

資料探勘 2011.03.29 練習

1. 一組資料 {20, 2, 2, 13, 3, 2, 9, 2, 41, 9, 1}下求列資料

1. 全距

$$\text{Max}\{20, 2, 2, 13, 3, 2, 9, 2, 41, 9, 1\} - \text{min}\{20, 2, 2, 13, 3, 2, 9, 2, 41, 9, 1\} = 41 - 1 = 40$$

2. 平均數

$$(20+2+2+13+3+2+9+2+41+9+1) / 11 = 9.45$$

3. 標準差

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

s

=

$\sqrt{\frac{1}{10} [(20 - 9.4)^2 + (2 - 9.4)^2 + (2 - 9.4)^2 + (13 - 9.4)^2 + (3 - 9.4)^2 + (2 - 9.4)^2 + (9 - 9.4)^2 + (2 - 9.4)^2 + (41 - 9.4)^2 + (9 - 9.4)^2 + (1 - 9.4)^2]}$

=

$\sqrt{\frac{1}{10} [111.21 + 55.57 + 55.57 + 12.57 + 41.66 + 55.57 + 0.21 + 55.57 + 995.12 + 0.21 + 71.48]} = 12.06$

4. 中位數

將資料排序過{1, 2, 2, 2, 2, 3, 9, 9, 13, 20, 41}

中位數為 3

5. 眾數

1 出現 1 次

2 出現 4 次

3 出現 1 次

9 出現 2 次

13 出現 1 次

20 出現 1 次

41 出現 1 次

眾數為 2

6. 絕對平均離差 (AAD)

$$\text{AAD}(x) = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

$$\text{AAD}(x) = (1/11) \times$$

$$(|1-9.45| + |2-9.45| + |2-9.45| + |2-9.45| + |2-9.45| + |3-9.45| + |9-9.45| + |9-9.45| + |13-9.45| + |20-9.45| + |41-9.45|) = 8.30$$

7. 中位數絕對離差 (MAD)

$$MAD(x) = \text{中位數} \left(\{|x_1 - \bar{x}|, \dots, |x_m - \bar{x}|\} \right)$$

$\{|1-9.45|, |2-9.45|, |2-9.45|, |2-9.45|, |2-9.45|, |3-9.45|, |9-9.45|, |9-9.45|, |13-9.45|, |20-9.45|, |41-9.45|\}$

$=\{10.55, 7.45, 7.45, 3.55, 6.45, 7.45, 0.45, 7.45, 31.55, 0.45, 8.45\}$

再經過排序

$\{0.45, 0.45, 3.55, 6.45, 7.45, \textcolor{red}{7.45}, 7.45, 7.45, 8.45, 10.55, 31.55\}$

$MAD = 7.45$

8. 四分位差 (IQR)

$$\text{四分位差}(x) = x_{75\%} - x_{25\%}$$

$x_{100\%} = \max\{20, 2, 2, 13, 3, 2, 9, 2, 41, 9, 1\} = 41$

$x_{0\%} = \min\{20, 2, 2, 13, 3, 2, 9, 2, 41, 9, 1\} = 1$

$x_{50\%} = (x_{100\%} + x_{0\%}) / 2 = (41 + 1) / 2 = 21$

$x_{75\%} = (x_{100\%} + x_{50\%}) / 2 = (41 + 21) / 2 = 31$

$x_{25\%} = (x_{50\%} + x_{0\%}) / 2 = (21 + 1) / 2 = 11$

$IQR = x_{75\%} - x_{25\%} = 31 - 11 = 20$