

MATH 11: Discussion Week1

Apr. 2019

Displaying and Summarizing Quantitative Data

1. Mean and variance

Goal: Consider there are some midterm scores of students,

59, 67, 78, 75, 64, 89, 72

(1) Compute the **mean** of this data set (you can use calculator).

(Hint: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.)

(2) Compute the **variance** of this data set (you can use calculator). Could you also show the **standard deviation**? (Hint: $Var(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$, $SD(x) = \sqrt{Var(x)}$.)

2. Median and inter-quartile range(IQR)

Consider the data set given above

(1) Compute the **median** of your data set.

(2) Compute the **lower quartile** Q_1 and **upper quartile** Q_3 of your data set. (Hint: Q_1 -median of the lower half, Q_3 -median of the upper half.)

(3) What is the **IQR** of this data set? (Hint: $IQR = Q_3 - Q_1$)

Consider a new data set given below

59, 67, 78, 75, 64, 89, 72, 12

(1) Recompute median, Q_1 and Q_3 again.

(2) Draw the boxplot (notice the outlier). (you need to recompute the IQR and the upper&lower fence.)

(3) Consider this new data set, given that the new mean and variance are 64.5 and 534.5714. Compared with median (69.5) and IQR (15), which pair of statistics would you like to use to describe the data set?

3. Correlation

(1) Determine that the following statements about the correlation is **true** or **false**:

$$\text{Hint: } r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

- ___ The range of correlation r is $-1 \leq r \leq 1$.
- ___ The + sign of correlation r indicates the positive associations.
- ___ The stronger association occurs when r is closer to 1 or -1 .
- ___ The choice of predictor/response doesn't matter, i.e. $\text{cor}(X, Y) = \text{cor}(Y, X)$.
- ___ The correlation is unaffected by linear scale changes, i.e. $\text{cor}(X, Y) = \text{cor}(X, 2Y) = \text{cor}(X, Y + 5)$.

(2) Could you provide two examples to illustrate that correlation is not causation?