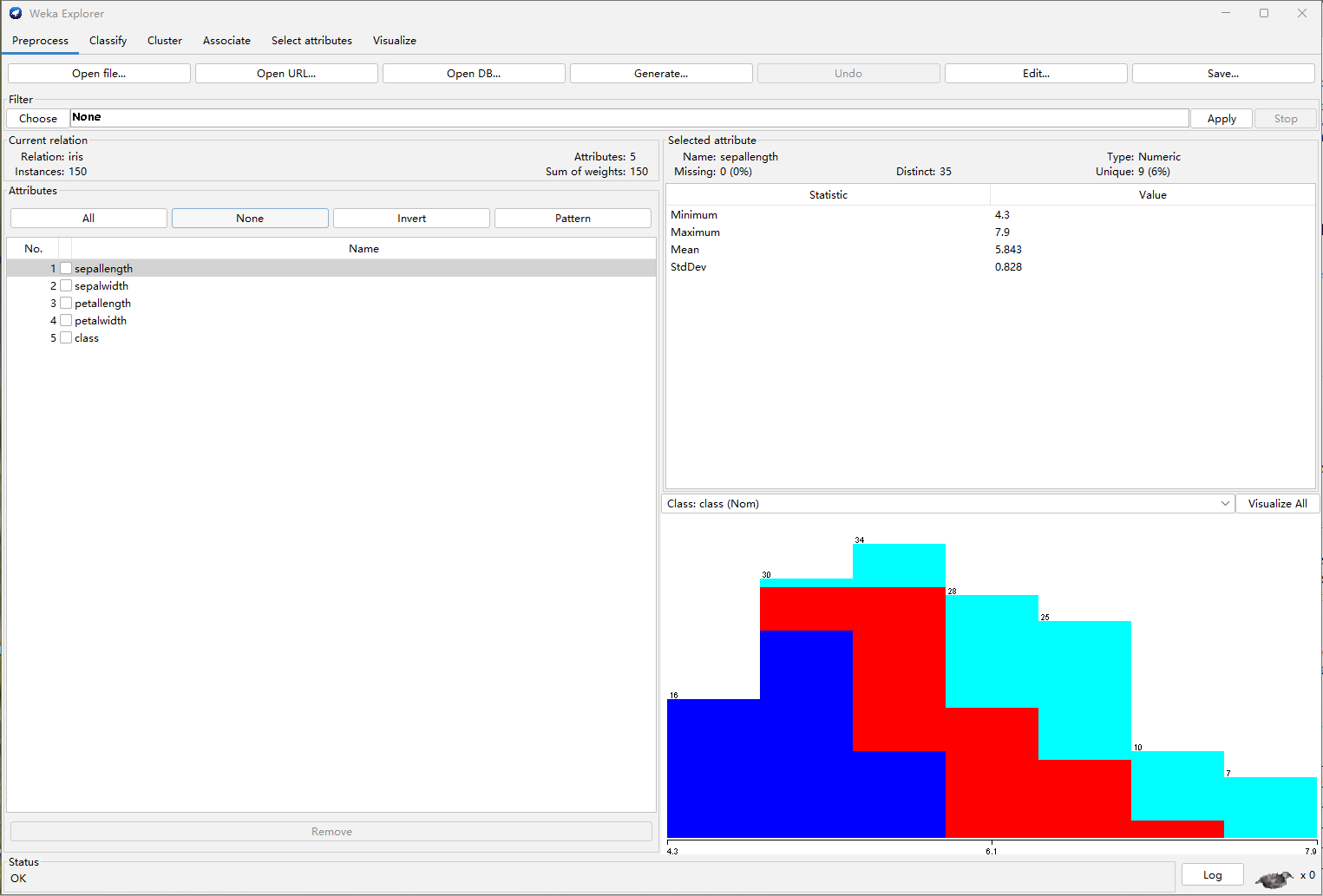
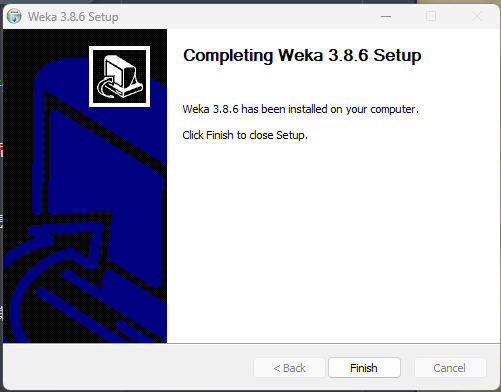
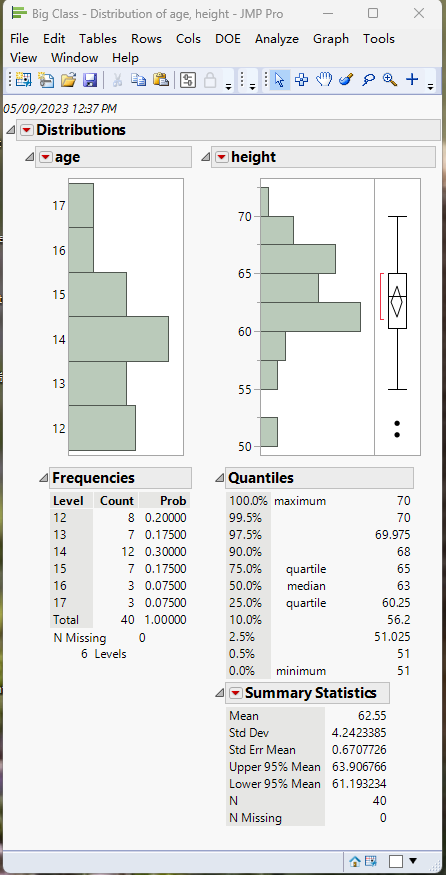
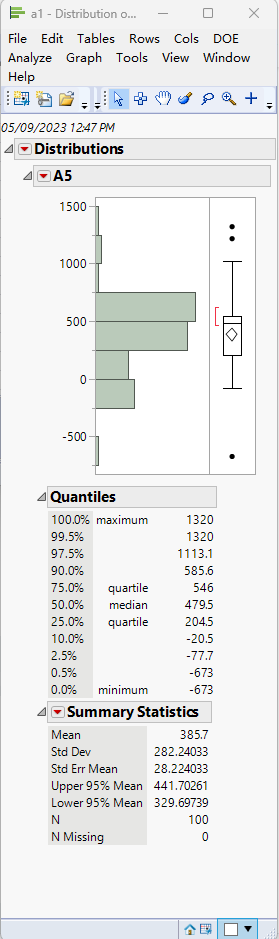
**Problem 1:Install Weka**

****

**Problem 2:Install JMP Pro**

****

**Problem 3:**

****

1. **.Calculate the mean, median, and standard deviation (sample) of the attribute A5.**

Mean = 385.7 median = 479.5 standard deviation = 282.24033

1. **. Determine Q1, Q2, and Q3 of A5.**

Q1 = 204.5 Q2 = 479.5 Q3 = 546

**(3). Detect outliers using the IQR method, which we discussed in the class, and show the**

**A5 values of the detected outliers. When detecting outliers, use only the A5 values.**

IQR = Q3 - Q1 = 546 - 204.5 = 341.5

Q1-1.5IQR = 204.5 - 1.5\*341.5 = -307.5

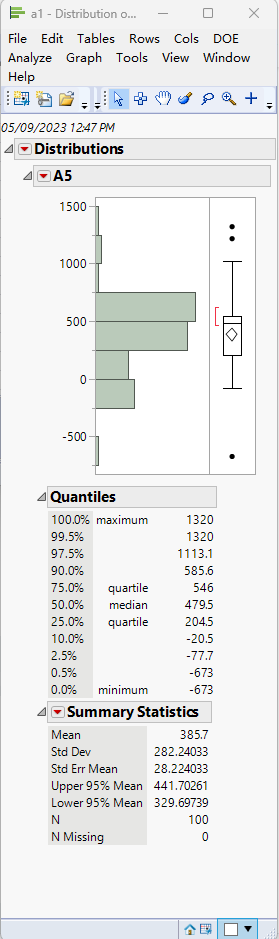
Q3+1.5IQR = 546 + 1.5\*341.5 = 1058.25

The outliers will exist greater than Q3+1.5IQR or smaller than Q1-1.5IQR

So the values of the outliers are 1216 1320 and -673

**(4). Plot the boxplot of the attribute A5. In your boxplot, you need to show outliers**

**separately.**

****

**Problem 4:**

**Calculate the distance between P4 and P5, *d*(P4, P5), and the distance between P4 and P6, *d*(P4, P6). Is P4 closer to P5 or P6? Here, all attributes are nominal attributes.**

P4 VS P5:

Job: management vs blue-collar (mismatch)

Marital: married vs single (mismatch)

Education: tertiary vs secondary (mismatch)

Default: no vs no (match)

Housing: yes vs yes (match)

Loan: yes vs no (mismatch)

Contact: unknown vs unknown (match)

*d*(P4, P5) = (4 mismatch) / (7 attribute) = 0.57

P4 VS P6:

Job: management vs management (match)

Marital: married vs single (mismatch)

Education: tertiary vs tertiary (match)

Default: no vs no (match)

Housing: yes vs no (mismatch)

Loan: yes vs yes (match)

Contact: unknown vs cellular (mismatch)

*d*(P4, P6) = (3 mismatch) / (7 attribute) = 0.43

P4 is closer to P6, because of *d*(P4, P6) < *d*(P4, P5)

**Problem 5:**

1. **. Calculate the distance between O1 and O2 using the Manhattan distance.**

m

1. **. Calculate the distance between O1 and O2 using the Euclidean distance.**

e

**Problem 6:**

**Calculate the distance between O1 and O2 using the method that we discussed in the class. Use the Euclidean distance measure.**

Change A2 A3 A4 to the numerical attribute according to the given,

Second -> 2 First ->1

Silver->2 gold->3

Small->1 large->3

so the table will change to:

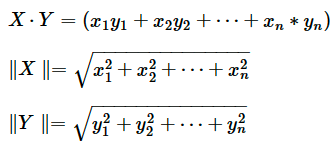
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A1 | A2 | A3 | A4 |
| O1 | 2 | 2 | 2 | 1 |
| O2 | 4 | 1 | 3 | 3 |
| O1-O2 | -2 | 1 | -1 | -2 |
| (O1-O2)2 | 4 | 1 | 1 | 4 |

e

**Problem 7:**

**Calculate the similarity between D1 and D2, cos(D1, D2), and the similarity between D1 and D3, cos(D1, D3), using the cosine similarity measure. Is D1 closer to D2 or D3? You must calculate the cosine similarity yourself (i.e., you must not use a built-in function of a software).**

**cos1**

****

D1 and D2

D1 and D3

is closer to 1, so D1 is closer D3