Sample Mean and Variance

Saturday 1 June 2024 8:51 AM

$D_1 \quad \{x_1, x_2, \ldots x_{n_1}\}$

n_1 : # points

$\bar{x}_1$: sample mean

s_1^2 : sample variance (biased)

$D_2 \quad \{x_{n_1+1}, \ldots, x_{n_1+n_2}\}$

n_2 : # points

$\bar{x}_2$: sample mean

s_2^2 : sample variance (biased)

D = combine D_1 & D_2

$\bar{x}$?

s^2 ?

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i$$

$$\bar{x}_2 = \frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2} x_i$$

$$\bar{x} = \frac{1}{n_1+n_2}\sum_{i=1}^{n_1+n_2} x_i = \frac{1}{n_1+n_2}\left(\overbrace{n_1\bar{x}_1 + n_2\bar{x}_2}\right)$$

$$\sum_{i=1}^{n_1} x_i + \sum_{i=n_1+1}^{n_1+n_2} x_i$$

$$s_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2$$

$$s_2^2 = \frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2}(x_i - \bar{x}_2)^2$$

$$s^2 = \frac{1}{n_1+n_2}\sum_{i=1}^{n_1+n_2}(x_i - \bar{x})^2$$

$$\sum_{i=1}^{n_1+n_2}(x_i - \bar{x})^2 = \sum_{i=1}^{n_1+n_2} x_i^2 + \bar{x}^2 - 2x_i\bar{x}$$

$$n_1 s_1^2 = \sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2 = \sum_{i=1}^{n_1} x_i^2 + \bar{x}_1^2 - 2\bar{x}_1 x_i$$

$$= \sum_{i=1}^{n_1} x_i^2 + n_1\bar{x}_1^2 - 2\bar{x}_1\sum_{i=1}^{n_1} x_i$$

$$= \sum_{i=1}^{n_1} x_i^2 + n_1\bar{x}_1^2 - 2\bar{x}_1(n_1\bar{x}_1)$$

$$= \sum_{i=1}^{n_1} x_i^2 + n_1\bar{x}_1^2 - 2n_1\bar{x}_1^2$$

$$= \sum_{i=1}^{n_1} x_i^2 - n_1\bar{x}_1^2$$

$$\boxed{\begin{aligned} n_1 s_1^2 + n_1\bar{x}_1^2 &= \sum_{i=1}^{n_1} x_i^2 \\ n_2 s_2^2 + n_2\bar{x}_2^2 &= \sum_{i=n_1+1}^{n_1+n_2} x_i^2 \end{aligned}}$$

$$(n_1+n_2)s^2 = \sum_{i=1}^{n_1+n_2} x_i^2 + (n_1+n_2)\bar{x}^2 - 2\bar{x}\sum_{i=1}^{n_1+n_2} x_i$$

$$= \left(n_1 s_1^2 + n_1\bar{x}_1^2\right) + \left(n_2 s_2^2 + n_2\bar{x}_2^2\right) + (n_1+n_2)\bar{x}^2 - 2\bar{x}^2(n_1+n_2)$$

$$(n_1+n_2)s^2 = n_1 s_1^2 + n_2 s_2^2 + n_1\bar{x}_1^2 + n_2\bar{x}_2^2 - (n_1+n_2)\bar{x}^2$$

$$s^2 = \frac{1}{\ldots}\left(n_1 s_1^2 + n_2 s_2^2 + n_1\bar{x}_1^2 + n_2\bar{x}_2^2 - (n_1+n_2)\bar{x}^2\right)$$

$$s^2 = \frac{1}{n_1+n_2}\left(n_1 s_1^2 + n_2 s_2^2 + n_1 x_1^2 + n_2 \bar{x}_2^2 - (n_1+n_2)\bar{x}^2\right)$$

$$n_2 s_2 + n_2 x_2 = \sum_{i=n_1+1} x_i$$

→ Extend to k datasets