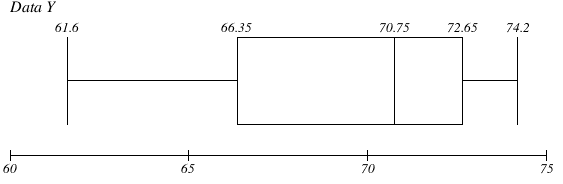
Q3 = (trung vị nửa phải) = (72.4 +72.9 ) = 72.65

IQR = Q3 – Q1 = 72.65 – 66.35 = 6.3

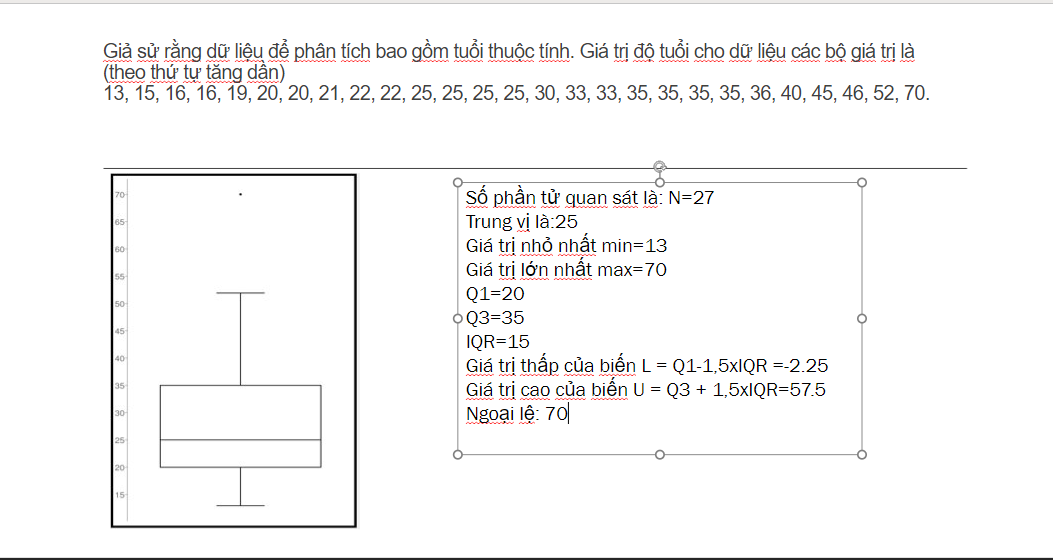
L = Q1 – 1.5 \* IQR = 66.35 – 1.5 \* 6.3 = 56.9

U = Q3 + 1.5 \* IQR = 72.65 + 1.5 \* 6.3 = 82.1

Ngoại lại: Không có (>U OR <L)









c. z-score Y

c. Chuẩn hoá Y về z-score

= = 3.96

v’ =

v’1 = = = 1.22

|  |  |  |
| --- | --- | --- |
| v’2 = 0.89 | v’6 = 0.21 | v’10 = -1.96 |
| v’3 = 1.07 | v’7 = -0.42 | v’11 = -0.15 |
| v’4 = 0.76 | v’8 = -1.1 | v’12 = 0.49 |
| v’5 = 0.49 | v’9 = -1.48 |  |

**d. làm trơn theo: trung bình, chia thùng: chiều rộng , số thùng: 3**

Sắp xếp dãy: 61.6 63.5 65 67.7 68.8 70.2 71.3 71.3 72.4 72.9 73.6 74.2

Khoảng dữ liệu: (74.2 – 61.6)/3 = 4.2

Bin 1: [61.6 ; 65.8) : 61.6 63.5 65

Bin 2: [65.8 ; 70) : 67.7 68.8

Bin 3: [70 ; 74.2] : 70.2 71.3 71.3 72.4 72.9 73.6 74.2

Làm trơn theo trung bình => Trung bình mỗi bin:

Bin 1: 63.37

Bin 2: 68.25

Bin 3: 72.27

Các bin sau khi làm trơn dữ liệu là:

Bin 1: 63.37 63.37 63.37

Bin 2: 68.25 68.25

Bin 3: 72.27 72.27 72.27 72.27 72.27 72.27 72.27

**d1. Làm trơn: theo trung vị, chia thùng: chiều rộng, số bin: 3**

Sắp xếp dãy: 61.6 63.5 65 67.7 68.8 70.2 71.3 71.3 72.4 72.9 73.6 74.2

Khoảng dữ liệu: (74.2 – 61.6)/3 = 4.2

Bin 1: [61.6 ; 65.8) : 61.6 63.5 65

Bin 2: [65.8 ; 70) : 67.7 68.8

Bin 3: [70 ; 74.2] : 70.2 71.3 71.3 72.4 72.9 73.6 74.2

Làm trơn theo trung vị => Trung vị mỗi bin:

Bin 1: 63.5

Bin 2: 68.25

Bin 3: 72.4

Các bin sau khi làm trơn dữ liệu là:

Bin 1: 63.5 63.5 63.5

Bin 2: 68.25 68.25

Bin 3: 72.4 72.4 72.4 72.4 72.4 72.4 72.4

**d1. Làm trơn: theo biên, chia thùng: chiều rộng, số bin: 3**

Sắp xếp dãy: 61.6 63.5 65 67.7 68.8 70.2 71.3 71.3 72.4 72.9 73.6 74.2

Khoảng dữ liệu: (74.2 – 61.6)/3 = 4.2

Bin 1: [61.6 ; 65.8) : 61.6 63.5 65

Bin 2: [65.8 ; 70) : 67.7 68.8

Bin 3: [70 ; 74.2] : 70.2 71.3 71.3 72.4 72.9 73.6 74.2

Các bin sau khi làm trơn dữ liệu theo biên là là: (giá trị cách 2 biên = nhau thì lấy tuỳ ý)

Bin 1: 61.6 65 65

Bin 2: 67.7 68.8

Bin 3: 70.2 70.2 70.2 74.2 74.2 74.2 74.2

**d1. Làm trơn: trung bình, chia thùng: chiều sâu, số bin: 3**

Sắp xếp dãy: 61.6 63.5 65 67.7 68.8 70.2 71.3 71.3 72.4 72.9 73.6 74.2

Chia thùng theo chiều sâu => Số dữ liệu mỗi bin là : 12/3 = 4

Bin 1: 61.6 63.5 65 67.7

Bin 2: 68.8 70.2 71.3 71.3

Bin 3: 72.4 72.9 73.6 74.2

Làm trơn theo trung bình => Trung bình mỗi bin

Bin 1: 64.45

Bin 2: 70.4

Bin 3: 73.275

Dữ liệu mỗi bin sau khi làm trơn là:

Bin 1: 64.45 64.45 64.45 64.45

Bin 2: 70.4 70.4 70.4 70.4

Bin 3: 73.275 73.275 73.275 73.275

**d1. Làm trơn: trung vị, chia thùng: chiều sâu, số bin: 3**

Sắp xếp dãy: 61.6 63.5 65 67.7 68.8 70.2 71.3 71.3 72.4 72.9 73.6 74.2

Chia thùng theo chiều sâu => Số dữ liệu mỗi bin là : 12/3 = 4

Bin 1: 61.6 63.5 65 67.7

Bin 2: 68.8 70.2 71.3 71.3

Bin 3: 72.4 72.9 73.6 74.2

Làm trơn theo trung vị => Trung vị mỗi bin

Bin 1: 64.25

Bin 2: 70.75

Bin 3: 73.25

Dữ liệu mỗi bin sau khi làm trơn là:

Bin 1: 64.25 64.25 64.25 64.25

Bin 2: 70.75 70.75 70.75 70.75

Bin 3: 73.25 73.25 73.25 73.25

**d1. Làm trơn: biên, chia thùng: chiều sâu, số bin: 3**

Sắp xếp dãy: 61.6 63.5 65 67.7 68.8 70.2 71.3 71.3 72.4 72.9 73.6 74.2

Chia thùng theo chiều sâu => Số dữ liệu mỗi bin là : 12/3 = 4

Bin 1: 61.6 63.5 65 67.7

Bin 2: 68.8 70.2 71.3 71.3

Bin 3: 72.4 72.9 73.6 74.2

Dữ liệu mỗi bin sau khi làm trơn là:

Bin 1: 61.6 61.6 67.7 67.7

Bin 2: 68.8 71.3 71.3 71.3

Bin 3: 72.4 72.4 74.2 74.2

e. Tương quan

= = 4.04

= = = 0.97

=> >0 => x và y là tương quan dương.

Bài 2: Tìm trung vị của tập dữ liệu tuổi

|  |  |
| --- | --- |
| **Tuổi** | **Số lượng** |
| 1 - 5 | 150 |
| 6 - 15 | 1500 |
| 16 - 22 | 300 |
| 23 - 50 | 450 |
| 51 - 80 | 650 |
| 81 - 100 | 60 |

Ảnh có chứa văn bản

Mô tả được tạo tự động

|  |  |  |
| --- | --- | --- |
| **Tuổi** | **Số lượng** | **Tần số tích luỹ** |
| 1 - 5 | 150 | 150 |
| 6 - 15 | 1500 | 1650 |
| 16 - 22 | 300 | 1950 |
| 23 - 50 | 450 | 2400 |
| 51 - 80 | 650 | 3050 |
| 81 - 100 | 60 | 3110 |

Tính:

N = 150 + 1500 + 300 + 450 + 650 + 60 = 3110

Khoảng trung vị: 3110/2 = 1555

(tần số tích luỹ(1650) so sánh >= 1555 ) => Khoảng trung vị hàng 2

= 150

= 1500

L1 = 6 (biên dưới khoảng trung vị)

width = 15 – 6 = 9 (hiệu 2 biên của khoảng trung vị)

median = L1 + = 6 + () x 9 = 14.43

**Câu 2.1: Euclide**

A1(1, 2), A2(2, 5), A3(8, 4), A4(5, 7), A5(7, 5), A6(6, 4), A7(4, 9), A8(2, 10)

Khởi tạo tâm :

Cụm 1: A1 => Tâm T1(1,2)

Cụm 2: A4 => Tâm T2(5,7)

Cụm 3: A7 => Tâm T3(4,9)

Euclid: d(i, j) =

Lặp 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(1,2) | T2(5,7) | T3(4,9) | Cụm |
| A1(1,2) | 0 | 6.4 | 7.61 | 1 |
| A2(2,5) | 3.16 | 3.6 | 4.47 | 1 |
| A3(8,4) | 7.28 | 4.24 | 6.4 | 2 |
| A4(5,7) | 6.4 | 0 | 2.23 | 2 |
| A5(7,5) | 6.71 | 2.83 | 5 | 2 |
| A6(6,4) | 5.38 | 3.16 | 5.38 | 2 |
| A7(4,9) | 7.61 | 2.23 | 0 | 3 |
| A8(2,10) | 8.06 | 4.24 | 2.23 | 3 |

Cụm 1: A1, A2 => Tâm T1 (1+2/2, 2+5/2) = (1.5, 3.5)

Cụm 2: A3, A4, A5, A6 => Tâm T2(8+5+7+6/4, 4+7+5+4/4) = (6.5, 5)

Cụm 3: A7, A8 => Tâm T3(4+2/2, 9+10/2) = (3, 9.5)

Lặp 2:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(1.5, 3.5) | T2(6.5, 5) | T3(3, 9.5) | Cụm |
| A1(1,2) | 1.58 | 6.26 | 7.76 | 1 |
| A2(2,5) | 1.58 | 4.5 | 4.61 | 1 |
| A3(8,4) | 6.52 | 1.8 | 7.43 | 2 |
| A4(5,7) | 4.95 | 2.5 | 3.2 | 2 |
| A5(7,5) | 5.7 | 0.5 | 6.02 | 2 |
| A6(6,4) | 4.53 | 1.11 | 6.26 | 2 |
| A7(4,9) | 6.04 | 4.72 | 1.11 | 3 |
| A8(2,10) | 6.52 | 6..73 | 1.11 | 3 |

Ta thấy tâm các cụm không thay đổi => Thuật toán dừng

Cụm 1: A1, A2

Cụm 2: A3, A4, A5, A6

Cụm 3: A7, A8

**Câu 2.2: Mahattan**

A1(1, 2), A2(2, 5), A3(8, 4), A4(5, 7), A5(7, 5), A6(6, 4), A7(4, 9), A8(2, 10)

Khởi tạo tâm :

Cụm 1: A1 => Tâm T1(1,2)

Cụm 2: A4 => Tâm T2(5,7)

Cụm 3: A7 => Tâm T3(4,9)

Mahattan: d(i, j) = |x1 – x2| + |y1 -y2|

Lặp 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(1,2) | T2(5,7) | T3(4,9) | Cụm |
| A1(1,2) | 0 | 9 | 10 | 1 |
| A2(2,5) | 4 | 5 | 6 | 1 |
| A3(8,4) | 9 | 6 | 9 | 2 |
| A4(5,7) | 9 | 0 | 3 | 2 |
| A5(7,5) | 9 | 4 | 7 | 2 |
| A6(6,4) | 7 | 4 | 7 | 2 |
| A7(4,9) | 10 | 3 | 0 | 3 |
| A8(2,10) | 9 | 6 | 3 | 3 |

Cụm 1: A1, A2 => Tâm T1 (1+2/2, 2+5/2) = (1.5, 3.5)

Cụm 2: A3, A4, A5, A6 => Tâm T2(8+5+7+6/4, 4+7+5+4/4) = (6.5, 5)

Cụm 3: A7, A8 => Tâm T3(4+2/2, 9+10/2) = (3, 9.5)

Lặp 2:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(1.5, 3.5) | T2(6.5, 5) | T3(3, 9.5) | Cụm |
| A1(1,2) | 2 | 8.5 | 9.5 | 1 |
| A2(2,5) | 2 | 4.5 | 5.5 | 1 |
| A3(8,4) | 7 | 2.5 | 10.5 | 2 |
| A4(5,7) | 7 | 3.5 | 4.5 | 2 |
| A5(7,5) | 7 | 0.5 | 8.5 | 2 |
| A6(6,4) | 5 | 1.5 | 8.5 | 2 |
| A7(4,9) | 8 | 6.5 | 1.5 | 3 |
| A8(2,10) | 7 | 9.5 | 1.5 | 3 |

Ta thấy tâm các cụm không thay đổi => Thuật toán dừng

Cụm 1: A1, A2

Cụm 2: A3, A4, A5, A6

Cụm 3: A7, A8

**Câu 2.3: Minkowski**

A1(1, 2), A2(2, 5), A3(8, 4), A4(5, 7), A5(7, 5), A6(6, 4), A7(4, 9), A8(2, 10)

Khởi tạo tâm :

Cụm 1: A1 => Tâm T1(1,2)

Cụm 2: A4 => Tâm T2(5,7)

Cụm 3: A7 => Tâm T3(4,9)

Minkowski:

Lặp 1:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(1,2) | T2(5,7) | T3(4,9) | Cụm |
| A1(1,2) | 0 | 5.74 | 7.18 | 1 |
| A2(2,5) | 3.04 | 3.27 | 4.16 | 1 |
| A3(8,4) | 7.05 | 3.78 | 5.74 | 2 |
| A4(5,7) | 5.73 | 0 | 2.08 | 2 |
| A5(7,5) | 6.24 | 2.52 | 4.49 | 2 |
| A6(6,4) | 5.1 | 3.04 | 5.1 | 2 |
| A7(4,9) | 7.18 | 2.08 | 0 | 3 |
| A8(2,10) | 8 | 3.78 | 2.08 | 3 |

Cụm 1: A1, A2 => Tâm T1 (1+2/2, 2+5/2) = (1.5, 3.5)

Cụm 2: A3, A4, A5, A6 => Tâm T2(8+5+7+6/4, 4+7+5+4/4) = (6.5, 5)

Cụm 3: A7, A8 => Tâm T3(4+2/2, 9+10/2) = (3, 9.5)

Lặp 2:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(1.5, 3.5) | T2(6.5, 5) | T3(3, 9.5) | Cụm |
| A1(1,2) | 1.52 | 5.78 | 7.55 | 1 |
| A2(2,5) | 1.52 | 4.5 | 4.51 | 1 |
| A3(8,4) | 6.5 | 1.63 | 6.63 | 2 |
| A4(5,7) | 4.41 | 2.25 | 2.87 | 2 |
| A5(7,5) | 5.54 | 0.5 | 5.37 | 2 |
| A6(6,4) | 4.5 | 1.04 | 5.78 | 2 |
| A7(4,9) | 5.67 | 4.3 | 1.04 | 3 |
| A8(2,10) | 6.5 | 6.0 | 1.04 | 3 |

Ta thấy tâm các cụm không thay đổi => Thuật toán dừng

Cụm 1: A1, A2

Cụm 2: A3, A4, A5, A6

Cụm 3: A7, A8

**Câu 2.4: Tính khoảng cách tối đa giữa 2 giá trị**

A1(1, 2), A2(2, 5), A3(8, 4), A4(5, 7), A5(7, 5), A6(6, 4), A7(4, 9), A8(2, 10)

Khởi tạo tâm :

Cụm 1: A1 => Tâm T1(1,2)

Cụm 2: A4 => Tâm T2(5,7)

Cụm 3: A7 => Tâm T3(4,9)

Khoảng cách tối đa:   
d(A1, T2) = max [|1-5|, |2-7|] = max [4, 5] = 5

Lặp 1:

Nếu 2 khoảng cách = nhau, lấy cái cụm nhỏ hơn.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(1,2) | T2(5,7) | T3(4,9) | Cụm |
| A1(1,2) | 0 | 5 | 7 | 1 |
| A2(2,5) | 3 | 3 | 4 | 1 |
| A3(8,4) | 7 | 3 | 5 | 2 |
| A4(5,7) | 5 | 0 | 2 | 2 |
| A5(7,5) | 6 | 2 | 4 | 2 |
| A6(6,4) | 5 | 3 | 5 | 2 |
| A7(4,9) | 7 | 2 | 0 | 3 |
| A8(2,10) | 8 | 3 | 2 | 3 |

Cụm 1: A1, A2 => Tâm T1 (1+2/2, 2+5/2) = (1.5, 3.5)

Cụm 2: A3, A4, A5, A6 => Tâm T2(8+5+7+6/4, 4+7+5+4/4) = (6.5, 5)

Cụm 3: A7, A8 => Tâm T3(4+2/2, 9+10/2) = (3, 9.5)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Điểm | T1(1.5, 3.5) | T2(6.5, 5) | T3(3, 9.5) | Cụm |
| A1(1,2) | 1.5 | 5.5 | 7.5 | 1 |
| A2(2,5) | 1.5 | 4.5 | 4.5 | 1 |
| A3(8,4) | 6.5 | 1.5 | 5.5 | 2 |
| A4(5,7) | 3.5 | 2 | 2.5 | 2 |
| A5(7,5) | 5.5 | 0 | 4.5 | 2 |
| A6(6,4) | 4.5 | 1 | 5.5 | 2 |
| A7(4,9) | 5.5 | 4 | 1 | 3 |
| A8(2,10) | 6.5 | 5 | 1 | 3 |

Ta thấy tâm các cụm không thay đổi => Thuật toán dừng

Cụm 1: A1, A2

Cụm 2: A3, A4, A5, A6

Cụm 3: A7, A8