



PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

I – PHƯƠNG SAI

Ví dụ 1. Cho biết giá trị thành phẩm quy ra tiền (nghìn đồng) trong một tuần lao động của 7 công nhân ở tổ 1 là

180, 190, 190, 200, 210, 210, 220, (1)

còn của 7 công nhân ở tổ 2 là

150, 170, 170, 200, 230, 230, 250. (2)

Ta thấy số trung bình cộng \bar{x} của dãy (1) và số trung bình cộng \bar{y} của dãy (2) bằng nhau

$$\bar{x} = \bar{y} = 200.$$

Tuy nhiên, khi so sánh dãy (1) và dãy (2) ta thấy các số liệu ở dãy (1) gần với số trung bình cộng hơn, nên chúng đồng đều hơn. Khi đó ta nói các số liệu thống kê ở dãy (1) ít phân tán hơn dãy (2).

Để tìm số đo độ phân tán (so với số trung bình cộng) của dãy (1) ta tính

Các độ lệch của mỗi số liệu thống kê đối với số trung bình cộng

(180 - 200) ; (190 - 200) ; (190 - 200) ; (200 - 200) ; (210 - 200) ;
(210 - 200) ; (220 - 200).

Bình phương các độ lệch và tính trung bình cộng của chúng, ta được

$$s_1^2 = \frac{(180 - 200)^2 + 2(190 - 200)^2 + (200 - 200)^2 + 2(210 - 200)^2 + (220 - 200)^2}{7}$$

$$\approx 171,4.$$

Số s_1^2 được gọi là **phương sai** của dãy (1).

Tương tự phương sai s_2^2 của dãy (2) là

$$s_2^2 = \frac{(150 - 200)^2 + 2(170 - 200)^2 + (200 - 200)^2 + 2(230 - 200)^2 + (250 - 200)^2}{7}$$

$$\approx 1228,6.$$

Ta thấy phương sai của dãy (1) nhỏ hơn phương sai của dãy (2). Điều đó biểu thị độ phân tán của các số liệu thống kê ở dãy (1) ít hơn ở dãy (2).

Ví dụ 2. Tính phương sai s^2 của các số liệu thống kê cho ở bảng 4, §1 (cũng gọi là phương sai của bảng 4).

Số trung bình cộng của bảng 4 là $\bar{x} = 162$ cm.

Mỗi số liệu thống kê thuộc một lớp được thay thế bởi giá trị đại diện của lớp đó.

a) Phương sai s^2 của bảng 4 (bảng phân bố tần số và tần suất ghép lớp) được tính như sau

$$s^2 = \frac{6(153 - 162)^2 + 12(159 - 162)^2 + 13(165 - 162)^2 + 5(171 - 162)^2}{36}$$

$$\approx 31.$$

(3)

Hệ thức (3) biểu thị cách tính gần đúng phương sai của bảng 4 theo tần số.

b) Từ (3) ta có

$$s^2 = \frac{6}{36}(153 - 162)^2 + \frac{12}{36}(159 - 162)^2 + \frac{13}{36}(165 - 162)^2 + \frac{5}{36}(171 - 162)^2$$

hay

$$s^2 \approx \frac{16,7}{100}(153 - 162)^2 + \frac{33,3}{100}(159 - 162)^2 + \frac{36,1}{100}(165 - 162)^2 + \frac{13,9}{100}(171 - 162)^2$$
$$\approx 31. \quad (4)$$

Hệ thức (4) biểu thị cách tính gần đúng phương sai của bảng 4 theo tần suất.

CHÚ Ý

a) Khi hai dãy số liệu thống kê có cùng đơn vị đo và có số trung bình cộng bằng nhau hoặc xấp xỉ nhau, nếu phương sai càng nhỏ thì mức độ phân tán (so với số trung bình cộng) của các số liệu thống kê càng bé.

b) Có thể tính phương sai theo các công thức sau đây

Trường hợp bảng phân bố tần số, tần suất

$$s^2 = \frac{1}{n} \left[n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2 \right]$$
$$= f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_k(x_k - \bar{x})^2.$$

trong đó n_i , f_i lần lượt là tần số, tần suất của giá trị x_i ; n là số các số liệu thống kê ($n = n_1 + n_2 + \dots + n_k$); \bar{x} là số trung bình cộng của các số liệu đã cho.

Trường hợp bảng phân bố tần số, tần suất ghép lớp

$$s^2 = \frac{1}{n} \left[n_1(c_1 - \bar{x})^2 + n_2(c_2 - \bar{x})^2 + \dots + n_k(c_k - \bar{x})^2 \right]$$
$$= f_1(c_1 - \bar{x})^2 + f_2(c_2 - \bar{x})^2 + \dots + f_k(c_k - \bar{x})^2.$$

trong đó c_i , n_i , f_i lần lượt là giá trị đại diện, tần số, tần suất của lớp thứ i ; n là số các số liệu thống kê ($n = n_1 + n_2 + \dots + n_k$); \bar{x} là số trung bình cộng của các số liệu thống kê đã cho.

Ngoài ra, người ta còn chứng minh được công thức sau

$$s^2 = \overline{x^2} - (\bar{x})^2$$

trong đó $\overline{x^2}$ là trung bình cộng của các bình phương số liệu thống kê, tức là

$$\overline{x^2} = \frac{1}{n}(n_1x_1^2 + n_2x_2^2 + \dots + n_kx_k^2) = f_1x_1^2 + f_2x_2^2 + \dots + f_kx_k^2$$

(đối với bảng phân bố tần số, tần suất),

$$\overline{x^2} = \frac{1}{n}(n_1c_1^2 + n_2c_2^2 + \dots + n_kc_k^2) = f_1c_1^2 + f_2c_2^2 + \dots + f_kc_k^2$$

(đối với bảng phân bố tần số, tần suất ghép lớp).



Hãy tính phương sai của bảng 6 (ở §2).

II – ĐỘ LỆCH CHUẨN

Trong ví dụ 2 ở trên, ta đã tính được phương sai của bảng 4 (ở §1) bằng $s^2 \approx 31$. Nếu để ý đến đơn vị đo thì ta thấy đơn vị đo của s^2 là cm^2 (bình phương đơn vị đo của dấu hiệu được nghiên cứu). Muốn tránh điều này, có thể dùng căn bậc hai của phương sai gọi là **độ lệch chuẩn** (của bảng 4) và kí hiệu là s . Vậy

$$s = \sqrt{s^2} \approx \sqrt{31} \approx 5,6 \text{ (cm)}.$$

Phương sai s^2 và độ lệch chuẩn s đều được dùng để đánh giá mức độ phân tán của các số liệu thống kê (so với số trung bình cộng). Nhưng khi cần chú ý đến đơn vị đo thì ta dùng s , vì s có cùng đơn vị đo với dấu hiệu được nghiên cứu.



Hãy tính độ lệch chuẩn của bảng 6 (ở §2).

BÀI ĐỌC THÊM



SỬ DỤNG MÁY TÍNH BỎ TÚI CASIO $fx - 500MS$ ĐỂ TÌM SỐ TRUNG BÌNH CỘNG VÀ ĐỘ LỆCH CHUẨN

Ví dụ. Cho bảng phân bố tần số

Khối lượng của 30 con thằn lằn

Khối lượng (gam)	140	150	160	170	180	190	Cộng
Tần số	2	3	5	9	8	3	30

Sử dụng máy tính bỏ túi CASIO $fx-500MS$, ta tìm số trung bình cộng \bar{x} và độ lệch chuẩn s của bảng phân bố đã cho như sau

1. Chọn MODE cho phép tính thống kê :

ấn **MODE** **2**

2. Xoá những bài thống kê cũ

ấn lần lượt **SHIFT** **CLR** **1** **=**.

3. Nhập dữ liệu

ấn liên tiếp 140 **SHIFT** **;** **2** **DT**

150 **SHIFT** **;** **3** **DT**

Tương tự đối với các cột 160, 170, 180, 190.

4. Gọi kết quả

a) Để tìm \bar{x} , ấn **SHIFT** **S-VAR** **1** **=**

Kết quả là $\bar{x} = 169$ (gam).

b) Để tìm s , ấn **SHIFT** **S-VAR** **2** **=**

Kết quả cho giá trị $s \approx 13,5$; đây chính là giá trị s cần tìm.

5. Chú ý

a) Không cần nhập đúng thứ tự của số liệu.

Để gọi dữ liệu (đã nhập), ấn **▲** hoặc **▼**.

Có thể hiệu chỉnh số liệu hoặc tần số như sau

Gọi số liệu (hay tần số) đó, rồi nhập giá trị mới và ấn **=**, giá trị mới sẽ thay thế giá trị cũ.

Có thể xoá một dữ liệu bằng cách gọi nó lên, rồi ấn **SHIFT** **CL** (các dữ liệu còn lại sẽ tự động dồn số thứ tự lại).

b) Đối với bảng phân bố tần số ghép lớp, ta sử dụng các giá trị đại diện của các lớp và làm tương tự.



Bài tập

1. Tính phương sai và độ lệch chuẩn của bảng phân bố tần số đã được lập ở bài tập 1 và của bảng phân bố tần số ghép lớp cho ở bài tập 2 của §1.
2. Hai lớp 10C, 10D của một trường Trung học phổ thông đồng thời làm bài thi môn Ngữ văn theo cùng một đề thi. Kết quả thi được trình bày ở hai bảng phân bố tần số sau đây

Điểm thi Ngữ văn của lớp 10C

Điểm thi	5	6	7	8	9	10	Cộng
Tần số	3	7	12	14	3	1	40

Điểm thi Ngữ văn của lớp 10D

Điểm thi	6	7	8	9	Cộng
Tần số	8	18	10	4	40

- a) Tính các số trung bình cộng, phương sai, độ lệch chuẩn của các bảng phân bố tần số đã cho.
 - b) Xét xem kết quả làm bài thi của môn Ngữ văn ở lớp nào là đồng đều hơn ?
3. Cho hai bảng phân bố tần số ghép lớp

Khối lượng của nhóm cá mè thứ 1

Lớp khối lượng (kg)	[0,6 ; 0,8)	[0,8 ; 1,0)	[1,0 ; 1,2)	[1,2 ; 1,4]	Cộng
Tần số	4	6	6	4	20

Khối lượng của nhóm cá mè thứ 2

Lớp khối lượng (kg)	[0,5 ; 0,7)	[0,7 ; 0,9)	[0,9 ; 1,1)	[1,1 ; 1,3)	[1,3 ; 1,5]	Cộng
Tần số	3	4	6	4	3	20

- a) Tính các số trung bình cộng của các bảng phân bố tần số ghép lớp đã cho.
- b) Tính phương sai của các bảng phân bố tần số ghép lớp đã cho.
- c) Xét xem nhóm cá nào có khối lượng đồng đều hơn ?