

Bài tập cuối chương VI

A. Lý thuyết

1. Số gần đúng

Trong đo đạc và tính toán, ta thường chỉ nhận được các số gần đúng.

Ví dụ: Dân số Việt Nam năm 2017 ước tính là 93,7 triệu người. Khi đó con số 93,7 triệu người là số gần đúng.

2. Sai số của số gần đúng

a. Sai số tuyệt đối

Nếu a là số gần đúng của số đúng \bar{a} thì $\Delta_a = |\bar{a} - a|$ được gọi là sai số tuyệt đối của số gần đúng a (Hình vẽ).



Chú ý: Sai số tuyệt đối của số gần đúng nhận được trong một phép đo đạc, tính toán càng bé thì kết quả của phép đo đạc, tính toán đó càng chính xác.

Ví dụ: Hai bạn Nam và Long muốn tính chu vi của một đường tròn có bán kính 1 cm. Bạn Nam lấy π là 3,14 còn Long lấy π là 3,1. Hỏi kết quả của bạn nào chính xác hơn.

Hướng dẫn giải

Gọi chu vi đường tròn bán kính $r = 1$ cm là $C = 2\pi r$ (cm).

Bạn Nam tính được chu vi của đường tròn khi lấy $\pi = 3,14$ là:

$$C_1 = 2\pi r = 2 \cdot 3,14 \cdot 1 = 6,28 \text{ (cm)}.$$

Bạn Long tính được chu vi của đường tròn khi lấy $\pi = 3,1$ là:

$$C_2 = 2\pi r = 2 \cdot 3,1 \cdot 1 = 6,2 \text{ (cm)}.$$

Ta thấy $3,1 < 3,14 < \pi$ nên $2 \cdot 3,1 \cdot 1 < 2 \cdot 3,14 \cdot 1 < 2 \cdot \pi \cdot 1$

Tức là $C_2 < C_1 < C$.

Suy ra $\Delta_{C_1} = |C - C_1| < |C - C_2| = \Delta_{C_2}$.

$\Rightarrow \Delta_{C_1} < \Delta_{C_2}$.

\Rightarrow Kết quả của bạn Nam chính xác hơn kết quả của bạn Long.

Vậy kết quả tính chu vi đường tròn của bạn Nam chính xác hơn kết quả của bạn Long.

b. Độ chính xác của một số gần đúng

Nhận xét:

- Giả sử a là số gần đúng của số đúng \bar{a} sao cho $\Delta_a = |\bar{a} - a| \leq d$.

Khi đó $\Delta_a = |\bar{a} - a| \leq d \Leftrightarrow -d \leq \bar{a} - a \leq d \Leftrightarrow a - d \leq \bar{a} \leq a + d$.

- Ta nói a là số gần đúng của số đúng \bar{a} với độ chính xác d nếu $\Delta_a = |\bar{a} - a| \leq d$ và quy ước viết gọn là $\bar{a} = a \pm d$.

- Nếu $\Delta_a \leq d$ thì số đúng \bar{a} nằm trong đoạn $[a - d; a + d]$. Bởi vậy, d càng nhỏ thì độ sai lệch của số gần đúng a so với số đúng \bar{a} càng ít. Điều đó giải thích vì sao d được gọi là độ chính xác của số gần đúng.

Ví dụ: Tính độ chính xác của kết quả phép tính chu vi đường tròn bán kính 1 cm khi lấy π là 3,14.

Hướng dẫn giải

Khi lấy π là 3,14 ta có chu vi đường tròn bán kính $r = 1$ cm là

$$C_1 = 2.3,14.1 = 6,28 \text{ (cm)}.$$

$$\text{Vì } 3,14 < \pi < 3,15 \text{ nên } 2.3,14.1 < 2\pi.1 < 2.3,15.1$$

$$\Rightarrow 6,28 < C < 6,3$$

$$\Delta_{C_1} = |C - 6,28| < 6,3 - 6,28 = 0,02.$$

Vậy độ chính xác của phép tính này là 0,02.

c. Sai số tương đối

Tỉ số $\delta_a = \frac{\Delta_a}{|a|}$ được gọi là sai số tương đối của số gần đúng a .

Nhận xét:

- Nếu $\bar{a} = a \pm d$ thì $\Delta_a \leq d$. Do đó $\delta_a \leq \frac{d}{|a|}$. Vì vậy, nếu $\frac{d}{|a|}$ càng bé thì chất lượng của

phép đo đạc, tính toán càng cao.

- Người ta thường viết sai số tương đối dưới dạng phần trăm.

Chẳng hạn, trong phép đo thời gian Trái Đất quay một vòng quanh Mặt Trời thì sai

số tương đối không vượt quá $\frac{1}{365} = \frac{1}{1\,460} \approx 0,068\%$.

Ví dụ: Trong phép đo chiều dài của một đoạn đường thu được kết quả là 13,1 m với độ chính xác là 0,1 m. Hãy đánh giá sai số tương đối của số gần đúng này.

Hướng dẫn giải

Ta có số gần đúng $a = 13,1$ m và độ chính xác $d = 0,1$ m.

Do đó sai số tương đối là: $\delta_a \leq \frac{d}{|a|} = \frac{0,1}{13,1} \approx 0,76\%$.

Vậy sai số tương đối không vượt quá 0,76%.

3. Số quy tròn. Quy tròn số đúng và số gần đúng

a. Số quy tròn

Khi quy tròn một số nguyên hoặc một số thập phân đến một hàng nào đó thì số nhận được gọi là số quy tròn của số ban đầu.

Ví dụ: Quy tròn số 5,123 đến hàng phần trăm ta được số 5,12. Khi đó số 5,12 được gọi là số quy tròn của số 5,123.

b. Quy tròn số đến một hàng cho trước

Nhận xét: Khi quy tròn số nguyên hoặc số thập phân đến một hàng cho trước thì sai số tuyệt đối của số quy tròn không vượt quá nửa đơn vị của hàng quy tròn. Như vậy, ta có thể lấy độ chính xác của số quy tròn bằng nửa đơn vị của hàng quy tròn.

Ví dụ: Quy tròn số 2,516 đến hàng phần trăm rồi ước lượng độ chính xác của số đó.

Hướng dẫn giải

Quy tròn số 2,516 đến hàng phần trăm ta được số 2,52.

Sai số tuyệt đối là $|2,516 - 2,52| = 0,004 < 0,005$.

Vậy số quy tròn 2,52 là số gần đúng của 2,516 với độ chính xác 0,005.

c. Quy tròn số gần đúng căn cứ vào độ chính xác cho trước

Quy ước: Cho a là số gần đúng với độ chính xác d . Giả sử a là số nguyên hoặc số thập phân. Khi được yêu cầu quy tròn số a mà không nói rõ quy tròn đến hàng nào thì ta quy tròn a đến hàng thấp nhất mà d nhỏ hơn một đơn vị của hàng đó.

Ví dụ: Viết số quy tròn của số 1 348 với $d = 300$.

Hướng dẫn giải

Vì độ chính xác $d = 300$ thỏa mãn $100 < d = 300 < 1\,000$ nên hàng thấp nhất mà d nhỏ hơn một đơn vị của hàng đó là hàng nghìn.

Vì vậy, ta quy tròn số 1 348 đến hàng nghìn.

Quy tròn số 1 348 đến hàng nghìn ta được số 1 000.

Vậy số quy tròn của số 1 348 với độ chính xác $d = 300$ là 1 000.

4. Số trung bình cộng (Số trung bình)

a. Định nghĩa

Số trung bình cộng của một mẫu n số liệu thống kê bằng tổng của các số liệu chia cho số các số liệu đó. Số trung bình cộng \bar{x} của mẫu số liệu x_1, x_2, \dots, x_n là:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Ví dụ: Tìm số trung bình cộng của các số 13, 15, 17, 20.

Hướng dẫn giải

Mẫu trên có 4 số liệu.

Khi đó, số trung bình cộng là $\bar{x} = \frac{13+15+17+20}{4} = 16,25$.

Vậy trung bình cộng của các số đã cho là 16,25.

Nhận xét:

- Đối với bảng tần số:

Giá trị	x_1	x_2	...	x_k
Tần số	n_1	n_2	...	n_k

Số trung bình cộng \bar{x} của mẫu số liệu thống kê trong bảng phân bố tần số là:

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{n_1 + n_2 + \dots + n_k}.$$

- Đối với bảng tần số tương đối:

Giá trị	x_1	x_2	...	x_k
Tần số tương đối	f_1	f_2	...	f_k

Số trung bình cộng \bar{x} của mẫu số liệu thống kê trong bảng phân bố tần số tương đối là:

$$\bar{x} = f_1x_1 + f_2x_2 + \dots + f_kx_k,$$

trong đó $f_1 = \frac{n_1}{n}$, $f_2 = \frac{n_2}{n}$, ..., $f_k = \frac{n_k}{n}$, với $n = n_1 + n_2 + \dots + n_k$.

Ví dụ:

a) Thời gian giải một bài toán (đơn vị: phút) của 30 học sinh được ghi lại trong bảng tần số sau:

Thời gian (phút)	5	6	7	8	9	10	12	13	15
Số học sinh	5	4	2	3	4	1	3	5	3

Tính thời gian trung bình để giải bài toán trên.

b) Số cân nặng (đơn vị: kg) của 20 học sinh được ghi lại trong bảng tần số tương đối sau:

Cân nặng (kg)	28	29	30	35	37	42
Tần số tương đối	$\frac{1}{10}$	$\frac{3}{20}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{1}{5}$	$\frac{1}{20}$

Hãy tính cân nặng trung bình của 20 học sinh.

Hướng dẫn giải

a) Thời gian trung bình để giải bài toán trên là

$$\bar{x} = \frac{5.5 + 4.6 + 2.7 + 3.8 + 4.9 + 1.10 + 3.12 + 5.13 + 3.15}{5 + 4 + 2 + 3 + 4 + 1 + 3 + 5 + 3} = 9,3.$$

Vậy thời gian trung bình để giải bài toán trên là 9,3 phút.

b) Ta có cân nặng trung bình của 20 học sinh là:

$$\bar{x} = \frac{1}{10}.28 + \frac{3}{20}.29 + \frac{1}{5}.30 + \frac{3}{10}.35 + \frac{1}{5}.37 + \frac{1}{20}.42 = 33,15.$$

Vậy cân nặng trung bình của 20 học sinh là 33,15 kg.

b. Ý nghĩa

Khi các số liệu trong mẫu ít sai lệch với số trung bình cộng, ta có thể giải quyết được vấn đề trên bằng cách lấy số trung bình cộng làm đại diện cho mẫu số liệu.

Ví dụ: Để dự báo lượng mưa trong tháng 8 tại Hà Nội người ta tiến hành đo lượng mưa của từng ngày của tháng 8 gồm 31 số liệu. Số trung bình cộng của mẫu số liệu đó được xem như lượng mưa trung bình tháng 8 tại Hà Nội. Thống kê lượng mưa trung bình tháng 8 tại Hà Nội trong nhiều năm liên tiếp sẽ cho ta những dự báo lượng mưa trung bình tháng 8 tại Hà Nội trong những năm sắp tới.

5. Trung vị

a. Định nghĩa

Sắp thứ tự mẫu số liệu gồm n số liệu thành một dãy không giảm (hoặc không tăng).

- Nếu n là số lẻ thì số liệu đứng ở vị trí thứ $\frac{n+1}{2}$ (số đứng chính giữa) gọi là *trung vị*.

- Nếu n là số chẵn thì số trung bình cộng của hai số liệu đứng ở vị trí thứ $\frac{n}{2}$ và $\frac{n}{2} + 1$ gọi là *trung vị*.

Trung vị kí hiệu là M_e .

Nhận xét:

- Trung vị không nhất thiết là một số trong mẫu số liệu và dễ tính toán.
- Khi các số liệu trong mẫu không có sự chênh lệch lớn thì số trung bình cộng và trung vị xấp xỉ nhau.

Ví dụ: Điểm kiểm tra Toán của 7 bạn học sinh tổ 1 lớp 10B như sau: 9; 5; 4; 5; 8; 7; 9. Tìm trung vị M_e của mẫu số liệu trên.

Hướng dẫn giải

- Sắp xếp số liệu của mẫu theo thứ tự không giảm:

4 5 5 7 8 9 9

- Xác định xem số các số liệu là chẵn hay lẻ để tìm số trung vị:

Mẫu có 7 số liệu. Giá trị chính giữa là 7. Vì thế, trung vị của mẫu là 7.

Vậy $M_e = 7$.

b. Ý nghĩa

Nếu những số liệu trong mẫu có sự chênh lệch lớn thì ta nên chọn thêm trung vị làm đại diện cho mẫu số liệu đó nhằm điều chỉnh một số hạn chế khi sử dụng số trung bình cộng. Những kết luận về đối tượng thống kê rút ra khi đó sẽ tin cậy hơn.

Ví dụ: Thời gian giải một bài tập (đơn vị: phút) của nhóm học sinh như sau:

20 3 2 5 6 1

Tính trung vị của mẫu và số trung bình cộng của mẫu.

Ta nên chọn trung vị hay số trung bình cộng để đại diện cho mẫu thì kết luận về thời gian giải một bài tập của nhóm học sinh sẽ đáng tin cậy hơn?

Hướng dẫn giải

Sắp xếp các số liệu theo thứ tự không giảm:

1 2 3 5 6 20

Mẫu có 6 số liệu, khi đó trung vị của mẫu là trung bình cộng của 3 và 5.

$$\text{Ta có } M_e = \frac{3+5}{2} = 4.$$

$$\text{Trung bình cộng của mẫu số liệu: } \bar{x} = \frac{1+2+3+5+6+20}{6} \approx 6,2.$$

Ta thấy nên lựa chọn trung vị $M_e = 4$ đại diện cho mẫu thì kết luận thời gian giải một bài tập của nhóm học sinh sẽ đáng tin cậy hơn.

Vậy trung vị của mẫu là $M_e = 4$; số trung bình cộng là 6,2 và nên lựa chọn trung vị $M_e = 4$ đại diện cho mẫu thì kết luận thời gian giải một bài tập của nhóm học sinh sẽ đáng tin cậy hơn.

6. Tứ phân vị

a. Định nghĩa

Sắp thứ tự mẫu số liệu gồm n số liệu thành một dãy không giảm.

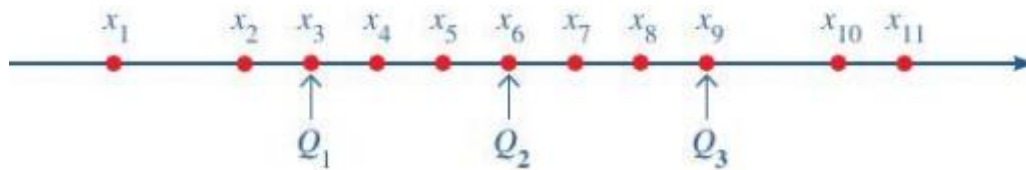
Tứ phân vị của mẫu số liệu trên là bộ ba giá trị: tứ phân vị thứ nhất, tứ phân vị thứ hai và tứ phân vị thứ ba; ba giá trị này chia mẫu số liệu thành bốn phần có số lượng phần tử bằng nhau.

- Tứ phân vị thứ hai Q_2 bằng trung vị.

- Nếu n là số chẵn thì tứ phân vị thứ nhất Q_1 bằng trung vị của nửa dãy phía dưới và tứ phân vị thứ ba Q_3 bằng trung vị của nửa dãy phía trên.

- Nếu n là số lẻ thì tứ phân vị thứ nhất Q_1 bằng trung vị của nửa dãy phía dưới (không bao gồm Q_2) và tứ phân vị thứ ba Q_3 bằng trung vị của nửa dãy phía trên (không bao gồm Q_2).

Ta minh họa tứ phân vị của mẫu số liệu gồm 11 số liệu trên trục số như sau:



Ví dụ: Tìm tứ phân vị của mẫu số liệu sau:

21 32 10 45 11 35 24 8

Hướng dẫn giải

Mẫu số liệu trên được sắp xếp theo thứ tự tăng dần như sau:

8 10 11 21 24 32 35 45

Dãy số liệu trên gồm 8 số liệu, là số chẵn.

Do đó ta có:

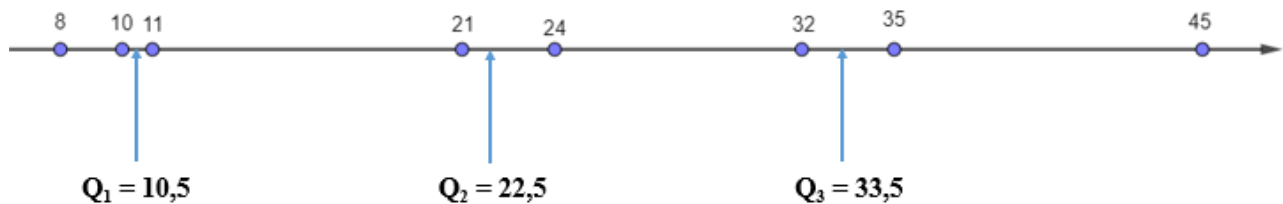
- Trung vị của mẫu số liệu trên là: $Q_2 = \frac{21 + 24}{2} = 22,5$.

• Trung vị của dãy 8, 10, 11, 21 là $Q_1 = \frac{10+11}{2} = 10,5$.

• Trung vị của dãy 24, 32, 35, 45 là $Q_3 = \frac{32+35}{2} = 33,5$.

Vậy $Q_1 = 10,5$, $Q_2 = 22,5$, $Q_3 = 33,5$.

Tứ phân vị đó được biểu diễn trên trục số như sau:



b. Ý nghĩa

- Trong thực tiễn, có những mẫu số liệu mà nhiều số liệu trong mẫu đó vẫn còn sự chênh lệch lớn so với trung vị. Ta nên chọn thêm những số khác cùng làm đại diện cho mẫu đó. Bằng cách lấy thêm trung vị của từng dãy số liệu tách ra bởi trung vị của mẫu nói trên, ta nhận được tứ phân vị đại diện cho mẫu số liệu đó.

- Bộ ba giá trị Q_1 , Q_2 , Q_3 trong tứ phân vị phản ánh độ phân tán của mẫu số liệu. Nhưng mỗi giá trị Q_1 , Q_2 , Q_3 lại đo xu thế trung tâm của phần số liệu tương ứng của mẫu đó.

7. Một

a. Định nghĩa

Mốt của mẫu số liệu là giá trị có tần số lớn nhất trong bảng phân bố tần số và kí hiệu là M_0 .

Chú ý: Một mẫu số liệu có thể có nhiều một.

Ví dụ: Cho bảng tần số sau:

Giá trị	1	2	3	4	5	6	7
Tần số	12	14	6	25	25	7	6

Tìm một của mẫu số liệu trên.

Hướng dẫn giải

Quan sát bảng tần số ta thấy giá trị 4 và 5 có tần số lớn nhất bằng 25.

Suy ra một của dấu hiệu là $M_o = 4$ và $M_o = 5$.

Vậy một của dấu hiệu là $M_o = 4$ và $M_o = 5$.

b. Ý nghĩa

Một của một mẫu số liệu đặc trưng cho số lần lặp đi lặp lại nhiều nhất tại một vị trí của mẫu số liệu đó. Dựa vào một, ta có thể đưa ra những kết luận (có ích) về đối tượng thống kê.

Ví dụ: Một cửa hàng bán 5 loại quạt với giá tiền là 150; 200; 350; 400; 500 (nghìn đồng). Số quạt bán ra trong mùa hè vừa qua được thống kê trong bảng sau:

Giá tiền	150	200	350	400	500
Số quạt bán được	25	80	100	123	75

Năm nay cửa hàng nên nhập nhiều số lượng loại quạt có giá tiền bao nhiêu để bán?

Hướng dẫn giải

Quan sát bảng thống kê trên ta thấy quạt có giá 400 nghìn đồng có số lượng bán được nhiều nhất, nghĩa là quạt giá 400 nghìn có tần số lớn nhất.

Suy ra $M_o = 400$.

Vậy năm nay cửa hàng nên nhập nhiều quạt có giá tiền 400 nghìn đồng về để bán.

8. Tính hợp lí của số liệu thống kê

Sau khi thu thập, tổ chức, phân loại và biểu diễn số liệu bằng bảng hoặc biểu đồ, ta cần phân tích và xử lí các số liệu đó để xem xét tính hợp lí của số liệu thống kê, đặc biệt chỉ ra được những số liệu bất thường (hay còn gọi là dị biệt, trong tiếng Anh là Outliers). Ta có thể sử dụng các số liệu đặc trưng đo xu thế trung tâm cho mẫu số liệu không ghép nhóm để thực hiện điều đó.

Ví dụ: Chiều cao của một nhóm học sinh nữ 6 tuổi (đơn vị cm) được ghi lại như sau:

100	112	113	115	115
116	118	117	120	122
122	125	123	162	97

a) Tìm tứ phân vị của mẫu trên.

b) Từ kết quả câu a) bước đầu xác định những giá trị bất thường của mẫu số liệu trên.

Hướng dẫn giải

a) Ta sắp xếp các số liệu theo thứ tự tăng dần như sau:

97	100	112	113	115
115	116	117	118	120
122	122	123	125	145

Mẫu trên có 15 số liệu.

Trung vị của mẫu số liệu trên là $Q_2 = 117$.

Trung vị nửa phía dưới 97, 100, 112, 113, 115, 115, 116 là $Q_1 = 113$.

Trung vị nửa phía trên 118, 120, 122, 122, 123, 125, 145 là $Q_3 = 122$.

Vậy tứ phân vị của mẫu là $Q_1 = 113$; $Q_2 = 117$; $Q_3 = 122$.

b) Dựa vào trung vị và tứ phân vị của mẫu số liệu, bước đầu ta thấy những số liệu bất thường trong mẫu là 97 và 145.

9. Khoảng biến thiên. Khoảng tứ phân vị

a. Định nghĩa

- Trong một mẫu số liệu, *khoảng biến thiên* là hiệu số giữa giá trị lớn nhất và giá trị nhỏ nhất của mẫu số liệu đó.

Ta có thể tính khoảng biến thiên R của mẫu số liệu theo công thức sau: $R = x_{\max} - x_{\min}$, trong đó x_{\max} là giá trị lớn nhất, x_{\min} là giá trị nhỏ nhất của mẫu số liệu đó.

- Giả sử Q_1, Q_2, Q_3 là tứ phân vị của mẫu số liệu. Ta gọi hiệu $\Delta_Q = Q_3 - Q_1$ là *khoảng tứ phân vị* của mẫu số liệu đó.

Chú ý: Khoảng tứ phân vị của mẫu số liệu còn gọi là khoảng trải giữa (tiếng Anh là InterQuartile Range – IQR) của mẫu số liệu đó.

Ví dụ: Mẫu số liệu thống kê cân nặng (đơn vị: kg) của 8 học sinh trong một tổ như sau:

45 52 41 37 50 52 66 49

a) Tìm khoảng biến thiên của mẫu số liệu trên.

b) Tìm khoảng tứ phân vị của mẫu số liệu trên.

Hướng dẫn giải

a) Trong mẫu có số liệu lớn nhất là 66, số liệu nhỏ nhất là 37.

Khi đó, khoảng biến thiên của mẫu là $R = x_{\max} - x_{\min} = 66 - 37 = 29$ (kg).

Vậy khoảng biến thiên của mẫu là $R = 29$ kg.

b) Sắp xếp mẫu theo thứ tự tăng dần, ta được:

37 41 45 49 50 52 52 66

Khi đó, trung vị của mẫu là: $Q_2 = \frac{49+50}{2} = 49,5$.

Q_1 là trung vị của mẫu 37, 41, 45, 49 nên $Q_1 = \frac{41+45}{2} = 43$.

Q_3 là trung vị của mẫu 50, 52, 52, 66 nên $Q_3 = \frac{52+52}{2} = 52$.

Khi đó, ta có khoảng tứ phân vị là:

$$\Delta_Q = Q_3 - Q_1 = 52 - 43 = 9 \text{ (kg)}.$$

Vậy khoảng tứ phân vị của mẫu là $\Delta_Q = 9 \text{ kg}$.

b. Ý nghĩa

- Ý nghĩa của khoảng biến thiên: *Khoảng biến thiên của mẫu số liệu phản ánh sự “dao động”, “sự dàn trải” của các số liệu trong mẫu đó. Khoảng biến thiên được sử dụng trong nhiều tình huống thực tiễn, chẳng hạn: tìm ra sự phân tán điểm kiểm tra của một lớp học hay xác định phạm vi giá cả của một dịch vụ ...*

Theo cách nhìn như ở trong vật lí, ở đó biên độ dao động phản ánh khoảng cách từ điểm cân bằng đến điểm xa nhất của dao động, nếu coi số trung bình cộng là “điểm cân bằng” của mẫu số liệu thì khoảng biến thiên của mẫu số liệu có thể xem như hai lần biên độ dao động của các số liệu trong mẫu đó quanh điểm cân bằng.

Trong các đại lượng đo mức độ phân tán của mẫu số liệu, khoảng biến thiên là đại lượng dễ hiểu, dễ tính toán và tương đối tốt đối với các mẫu số liệu nhỏ. Tuy nhiên, do khoảng biến thiên chỉ sử dụng hai giá trị x_{\max} và x_{\min} của mẫu số liệu nên đại lượng đó chưa diễn giải đầy đủ sự phân tán của số liệu trong mẫu. Ngoài ra, giá trị của khoảng biến thiên sẽ bị ảnh hưởng bởi các giá trị bất thường của mẫu số liệu đó.

Trong những trường hợp như vậy, khoảng biến thiên của mẫu số liệu không phản ánh chính xác độ dàn trải của mẫu số liệu.

- **Ý nghĩa của khoảng tứ phân vị:** Khoảng tứ phân vị là đại lượng cho biết mức độ phân tán của 50% số liệu chính giữa của mẫu số liệu đã sắp xếp và có thể giúp xác định các giá trị bất thường của mẫu số liệu đó. Khoảng tứ phân vị thường được sử dụng thay cho khoảng biến thiên vì nó loại trừ hầu hết giá trị bất thường của mẫu số liệu.

10. Phương sai

a. Định nghĩa

- Mỗi hiệu số giữa số liệu và số trung bình cộng gọi là *độ lệch* của số liệu đó đối với số trung bình cộng.

- Cho mẫu số liệu thống kê có n giá trị x_1, x_2, \dots, x_n và số trung bình cộng là \bar{x} .

Ta gọi số
$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$
 là *phương sai* của mẫu số liệu trên.

Nhận xét:

- Khi có các số liệu bằng nhau, ta có thể tính phương sai theo công thức sau:

+ Đối với bảng tần số:

Giá trị	x_1	x_2	...	x_k
Tần số	n_1	n_2	...	n_k

Phương sai của mẫu số liệu thống kê trong bảng phân bố tần số là:

$$s^2 = \frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{n},$$

trong đó $n = n_1 + n_2 + \dots + n_k$; \bar{x} là số trung bình cộng của các số liệu đã cho.

+ Đối với bảng phân bố tần số tương đối:

Giá trị	x_1	x_2	...	x_k
Tần số tương đối	f_1	f_2	...	f_k

Phương sai của mẫu số liệu thống kê trong bảng phân bố tần số tương đối là:

$$s^2 = f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_k(x_k - \bar{x})^2,$$

trong đó \bar{x} là số trung bình cộng của số liệu đã cho.

- Trong thực tế, người ta còn dùng công thức sau để tính phương sai của mẫu số liệu:

$$\hat{s}^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1},$$

trong đó: x_i là giá trị của quan sát thứ i ; \bar{x} là giá trị trung bình và n là số quan sát trong mẫu số liệu đó.

Ví dụ: Hai lớp 10A và 10B của một trường THPT đồng thời làm bài thi môn Toán theo cùng một đề thi. Kết quả được ghi lại trong bảng tần số sau:

Điểm thi của lớp 10A:

Điểm thi	5	6	7	8	9	10
----------	---	---	---	---	---	----

Số học sinh	3	7	12	14	3	1
-------------	---	---	----	----	---	---

Điểm thi của lớp 10B:

Điểm thi	6	7	8	9
Số học sinh	8	18	10	4

a) Tính phương sai của từng mẫu số liệu ở hai bảng trên.

b) Xét xem kết quả bài thi của lớp nào đồng đều hơn.

Hướng dẫn giải

Ta có điểm thi trung bình của lớp 10A là:

$$\bar{x}_{10A} = \frac{3.5 + 7.6 + 12.7 + 14.8 + 3.9 + 1.10}{3 + 7 + 12 + 14 + 3 + 1} = 7,25$$

Điểm thi trung bình của lớp 10A là:

$$\bar{x}_{10B} = \frac{8.6 + 18.7 + 10.8 + 4.9}{8 + 18 + 10 + 4} = 7,25$$

Ta có phương sai của mẫu số liệu lớp 10A là:

$$s_{10A}^2 = \frac{3(5 - 7,25)^2 + 7(6 - 7,25)^2 + \dots + 1.(10 - 7,25)^2}{3 + 7 + 12 + 14 + 3 + 1} = 1,2875$$

$$\Rightarrow s_{10A}^2 = 1,2875$$

Ta có phương sai của mẫu số liệu lớp 10B là:

$$s_{10B}^2 = \frac{8(6 - 7,25)^2 + 18(7 - 7,25)^2 + 10.(8 - 7,25)^2 + 4.(9 - 7,25)^2}{8 + 18 + 10 + 4} = 0,7875$$

$$\Rightarrow s_{10B}^2 = 0,7875.$$

Ta thấy điểm thi trung bình của lớp 10A và 10B bằng nhau đều bằng 7,25. Nhưng phương sai bảng điểm của lớp 10A lại lớn hơn phương sai bảng điểm lớp 10B nên kết quả làm bài thi của lớp 10B đồng đều hơn lớp 10A.

b. Ý nghĩa

Phương sai là số đặc trưng đo mức độ phân tán của mẫu số liệu. Mẫu số liệu nào có phương sai nhỏ hơn thì mức độ phân tán (so với số trung bình cộng) của các số liệu trong mẫu đó sẽ thấp hơn.

11. Độ lệch chuẩn

a. Định nghĩa

Căn bậc hai (số học) của phương sai gọi là *độ lệch chuẩn* của mẫu số liệu thống kê.

Nhận xét: Vì độ lệch chuẩn có cùng đơn vị đo với số liệu thống kê nên khi cần chú ý đến đơn vị đo thì ta sử dụng độ lệch chuẩn mà không sử dụng phương sai.

Ví dụ: Cho mẫu số liệu:

23 22 20 12 35

Tính độ lệch chuẩn của mẫu trên.

Hướng dẫn giải

Mẫu trên có 5 số liệu.

Số trung bình của mẫu trên là: $\bar{x} = \frac{23 + 22 + 20 + 12 + 35}{5} = 22,4$.

Phương sai của mẫu số liệu đó là:

$$s^2 = \frac{(23 - 22,4)^2 + (22 - 22,4)^2 + (20 - 22,4)^2 + (12 - 22,4)^2 + (35 - 22,4)^2}{5} = 54,64$$

$$\Rightarrow s = \sqrt{s^2} = \sqrt{54,64} \approx 7,39.$$

Vậy độ lệch chuẩn của mẫu là 7,39.

b. Ý nghĩa

Cũng như phương sai, khi hai mẫu số liệu thống kê có cùng đơn vị đo và có số trung bình cộng bằng nhau (hoặc xấp xỉ nhau), mẫu số liệu nào có độ lệch chuẩn nhỏ hơn thì mức độ phân tán (so với số trung bình cộng) của các số liệu trong mẫu đó sẽ thấp hơn. Độ lệch chuẩn là số đặc trưng đo mức độ phân tán của mẫu số liệu thống kê có cùng đơn vị đo.

12. Giá trị bất thường của số liệu thống kê

Ta có thể sử dụng các số đặc trưng đo mức độ phân tán cho mẫu số liệu không ghép nhóm để chỉ ra được những số liệu bất thường của mẫu số liệu đó. Ta thường sử dụng khoảng tứ phân vị để xác định số liệu bất thường của mẫu số liệu. Cụ thể như sau:

Giả sử Q_1, Q_2, Q_3 là tứ phân vị của mẫu số liệu và hiệu $\Delta_Q = Q_3 - Q_1$ là khoảng tứ phân vị của mẫu số liệu đó. Một giá trị trong mẫu số liệu được coi là một giá trị bất thường nếu nó nhỏ hơn $Q_1 - \frac{3}{2}\Delta_Q$ hoặc lớn hơn $Q_3 + \frac{3}{2}\Delta_Q$. Như vậy, khoảng tứ phân vị cho ta cách nhận biết giá trị bất thường của mẫu số liệu.

Chú ý: Ta cũng có thể xác định số liệu bất thường của mẫu số liệu bằng số trung bình cộng và độ lệch chuẩn. Cụ thể như sau:

Giả sử \bar{x} , s lần lượt là số trung bình cộng và độ lệch chuẩn của mẫu số liệu. Một giá trị trong mẫu số liệu cũng được coi là một giá trị bất thường nếu nó nhỏ hơn $\bar{x} - 3s$ hoặc lớn hơn $\bar{x} + 3s$. Như vậy, số trung bình cộng và độ lệch chuẩn cho ta cách nhận ra giá trị bất thường của mẫu số liệu.

Ví dụ: Hãy tìm các giá trị bất thường của mẫu số liệu sau:

12 4 10 -5 6 7 9 30

Hướng dẫn giải

Mẫu được sắp xếp theo thứ tự tăng dần là:

-5 4 6 7 9 10 12 30

Mẫu có 8 số liệu

Trung vị của mẫu là: $\frac{7+9}{2} = 8$. Suy ra $Q_2 = 8$.

Trung vị nửa dưới $-5, 4, 6, 7$ là $\frac{4+6}{2} = 5$. Suy ra $Q_1 = 5$.

Trung vị nửa trên $9, 10, 12, 30$ là $\frac{10+12}{2} = 11$. Suy ra $Q_3 = 11$.

Khoảng tứ phân vị là $\Delta_Q = Q_3 - Q_1 = 11 - 5 = 6$.

Ta có: $Q_1 - \frac{3}{2}\Delta_Q = 5 - \frac{3}{2}.6 = -4$; $Q_3 + \frac{3}{2}\Delta_Q = 11 + \frac{3}{2}.6 = 20$.

Ta thấy $-5 < -4$ và $30 > 20$ nên các giá trị -5 và 30 là các giá trị bất thường của mẫu. Vậy mẫu có hai giá trị bất thường là -5 và 30 .

13. Xác suất của biến cố trong trò chơi tung đồng xu

- Trong trò chơi tung đồng xu, ta quy ước đồng xu là cân đối và đồng chất.

Nhận xét:

Khi tung một đồng xu hai lần liên tiếp:

+ Tập hợp Ω các kết quả có thể xảy ra đối với mặt xuất hiện của đồng xu sau hai lần tung là $\Omega = \{SS; SN; NS; NN\}$, trong đó, chẳng hạn SN là kết quả “Lần thứ nhất đồng xu xuất hiện mặt sấp, lần thứ hai đồng xu xuất hiện mặt ngửa”.

+ Tập hợp Ω gọi là *không gian mẫu* trong trò chơi tung một đồng xu hai lần liên tiếp.

- Trong trò chơi tung một đồng xu hai lần liên tiếp, đối với mỗi biến cố A ta có định nghĩa cổ điển của xác suất như sau:

Xác suất của biến cố A, kí hiệu $P(A)$, là tỉ số giữa số các kết quả thuận lợi cho biến cố A và số phần tử của không gian mẫu Ω :

$$P(A) = \frac{n(A)}{n(\Omega)},$$

ở đó $n(A)$, $n(\Omega)$ lần lượt là số phần tử của hai tập hợp A và Ω .

Ví dụ: Tung một đồng xu hai lần liên tiếp.

- a) Viết tập hợp Ω là không gian mẫu của trò chơi trên.
- b) Xét biến cố A: “Có một lần đồng xu xuất hiện mặt sấp”. Tính xác suất của biến cố A.

Hướng dẫn giải

a) Không gian mẫu của trò chơi trên là: $\Omega = \{SS; SN; NS; NN\}$.

Vậy không gian mẫu là $\Omega = \{SS; SN; NS; NN\}$.

b) Từ $\Omega = \{SS; SN; NS; NN\} \Rightarrow n(\Omega) = 4$.

Có hai kết quả thuận lợi cho biến cố A là SN, NS, tức là $A = \{SN; NS\} \Rightarrow n(A) = 2$.

$$\Rightarrow \text{Xác suất của biến cố A là } P(A) = \frac{n(A)}{n(\Omega)} = \frac{2}{4} = \frac{1}{2}.$$

Vậy xác suất của biến cố A bằng $\frac{1}{2}$.

14. Xác suất của biến cố trong trò chơi gieo xúc xắc

- Trong trò chơi gieo xúc xắc, ta quy ước xúc xắc là cân đối và đồng chất.

Nhận xét: Khi gieo một xúc xắc hai lần liên tiếp, có 36 kết quả có thể xảy ra đối với mặt xuất hiện của xúc xắc sau hai lần gieo, đó là:

(1; 1)	(1; 2)	(1; 3)	(1; 4)	(1; 5)	(1; 6)
(2; 1)	(2; 2)	(2; 3)	(2; 4)	(2; 5)	(2; 6)
(3; 1)	(3; 2)	(3; 3)	(3; 4)	(3; 5)	(3; 6)
(4; 1)	(4; 2)	(4; 3)	(4; 4)	(4; 5)	(4; 6)
(5; 1)	(5; 2)	(5; 3)	(5; 4)	(5; 5)	(5; 6)
(6; 1)	(6; 2)	(6; 3)	(6; 4)	(6; 5)	(6; 6)

+ Tập hợp Ω các kết quả có thể xảy ra đối với mặt xuất hiện của xúc xắc sau hai lần gieo là $\Omega = \{(i; j) | i, j = 1, 2, 3, 4, 5, 6\}$, trong đó $(i; j)$ là kết quả “Lần đầu xuất hiện mặt i chấm, lần sau xuất hiện mặt j chấm”.

+ Tập hợp Ω gọi là *không gian mẫu* trong trò chơi gieo một xúc xắc hai lần liên tiếp.

- Trong trò chơi gieo một xúc xắc hai lần liên tiếp, đối với mỗi biến cố C ta có định nghĩa cổ điển của xác suất như sau:

Xác suất của biến cố C, kí hiệu $P(C)$, là tỉ số giữa số các kết quả thuận lợi cho biến cố C và số phần tử của không gian mẫu Ω :

$$P(C) = \frac{n(C)}{n(\Omega)},$$

ở đó $n(C)$, $n(\Omega)$ lần lượt là số phần tử của hai tập hợp C và Ω .

Ví dụ: Gieo một con xúc xắc hai lần liên tiếp.

a) Viết tập hợp Ω là không gian mẫu của trò chơi trên.

b) Xét biến cố C: “Hai lần gieo xuất hiện ít nhất một mặt là số chẵn”. Tính xác suất của biến cố C.

Hướng dẫn giải

Không gian mẫu của trò chơi gieo một con xúc xắc hai lần là: $\Omega = \{(i; j) | i, j = 1, 2, 3, 4, 5, 6\}$, trong đó $(i; j)$ là kết quả “Lần đầu xuất hiện mặt i chấm, lần sau xuất hiện mặt j chấm”.

Vậy ta có không gian mẫu của trò chơi gieo một con xúc xắc hai lần là:

$\Omega = \{(1; 1); (1; 2); (1; 3); (1; 4); (1; 5); (1; 6); (2; 1); (2; 2); (2; 3); (2; 4); (2; 5); (2; 6); (3; 1); (3; 2); (3; 3); (3; 4); (3; 5); (3; 6); (4; 1); (4; 2); (4; 3); (4; 4); (4; 5); (4; 6); (5; 1); (5; 2); (5; 3); (5; 4); (5; 5); (5; 6); (6; 1); (6; 2); (6; 3); (6; 4); (6; 5); (6; 6)\}$.

b) Từ $\Omega = \{(1; 1); (1; 2); (1; 3); (1; 4); (1; 5); (1; 6); (2; 1); (2; 2); (2; 3); (2; 4); (2; 5); (2; 6); (3; 1); (3; 2); (3; 3); (3; 4); (3; 5); (3; 6); (4; 1); (4; 2); (4; 3); (4; 4); (4; 5); (4; 6); (5; 1); (5; 2); (5; 3); (5; 4); (5; 5); (5; 6); (6; 1); (6; 2); (6; 3); (6; 4); (6; 5); (6; 6)\}$.

$\Rightarrow n(\Omega) = 36$.

Các kết quả thuận lợi cho biến cố C là: $(1; 2); (1; 4); (1; 6); (2; 1); (2; 2); (2; 3); (2; 4); (2; 5); (2; 6); (3; 2); (3; 4); (3; 6); (4; 1); (4; 2); (4; 3); (4; 4); (4; 5); (4; 6); (5; 2); (5; 4); (5; 6); (6; 1); (6; 2); (6; 3); (6; 4); (6; 5); (6; 6)$.

$\Rightarrow C = \{(1; 2); (1; 4); (1; 6); (2; 1); (2; 2); (2; 3); (2; 4); (2; 5); (2; 6); (3; 2); (3; 4); (3; 6); (4; 1); (4; 2); (4; 3); (4; 4); (4; 5); (4; 6); (5; 2); (5; 4); (5; 6); (6; 1); (6; 2); (6; 3); (6; 4); (6; 5); (6; 6)\}.$

$\Rightarrow n(C) = 27.$

$$\Rightarrow P(C) = \frac{n(C)}{n(\Omega)} = \frac{27}{36} = \frac{3}{4}.$$

Vậy xác suất của biến cố C là $\frac{3}{4}$.

15. Một số khái niệm về xác suất

a. Phép thử ngẫu nhiên và không gian mẫu

Có những phép thử mà ta không thể đoán được kết quả của nó, mặc dù đã biết tập hợp các kết quả có thể của phép thử đó. Những phép thử như thế gọi là phép thử ngẫu nhiên (gọi tắt là phép thử).

Tập hợp Ω các kết quả có thể xảy ra của một phép thử gọi là không gian mẫu của phép thử đó.

Ví dụ: Viết không gian mẫu của phép thử tung một đồng xu ba lần.

Hướng dẫn giải

Khi tung một đồng xu thì có hai kết quả có thể là đồng xu xuất hiện mặt sấp (S) hoặc đồng xu xuất hiện mặt ngửa (N).

Khi đó, tung ba đồng xu thì có các kết quả có thể là: SSS; SSN; SNN; SNS; NSS; NSN; NNS; NNN.

Suy ra không gian mẫu của phép thử là $\Omega = \{SSS; SSN; SNN; SNS; NSS; NSN; NNS; NNN\}.$

Vậy không gian mẫu của phép thử tung đồng xu ba lần là: $\Omega = \{SSS; SSN; SNN; SNS; NSS; NSN; NNS; NNN\}.$

b. Biến cố

* Định nghĩa

Nhận xét:

- Mỗi sự kiện liên quan đến phép thử T tương ứng với một (và chỉ một) tập con A của không gian mẫu Ω .
- Ngược lại, mỗi tập con A của không gian mẫu Ω có thể phát biểu dưới dạng mệnh đề nêu sự kiện liên quan đến phép thử T .

Định nghĩa:

Biến cố ngẫu nhiên (gọi tắt là biến cố) là một tập con của không gian mẫu.

Chú ý: Vì sự kiện chỉ ra tính chất đặc trưng cho các phần tử của một biến cố nên ta cũng gọi sự kiện là biến cố. Chẳng hạn “Kết quả của hai lần tung là giống nhau” trong phép thử “Tung một đồng xu hai lần liên tiếp” là một biến cố.

Ví dụ: Với phép thử tung đồng xu ba lần liên tiếp. Biến cố A : “Có ít nhất hai lần xuất hiện mặt sấp” là tập con nào của không gian mẫu.

Hướng dẫn giải

Phép thử tung đồng xu ba lần có không gian mẫu là: $\Omega = \{SSS; SSN; SNN; SNS; NSS; NSN; NNS; NNN\}$.

Khi đó, biến cố A tương ứng với tập con $\{SSS; SSN; SNS; NSS\}$.

Vậy $A = \{SSS; SSN; SNS; NSS\}$.

*** Biến cố không. Biến cố chắc chắn**

Xét phép thử T với không gian mẫu Ω . Mỗi biến cố là một tập con của tập Ω . Vì thế, tập hợp \emptyset cũng là một biến cố, gọi là *biến cố không thể* (gọi tắt là *biến cố không*). Còn tập hợp Ω gọi là *biến cố chắc chắn*.

Ví dụ: Khi gieo một con xúc xắc hai lần liên tiếp. Biến cố A : “Tổng số chấm của hai lần gieo bằng 1” là biến cố không. Biến cố B : “Tổng số chấm hai lần gieo nhỏ hơn 13” là biến cố chắc chắn.

*** Biến cố đối**

Tập con $\Omega \setminus A$ xác định một biến cố, gọi là biến cố đối của biến cố A , kí hiệu là \bar{A} .

Chú ý: Nếu biến cố A được mô tả dưới dạng mệnh đề toán học Q thì biến cố đối \bar{A} được mô tả bằng mệnh đề phủ định của mệnh đề Q (tức là mệnh đề \bar{Q}).

Ví dụ: Xét phép thử “Tung một đồng xu”. Hãy xác định biến cố đối của biến cố A: “Đồng xu xuất hiện mặt ngửa”.

Hướng dẫn giải

Khi tung một đồng xu thì sẽ xuất hiện mặt sấp (S) hoặc mặt ngửa (N).

Khi đó biến cố đối của biến cố A: “Đồng xu xuất hiện mặt ngửa” là \bar{A} : “Đồng xu xuất hiện mặt sấp”.

Vậy biến cố đối của biến cố A là \bar{A} : “Đồng xu xuất hiện mặt sấp”.

c. Xác suất của biến cố

Xác suất của biến cố A, kí hiệu là $P(A)$, bằng tỉ số $\frac{n(A)}{n(\Omega)}$, ở đó $n(A)$, $n(\Omega)$ lần lượt

là số phần tử của hai tập hợp A và Ω . Như vậy $P(A) = \frac{n(A)}{n(\Omega)}$.

Ví dụ: Với phép thử tung đồng xu ba lần liên tiếp. Tính xác suất của biến cố A: “Có ít nhất hai lần xuất hiện mặt sấp”.

Hướng dẫn giải

Phép thử tung đồng xu ba lần có không gian mẫu là: $\Omega = \{SSS; SSN; SNN; SNS; NSS; NSN; NNS; NNN\}$.

$$\Rightarrow n(\Omega) = 8.$$

Khi đó, các kết quả thuận lợi cho biến cố A là: SSS; SSN; SNS; NSS.

$$\Rightarrow A = \{SSS; SSN; SNS; NSS\}.$$

$$\Rightarrow n(A) = 4.$$

$$\Rightarrow P(A) = \frac{n(A)}{n(\Omega)} = \frac{4}{8} = \frac{1}{2}.$$

Vậy xác suất của biến cố A: “Có ít nhất hai lần xuất hiện mặt sấp” là $\frac{1}{2}$.

16. Tính chất của xác suất

Xét phép thử T với không gian mẫu là Ω . Khi đó, ta có các tính chất sau:

- +) $P(\emptyset) = 0$; $P(\Omega) = 1$;
- +) $0 \leq P(A) \leq 1$ với mỗi biến cố A;
- +) $P(\overline{A}) = 1 - P(A)$ với mỗi biến cố A.

Ví dụ: Trong túi có 3 quả bóng màu xanh và 2 quả bóng màu vàng, các quả bóng có kích thước và khối lượng giống nhau. Lấy đồng thời ngẫu nhiên 2 quả bóng.

Tính xác suất của các biến cố:

A: “Hai quả bóng lấy ra không phải màu xanh và màu vàng”.

B: “Hai quả bóng lấy ra là màu xanh hoặc màu vàng”.

C: “Hai quả bóng lấy ra khác màu”.

Hướng dẫn giải

Do trong túi chỉ có hai loại bóng màu xanh và màu vàng nên khi lấy ngẫu nhiên hai quả bóng trong túi thì hai quả bóng lấy ra phải là bóng màu xanh hoặc màu vàng.

Do đó biến cố A: “Hai quả bóng lấy ra không phải màu xanh và màu vàng” là biến cố không thể, tức là $A = \emptyset$.

Suy ra $P(A) = P(\emptyset) = 0$.

Biến cố B: “Hai quả bóng lấy ra là màu xanh hoặc màu vàng” luôn luôn xảy ra.

$\Rightarrow B = \Omega$

$\Rightarrow P(B) = P(\Omega) = 1$.

Ta có 3 quả bóng màu xanh, 2 quả bóng màu vàng, nên trong túi có $3 + 2 = 5$ quả bóng.

Khi lấy ngẫu nhiên ra 2 trong 5 quả bóng, ta có $C_5^2 = 10$ (cách).

Suy ra không gian mẫu Ω có 10 phần tử.

$$\Rightarrow n(\Omega) = 10.$$

Xét biến cố C: “Hai quả bóng lấy ra khác màu”.

Ta có biến cố đối của C là \bar{C} : “Hai quả bóng lấy ra cùng màu”.

Suy ra hai quả bóng lấy ra cùng là màu xanh hoặc cùng là màu vàng.

+ Hai quả bóng lấy ra cùng là màu xanh, tức là lấy được 2 trong 3 quả bóng màu xanh, có $C_3^2 = 3$ (cách).

+ Hai quả bóng lấy ra cùng là màu vàng, tức là lấy được 2 trong 2 quả bóng màu vàng, có $C_2^2 = 1$ (cách).

Suy ra số cách lấy được hai quả bóng cùng màu là: $3 + 1 = 4$ (cách)

$$\Rightarrow n(\bar{C}) = 4.$$

$$\Rightarrow P(\bar{C}) = \frac{n(\bar{C})}{n(\Omega)} = \frac{4}{10} = \frac{2}{5}.$$

$$\text{Mặt khác } P(\bar{C}) = 1 - P(C)$$

$$\Rightarrow P(C) = 1 - P(\bar{C}) = 1 - \frac{2}{5} = \frac{3}{5}.$$

Vậy xác suất của biến cố C là $\frac{3}{5}$.

17. Nguyên lí xác suất bé

- Nếu một biến cố ngẫu nhiên có xác suất rất bé thì thực tế có thể cho rằng trong một phép thử biến cố đó sẽ không xảy ra.

- Một xác suất như thế nào được xem là bé phải tùy thuộc vào từng bài toán cụ thể.

Ví dụ:

- Mỗi chuyến bay đều có một xác suất rất bé bị xảy ra tai nạn. Nhưng thực tế, tai nạn của một chuyến bay gần như sẽ không xảy ra.

- Xác suất để dù không mở là 0,01(dùng cho nhảy dù) thì không thể coi là bé và không thể dùng loại dù đó. Xác suất để tàu về ga chậm là 0,01 thì có thể xem là tàu về ga đúng giờ.

B. Bài tập tự luyện

B.1 Bài tập tự luận

Bài 1. Hãy viết số quy tròn của mỗi số gần đúng sau với độ chính xác d sau đó ước lượng sai số tương đối của số gần đúng đó.

a) $a = 32\,564$ với $d = 20$;

b) $a = 0,7612309$ với $d = 0,001$.

Hướng dẫn giải

a) Vì độ chính xác $d = 20$ thỏa mãn $10 < d = 20 < 100$ nên hàng thấp nhất mà d nhỏ hơn một đơn vị của hàng đó là hàng trăm.

Vì vậy, ta quy tròn số $a = 32\,564$ đến hàng trăm.

Quy tròn số $a = 32\,564$ đến hàng trăm ta được số quy tròn là 32 600.

Do đó sai số tương đối là: $\delta_a \leq \frac{d}{|a|} = \frac{20}{32\,564} \approx 0,06\%$.

Suy ra, sai số tương đối không vượt quá 0,06%.

Vậy số quy tròn của số gần đúng $a = 32\,564$ với độ chính xác $d = 20$ là 32 600 và sai số tương đối không vượt quá 0,06%.

b) Vì độ chính xác $d = 0,001$ thỏa mãn $d = 0,001 < 0,01$ nên hàng thấp nhất mà d nhỏ hơn một đơn vị của hàng đó là hàng phần trăm.

Vì vậy, ta quy tròn số 0,7612309 đến hàng phần trăm.

Quy tròn số $a = 0,7612309$ đến hàng phần trăm ta được số quy tròn là 0,76.

Do đó sai số tương đối là: $\delta_a \leq \frac{d}{|a|} = \frac{0,001}{0,7612309} \approx 0,13\%$.

Suy ra, sai số tương đối không vượt quá 0,13%.

Vậy số quy tròn của số gần đúng $a = 0,7612309$ với độ chính xác $d = 0,001$ là 0,76 và sai số tương đối không vượt quá 0,13%.

Bài 2. Một cửa hàng thống kê số lượng một loại giày bán được trong 1 tháng trong bảng sau:

Size giày	35	36	37	38	39
Số giày bán được	12	50	31	32	22

- a) Tìm một của mẫu số liệu trên?
- b) Cửa hàng nên nhập nhiều hơn size giày nào để bán?

Hướng dẫn giải

a) Ta thấy số giày bán được của size giày 36 là lớn nhất bằng 50 nên một của mẫu số liệu là $M_o = 36$.

Vậy một của mẫu số liệu là $M_o = 36$.

b) Vì $M_o = 36$ nên số lượng bán được của size 36 là lớn nhất.

Do đó, cửa hàng nên nhập nhiều số lượng giày size 36.

Vậy cửa hàng nên nhập nhiều số lượng giày size 36.

Bài 3. Điểm số bài kiểm tra Toán của các bạn trong 2 tổ của lớp 10A được ghi lại trong bảng sau:

Tổ 1	8	7	8	8	7	9
Tổ 2	7	8	9	7	6	9

Tổ nào có kết quả tốt hơn? Tại sao?

Hướng dẫn giải

Ta có số bạn trong mỗi tổ 1 và 2 là 6 (người).

Điểm trung bình của tổ 1 là : $\overline{x_1} = \frac{8+7+8+8+7+9}{6} \approx 7,83$

Điểm trung bình của tổ 2 là : $\overline{x_2} = \frac{7+8+9+7+6+9}{6} \approx 7,67$

Vì $\overline{x_2} < \overline{x_1}$ ($7,67 < 7,83$) nên điểm trung bình của tổ 2 nhỏ hơn của tổ 1.

Vậy tổ 1 có kết quả bài kiểm tra tốt hơn tổ 2.

Bài 4. Cân nặng của 12 học sinh (đơn vị: kg) của một lớp 5 được cho trong bảng sau:

25	34	36	41	19	35
27	57	36	37	29	30

a) Xác định trung vị, tứ phân vị của mẫu số liệu trên.

b) Từ kết quả của câu a), bước đầu xác định những số liệu bất thường trong mẫu số liệu trên.

Hướng dẫn giải

a) Ta sắp xếp các số liệu theo thứ tự tăng dần như sau:

19	25	27	29	30	34
35	36	36	37	41	51

Mẫu trên có 12 số liệu, là số chẵn

Trung vị của mẫu số liệu trên là $Q_2 = \frac{34+35}{2} = 34,5$.

Trung vị nửa phía dưới 19, 25, 27, 29, 30, 34 là $Q_1 = \frac{27+29}{2} = 28$.

Trung vị nửa phía trên 35, 36, 36, 37, 41, 51 là $Q_3 = \frac{36+37}{2} = 36,5$.

Vậy tứ phân vị của mẫu là $Q_1 = 28$; $Q_2 = 34,5$; $Q_3 = 36,5$.

b) Dựa vào trung vị và tứ phân vị của mẫu số liệu, bước đầu ta thấy số liệu bất thường trong mẫu là 51.

Bài 5. Cho mẫu số liệu thống kê chiều cao của 13 cây xoan đào (đơn vị: mét) như sau:

2,4 3 2,5 4 0,9 5 4,3 4 3,7 3,9 4,4 7,3 1,7

a) Tìm khoảng biến thiên của mẫu số liệu trên.

b) Tìm khoảng tứ phân vị của mẫu số liệu.

c) Tìm các giá trị bất thường của mẫu.

Hướng dẫn giải

a) Sắp xếp mẫu theo thứ tự tăng dần:

0,9 1,7 2,4 2,5 3 3,7 3,9 4 4 4,3 4,4 5 7,3

Mẫu có 13 số liệu, số liệu nhỏ nhất là 0,9; số liệu lớn nhất là 8.

Khi đó khoảng biến thiên của mẫu là $R = 8 - 0,9 = 7,1$ (m).

Vậy khoảng biến thiên của mẫu là 7,1 mét.

b) Trung vị của mẫu là 3,9; tức là $Q_2 = 3,9$.

Trung vị nửa dưới của mẫu 0,9; 1,7; 2,4; 2,5; 3; 3,7 là $\frac{2,4+2,5}{2} = 2,45$. Suy ra $Q_1 = 2,45$ (m).

Trung vị nửa trên của mẫu 4; 4; 4,3; 4,4; 5; 8 là $\frac{4,3+4,4}{2} = 4,35$. Suy ra $Q_3 = 4,35$ (m).

Khi đó khoảng tứ phân vị của mẫu là: $\Delta_Q = Q_3 - Q_1 = 4,35 - 2,45 = 1,9$ (m).

Vậy khoảng tứ phân vị là 1,9 m.

c) Ta có: $Q_1 - \frac{3}{2}\Delta_Q = 2,45 - \frac{3}{2} \cdot 1,9 = -0,4$; $Q_3 + \frac{3}{2}\Delta_Q = 4,35 + \frac{3}{2} \cdot 1,9 = 7,2$.

Ta thấy $7,3 > 7,2$ nên giá trị 7,3 là các giá trị bất thường của mẫu.

Vậy mẫu có giá trị bất thường là 7,3.

Bài 6. Thời gian 5 lần chạy cự li 100 m (đơn vị: giây) của hai bạn Long và Nam được ghi lại trong bảng sau:

Long	20	23	19	21	22
Nam	18	21	20	22	22

a) Tính thời gian chạy trung bình của mỗi bạn Long và Nam. Bạn nào có kết quả chạy tốt hơn.

b) Tính phương sai, độ lệch chuẩn của mẫu số liệu thống kê kết quả 5 lần chạy của mỗi bạn. Từ đó cho biết bạn nào có kết quả chạy ổn định hơn.

Hướng dẫn giải

a) Thời gian chạy trung bình của Long là: $\overline{x_L} = \frac{20 + 23 + 19 + 21 + 22}{5} = 21$ (s).

Thời gian chạy trung bình của Nam là: $\overline{x_N} = \frac{18 + 21 + 20 + 22 + 22}{5} = 20,6$ (s).

Do $\overline{x_L} > \overline{x_N}$ ($21 > 20,6$) nên kết quả chạy của Nam tốt hơn của Long.

Vậy thời gian chạy trung bình của Long là 21 giây, của Nam là 20,6 giây và kết quả chạy của Nam tốt hơn của Long.

b) Ta có phương sai của mẫu số liệu thống kê kết quả 5 lần chạy của Long là:

$$s_L^2 = \frac{(20 - 21)^2 + (23 - 21)^2 + (19 - 21)^2 + (21 - 21)^2 + (22 - 21)^2}{5} = 2$$

$$\Rightarrow s_L = \sqrt{s_L^2} = \sqrt{2} \approx 1,41.$$

Phương sai của mẫu số liệu thống kê kết quả 5 lần chạy của Nam là:

$$s_N^2 = \frac{(18-20,6)^2 + (21-20,6)^2 + (20-20,6)^2 + (22-20,6)^2 + (22-20,6)^2}{5} = 2,24.$$

$$\Rightarrow s_N = \sqrt{s_N^2} = \sqrt{2,24} \approx 1,5.$$

Ta thấy $s_L^2 < s_N^2$ ($2 < 2,24$) nên bạn Long có kết quả chạy ổn định hơn bạn Nam.

Vậy phương sai của mẫu số liệu thống kê kết quả 5 lần chạy của Long và Nam lần lượt là 2 và 2,24. Độ lệch chuẩn của mẫu số liệu thống kê kết quả 5 lần chạy của Long và Nam lần lượt là 1,41 và 1,5. Bạn Long có kết quả chạy ổn định hơn bạn Nam.

Bài 7. Tung một đồng xu hai lần liên tiếp.

- Viết tập hợp Ω là không gian mẫu của trò chơi trên.
- Xét biến cố A: “Có ít nhất một lần xuất hiện mặt sấp”. Tính xác suất của biến cố A.

Hướng dẫn giải

- Không gian mẫu của trò chơi trên là: $\Omega = \{SS; SN; NS; NN\}$.

Vậy không gian mẫu là $\Omega = \{SS; SN; NS; NN\}$.

- Từ $\Omega = \{SS; SN; NS; NN\} \Rightarrow n(\Omega) = 4$.

Có ba kết quả thuận lợi cho biến cố A là SS; SN, NS, tức là $A = \{SS; SN; NS\} \Rightarrow n(A) = 3$.

$$\Rightarrow \text{Xác suất của biến cố A là } P(A) = \frac{n(A)}{n(\Omega)} = \frac{3}{4}.$$

Vậy xác suất của biến cố A bằng $\frac{3}{4}$.

Bài 8. Một tổ trong lớp 10A có 5 bạn nữ và 4 bạn nam. Giáo viên chọn ngẫu nhiên hai bạn trong tổ đó tham gia đội làm báo tường của lớp. Tính xác suất để hai bạn được chọn có một bạn nam và một bạn nữ.

Hướng dẫn giải

Vì tổ có 5 bạn nữ và 4 bạn nam nên tổ đó có $5 + 4 = 9$ (học sinh).

Chọn 2 trong 9 bạn học sinh của tổ đó, ta có $C_9^2 = 36$ (cách chọn).

Gọi A là biến cố “hai bạn được chọn có một bạn nam và một bạn nữ”.

+ Để chọn được 1 bạn nữ trong 5 bạn nữ, ta có $C_5^1 = 5$ (cách chọn).

+ Để chọn được 1 bạn nam trong 4 bạn nam, ta có $C_4^1 = 4$ (cách chọn).

Áp dụng quy tắc nhân ta có $5.4 = 20$ cách chọn 1 bạn nữ và 1 bạn nam.

Suy ra $n(A) = 20$.

$$\text{Khi đó } P(A) = \frac{n(A)}{n(\Omega)} = \frac{20}{36} = \frac{5}{9}.$$

Vậy xác suất để hai bạn được chọn có một bạn nam và một bạn nữ là $\frac{5}{9}$.

Bài 9. Chọn ngẫu nhiên 4 viên bi từ một túi đựng 5 viên bi đỏ và 6 viên bi xanh, các viên bi có kích thước và khối lượng giống nhau. Gọi A là biến cố: “Trong bốn viên bi đó có cả bi đỏ và cả bi xanh”. Tính $P(A)$ và $P(\bar{A})$.

Hướng dẫn giải

Có 5 viên bi đỏ và 6 viên bi xanh thì có tất cả $5 + 6 = 11$ viên bi.

Chọn 4 viên bi từ 11 viên bi, thì số cách là: $C_{11}^4 = 330$ (cách).

$$\Rightarrow n(\Omega) = 330.$$

Xét biến cố A: “Trong bốn viên bi đó có cả bi đỏ và cả bi xanh”.

Khi đó, biến cố \bar{A} : “trong bốn viên bi chỉ có bi đỏ hoặc chỉ có bi xanh”.

- Nếu 4 viên bi đều là bi đỏ, thì ta chọn 4 trong 5 viên bi đỏ, ta có $C_5^4 = 5$ (cách).

- Nếu 4 viên bi đều là bi xanh, thì ta chọn 4 trong 6 viên bi xanh, ta có $C_6^4 = 15$ (cách).

Khi đó, ta có $5 + 15 = 20$ cách chọn bốn viên bi chỉ có bi đỏ hoặc chỉ có bi xanh.

$$\Rightarrow n(\bar{A}) = 20.$$

$$\Rightarrow P(\bar{A}) = \frac{n(\bar{A})}{n(\Omega)} = \frac{20}{330} = \frac{2}{33}.$$

$$\text{Mặt khác } P(\bar{A}) = 1 - P(A) \Rightarrow P(A) = 1 - P(\bar{A}) = 1 - \frac{2}{33} = \frac{31}{33}.$$

$$\text{Vậy } P(A) = \frac{31}{33} \text{ và } P(\bar{A}) = \frac{2}{33}.$$

B.2 Bài tập trắc nghiệm

Câu 1. Hãy viết số quy tròn của số gần đúng $a = 17\,658$ biết $\bar{a} = 17\,658 \pm 16$.

A. 17 700;

B. 17 800;

C. 17 500;

D. 17 600.

Hướng dẫn giải

Đáp án đúng là: A

Ta có: $\bar{a} = 17\,658 \pm 16$, suy ra $d = 16$ (hàng chục)

Do đó, ta làm tròn số $a = 17658$ đến hàng trăm, kết quả là: 17 700.

Câu 2. Số áo bán được trong một quý ở cửa hàng bán áo sơ mi được thống kê như sau:

Cỡ áo	36	37	38	39	40	41	42
Tần số (số áo bán được)	13	45	126	125	110	40	12

Giá trị một của bảng phân bố tần số trên bằng

- A. 38;
- B. 126;
- C. 39;
- D. 12.

Hướng dẫn giải

Đáp án đúng là: A

Theo bảng tần số trên thì cỡ áo 38 bán được với số lượng nhiều nhất trong một quý . Nên một của mẫu số liệu là 38.

Câu 3. Kết quả kiểm tra 15 phút môn toán của 100 học sinh được trình bày ở bảng sau:

Điểm	3	4	5	6	7	8	9	10
Số học sinh	3	5	11	17	30	19	10	5

Điểm trung bình môn Toán của các học sinh nói trên là:

- A. 6,88;
- B. 7,12;
- C. 6,5;
- D. 7,22.

Hướng dẫn giải

Đáp án đúng là: A

$$\bar{x} = \frac{3.3 + 4.5 + 5.11 + 6.17 + 7.30 + 8.19 + 9.10 + 10.5}{100} = 6,88.$$

Câu 4. Tìm tứ phân vị của mẫu số liệu sau:

200 240 220 210 225 235 225 270 250 280.

A. $Q_1 = 210$; $Q_2 = 230$; $Q_3 = 250$;

B. $Q_1 = 220$; $Q_2 = 230$; $Q_3 = 250$;

C. $Q_1 = 220$; $Q_2 = 230$; $Q_3 = 280$;

D. $Q_1 = 200$; $Q_2 = 250$; $Q_3 = 280$.

Hướng dẫn giải

Đáp án đúng là: B

Mẫu số liệu trên được sắp xếp theo thứ tự tăng dần như sau:

200 210 220 225 225 235 240 250 270 280

Trung vị của mẫu số liệu trên là: $\frac{225 + 235}{2} = 230 \Rightarrow Q_2 = 230$

Nửa dãy phía dưới số 230 (nghĩa là những số nhỏ hơn 230) gồm: 200 210 220 225 225 có trung vị là 220 $\Rightarrow Q_1 = 220$.

Nửa dãy phía trên số 230 (nghĩa là những số lớn hơn 230) gồm: 235 240 250 270 280 có trung vị là 250 $\Rightarrow Q_3 = 250$.

Vậy tứ phân vị của mẫu số liệu: $Q_1 = 220$; $Q_2 = 230$; $Q_3 = 250$.

Câu 5. Điều tra về số học sinh của một trường THPT như sau:

Khối lớp	10	11	12
Số học sinh	1 120	1 075	900

Khoảng biến thiên của mẫu số liệu trên là.

- A. 220;
- B. 45;
- C. 175;
- D. 3 095.

Hướng dẫn giải

Đáp án đúng là: A

Giá trị lớn nhất của mẫu số liệu là 1 120 và giá trị nhỏ nhất của mẫu số liệu là 900.

Vậy khoảng biến thiên: $R = 1\,120 - 900 = 220$.

Câu 6. Cho mẫu số liệu: 10 7 8 5 4. Tính độ lệch chuẩn của mẫu số liệu đó.

- A. 6,8;
- B. 2,14;
- C. 4,56;
- D. 20,79.

Hướng dẫn giải

Đáp án đúng là: B

$$\text{Ta có: } \bar{x} = \frac{10 + 7 + 8 + 5 + 4}{5} = 6,8$$

$$s^2 = \frac{(10 - 6,8)^2 + (7 - 6,8)^2 + (8 - 6,8)^2 + (5 - 6,8)^2 + (4 - 6,8)^2}{5} = 4,56$$

$$\Rightarrow s = \sqrt{4,56} \approx 2,14.$$

Câu 7. Gieo xúc xắc hai lần. Tính xác suất để tổng hai số chấm xuất hiện trên hai

con xúc xắc là một số nguyên tố.

A. $\frac{1}{3}$;

B. $\frac{13}{36}$;

C. $\frac{17}{36}$;

D. $\frac{1}{6}$.

Hướng dẫn giải

Đáp án đúng là: C

Ta có: $n(\Omega) = 6 \cdot 6 = 36$

Gọi M là biến cố tổng hai số chấm xuất hiện trên hai con xúc xắc là một số nguyên tố.

$$\Rightarrow M = \{(1; 1); (1; 2); (2; 1); (1; 4); (4; 1); (1; 6); (6; 1); (2; 3); (3; 2); (2; 5); (5; 2); (3; 4); (4; 3); (4; 5); (5; 4); (5; 6); (6; 5)\} \Rightarrow n(M) = 17$$

Vậy xác suất của biến cố M là : $\frac{n(M)}{n(\Omega)} = \frac{17}{36}$.

Câu 8. Gieo một đồng xu cân đối 3 lần liên tiếp. Gọi H là biến cố có hai lần xuất hiện mặt sấp và một lần xuất hiện mặt ngửa. Xác suất biến cố H là:

A. $\frac{3}{8}$;

B. $\frac{1}{8}$;

C. $\frac{5}{8}$;

D. $\frac{1}{6}$.

Hướng dẫn giải

Đáp án đúng là: A

Ta có: $n(\Omega) = 2 \cdot 2 \cdot 2 = 8$.

Gọi H là biến cố có hai lần xuất hiện mặt sấp và một lần xuất hiện mặt ngửa.

Ta có: $H = \{SSN; SNS; NSS\} \Rightarrow n(H) = 3$

Vậy xác suất của biến cố H là : $\frac{n(H)}{n(\Omega)} = \frac{3}{8}$.

Câu 9. Một hộp có 5 viên bi đỏ và 9 viên bi xanh. Chọn ngẫu nhiên 2 viên bi. Xác suất để chọn được hai viên bi khác màu là:

A. $\frac{14}{45}$;

B. $\frac{45}{91}$;

C. $\frac{46}{91}$;

D. $\frac{15}{22}$.

Hướng dẫn giải

Đáp án đúng là: B

Mỗi lần chọn 2 viên bi ngẫu nhiên từ $5 + 9 = 14$ viên bi cho ta một tổ hợp chập 2 của 14 nên $n(\Omega) = C_{14}^2 = 91$.

Gọi A là biến cố: “Hai viên bi được chọn khác màu”.

Việc chọn 2 viên bi từ hộp sao cho hai viên bi được chọn khác màu có thể xem là một công việc gồm 2 công đoạn:

+ Công đoạn 1: Chọn 1 viên bi màu đỏ có 5 cách

+ Công đoạn 2: Chọn 1 viên bi màu xanh có 9 cách

$$\Rightarrow n(A) = 5 \cdot 9 = 45.$$

$$\text{Vậy } P(A) = \frac{n(A)}{n(\Omega)} = \frac{45}{91}.$$

Câu 10. Cho tập hợp $A = \{2; 3; 4; 5; 6; 7; 8\}$. Gọi S là tập hợp các số tự nhiên có 4 chữ số đôi một khác nhau được lập thành từ các chữ số của tập A. Chọn ngẫu nhiên một số từ S, tính xác suất để số được chọn mà trong mỗi số luôn có mặt hai chữ số chẵn và hai chữ số lẻ.

A. $\frac{1}{5}$;

B. $\frac{3}{35}$;

C. $\frac{17}{35}$;

D. $\frac{18}{35}$.

Hướng dẫn giải

Đáp án đúng là: D

Ta có : $n(\Omega) = A_7^4 = 840$.

Gọi E là biến cố: “Số được chọn luôn có mặt hai chữ số chẵn và hai chữ số lẻ”.

Việc số được chọn luôn có mặt hai chữ số chẵn và hai chữ số lẻ là một công việc gồm 3 công đoạn:

+ Công đoạn 1: Chọn 2 chữ số chẵn từ các chữ số 2; 4; 6; 8 có $C_4^2 = 6$ cách.

+ Công đoạn 2: Chọn 2 chữ số lẻ từ các chữ số 3; 5; 7 có $C_3^2 = 3$ cách.

+ Công đoạn 3: Từ 4 chữ số được chọn ta lập số có 4 chữ số khác nhau, mỗi cách lập một số tương ứng với một hoán vị của 4, do đó ta có $4! = 24$ cách.

$$\Rightarrow n(E) = 6 \cdot 3 \cdot 24 = 432.$$

$$\Rightarrow P(E) = \frac{n(E)}{n(\Omega)} = \frac{432}{840} = \frac{18}{35}.$$