Elemente de statistică

CURS 10 - 30.04.2014

Facultatea de Automatică și Calculatoare

Statistică și prelucrarea datelor

Eșantioane aleatoare

- **Populația statistică** este o mulțime de elemente supusă cercetării statistice. Un element al acestei populații se numește **unitate statistică**.

Eșantioane aleatoare

- **Populația statistică** este o mulțime de elemente supusă cercetării statistice. Un element al acestei populații se numește **unitate statistică**.
- O **caracteristică** a unei populații statistice este **cantitativă** *dacă poate fi cuantificată printr-un număr.*

Eşantioane aleatoare

- **Populaţia statistică** este o mulţime de elemente supusă cercetării statistice. Un element al acestei populaţii se numeşte **unitate statistică**.
- O **caracteristică** a unei populaţii statistice este **cantitativă** *dacă poate fi cuantificată printr-un număr.*
- O **caracteristică calitativă** a unei populaţii statistice *nu poate fi cuantificată printr-un număr* ci prin aprecieri de tipul bun, f. bun, mult, puţin.

Eșantioane aleatoare

- **Populația statistică** este o mulțime de elemente supusă cercetării statistice. Un element al acestei populații se numește **unitate statistică**.
- O **caracteristică** a unei populații statistice este **cantitativă** *dacă poate fi cuantificată printr-un număr.*
- O **caracteristică calitativă** a unei populații statistice *nu poate fi cuantificată printr-un număr* ci prin aprecieri de tipul bun, f. bun, mult, puțin.
- Valoarea numerică a unei caracteristici cantitative se numește **variabilă statistică**.

Eșantioane aleatoare

- **Populația statistică** este o mulțime de elemente supusă cercetării statistice. Un element al acestei populații se numește **unitate statistică**.
- O **caracteristică** a unei populații statistice este **cantitativă** *dacă poate fi cuantificată printr-un număr.*
- O **caracteristică calitativă** a unei populații statistice *nu poate fi cuantificată printr-un număr* ci prin aprecieri de tipul bun, f. bun, mult, puțin.
- Valoarea numerică a unei caracteristici cantitative se numește **variabilă statistică**.
- O submulțime a unei populații statistice se numește **selecție** sau **eșantion**. De exemplu, dacă se solicită opinia unei populații într-o problemă oarecare este dificil să fie consultată întreaga populație și se recurge la extragerea unui eșantion.

Eșantioane aleatoare

- Un eșantion ale cărui unități au fost alese la întâmplare se numește **eșantion reprezentativ (aleator)**. Extragerea unui eșantion reprezentativ se realizează astfel încât elementele să aibă șanse egale de a fi extrase. În multe situații se folosesc numere aleatoare. De exemplu dacă avem o listă a populației inițiale $x_1, x_2, ..., x_N$ se aleg numere aleatoare $k_1, k_2, ..., k_n$ între 1 și N și un eșantion aleator este $x_{k_1}, x_{k_2}, ..., x_{k_n}$.

Eșantioane aleatoare

- Un eșantion ale cărui unități au fost alese la întâmplare se numește **eșantion reprezentativ (aleator)**. Extragerea unui eșantion reprezentativ se realizează astfel încât elementele să aibă șanse egale de a fi extrase. În multe situații se folosesc numere aleatoare. De exemplu dacă avem o listă a populației inițiale $x_1, x_2, ..., x_N$ se aleg numere aleatoare $k_1, k_2, ..., k_n$ între 1 și N și un eșantion aleator este $x_{k_1}, x_{k_2}, ..., x_{k_n}$.
- Variabilele statistice pot fi **discrete** (numărul de becuri care se ard după 1000 de ore de întrebuințare) sau **continue** (timpii de defectare a unui număr fixat de n becuri).

Eșantioane aleatoare

- Un eșantion ale cărui unități au fost alese la întâmplare se numește **eșantion reprezentativ (aleator)**. Extragerea unui eșantion reprezentativ se realizează astfel încât elementele să aibă șanse egale de a fi extrase. În multe situații se folosesc numere aleatoare. De exemplu dacă avem o listă a populației inițiale $x_1, x_2, ..., x_N$ se aleg numere aleatoare $k_1, k_2, ..., k_n$ între 1 și N și un eșantion aleator este $x_{k_1}, x_{k_2}, ..., x_{k_n}$.
- Variabilele statistice pot fi **discrete** (numărul de becuri care se ard după 1000 de ore de întrebuințare) sau **continue** (timpii de defectare a unui număr fixat de n becuri).
- **În cazul variabilelor discrete** rezultatele sunt prezentate sub forma unui tablou de forma următoare:

X	x_1	x_2	x_3	x_4	$\cdots$	x_k
n	n_1	n_2	n_3	n_4	$\cdots$	n_k

(1)

cu $n = \sum_{j=1}^{k} n_j$, n_j se numește **frecvența absolută** și reprezintă numărul de apariții a valorii x_j a variabilei statistice X,

Eșantioane aleatoare

- Un eșantion ale cărui unități au fost alese la întâmplare se numește **eșantion reprezentativ (aleator)**. Extragerea unui eșantion reprezentativ se realizează astfel încât elementele să aibă șanse egale de a fi extrase. În multe situații se folosesc numere aleatoare. De exemplu dacă avem o listă a populației inițiale $x_1, x_2, ..., x_N$ se aleg numere aleatoare $k_1, k_2, ..., k_n$ între 1 și N și un eșantion aleator este $x_{k_1}, x_{k_2}, ..., x_{k_n}$.
- Variabilele statistice pot fi **discrete** (numărul de becuri care se ard după 1000 de ore de întrebuințare) sau **continue** (timpii de defectare a unui număr fixat de n becuri).
- **În cazul variabilelor discrete** rezultatele sunt prezentate sub forma unui tablou de forma următoare:

X	x_1	x_2	x_3	x_4	$\cdots$	x_k
n	n_1	n_2	n_3	n_4	$\cdots$	n_k

(1)

cu $n = \sum_{j=1}^{k} n_j$, n_j se numește **frecvența absolută** și reprezintă numărul de apariții a valorii x_j a variabilei statistice X, n se numește **volumul selecției**.

Eșantioane aleatoare

- **Frecvența relativă** este $\dfrac{n_j}{n}$.

Eșantioane aleatoare

- **Frecvența relativă** este $\frac{n_j}{n}$.
- **Frecvența absolută cumulată** este: $\sum_{j=1}^{r} n_j, r \leq k$.

Eșantioane aleatoare

- **Frecvența relativă** este $\dfrac{n_j}{n}$.
- **Frecvența absolută cumulată** este: $\sum\limits_{j=1}^{r} n_j, r \leq k$.
- **Frecvența relativă cumulată** este: $\dfrac{\sum\limits_{j=1}^{r} n_j}{n}, r \leq k$.

Eșantioane aleatoare

- **Frecvența relativă** este $\frac{n_j}{n}$.
- **Frecvența absolută cumulată** este: $\sum_{j=1}^{r} n_j, r \leq k$.
- **Frecvența relativă cumulată** este: $\frac{\sum_{j=1}^{r} n_j}{n}, r \leq k$.
- O reprezentare a variabilei statistice este sub forma unei diagrame în batoane. În acest tip de prezentare, pe axa orizontală se așează valorile luate de v.a. iar pe verticală se așează frecvențele absolute corespunzătoare.

Eșantioane aleatoare

- **Frecvența relativă** este $\frac{n_j}{n}$.
- **Frecvența absolută cumulată** este: $\sum_{j=1}^{r} n_j, r \leq k$.
- **Frecvența relativă cumulată** este: $\frac{\sum_{j=1}^{r} n_j}{n}, r \leq k$.
- O reprezentare a variabilei statistice este sub forma unei diagrame în batoane. În acest tip de prezentare, pe axa orizontală se așează valorile luate de v.a. iar pe verticală se așează frecvențele absolute corespunzătoare.

Example 1

Punctele obţinute de studenţi care au promovat examenul de matematică şi care cuantifică cunoştinţele lor sunt: 64, 62, 76, 82, 66, 76, 72, 71, 74, 72, 71, 73, 70, 75, 77, 84, 92, 86, 62, 58, 78, 80, 79, 84, 83, 82, 66, 68, 68, 82, 84, 78, 76, 69, 77, 58, 62, 82, 85, 58, 78, 84, 94, 88, 77, 78, 88, 91, 70, 71, 78, 58, 65, 53, 60, 49, 68, 74, 71, 66, 68, 71, 73, 70, 85, 78, 65, 54, 51, 78, 89, 66, 68, 95, 94, 99, 81, 81, 92, 88, 99, 81, 81.

Eşantioane aleatoare

- **În cazul variabilelor continue**, datele se grupează într-un tablou de forma:

X	$[t_1, t_2)$	$[t_2, t_3)$	$\cdots$	$[t_{m-1}, t_m)$
n	n_1	n_2	$\cdots$	n_{m-1}

.

Eșantioane aleatoare

- **În cazul variabilelor continue**, datele se grupează într-un tablou de forma:

X	$[t_1, t_2)$	$[t_2, t_3)$	$\cdots$	$[t_{m-1}, t_m)$
n	n_1	n_2	$\cdots$	n_{m-1}

.

- În acest caz X este variabila statistică iar n_j este frecvența absolută corespunzătoare intervalului $[t_j, t_{j+1})$. La fel $n = \sum_{j=1}^{m-1} n_j$ este volumul selecției.

Eșantioane aleatoare

- **În cazul variabilelor continue**, datele se grupează într-un tablou de forma:

X	$[t_1, t_2)$	$[t_2, t_3)$	$\cdots$	$[t_{m-1}, t_m)$
n	n_1	n_2	$\cdots$	n_{m-1}

.

- În acest caz X este variabila statistică iar n_j este frecvența absolută corespunzătoare intervalului $[t_j, t_{j+1})$. La fel $n = \sum_{j=1}^{m-1} n_j$ este volumul selecției.
- **Frecvența relativă** este $\frac{n_j}{n}$ corespunzătoare intervalului $[t_j, t_{j+1})$.

Eșantioane aleatoare

- **În cazul variabilelor continue**, datele se grupează într-un tablou de forma:

X	$[t_1, t_2)$	$[t_2, t_3)$	$\cdots$	$[t_{m-1}, t_m)$
n	n_1	n_2	$\cdots$	n_{m-1}

.

- În acest caz X este variabila statistică iar n_j este frecvența absolută corespunzătoare intervalului $[t_j, t_{j+1})$. La fel $n = \sum_{j=1}^{m-1} n_j$ este volumul selecției.
- **Frecvența relativă** este $\frac{n_j}{n}$ corespunzătoare intervalului $[t_j, t_{j+1})$.
- **Frecvența absolută cumulată** este: $\sum_{j=1}^{k} n_j,\ k \leq m-1$.

Eșantioane aleatoare

- **În cazul variabilelor continue**, datele se grupează într-un tablou de forma:

X	$[t_1, t_2)$	$[t_2, t_3)$	$\cdots$	$[t_{m-1}, t_m)$
n	n_1	n_2	$\cdots$	n_{m-1}

.

- În acest caz X este variabila statistică iar n_j este frecvența absolută corespunzătoare intervalului $[t_j, t_{j+1})$. La fel $n = \sum_{j=1}^{m-1} n_j$ este volumul selecției.
- **Frecvența relativă** este $\frac{n_j}{n}$ corespunzătoare intervalului $[t_j, t_{j+1})$.
- **Frecvența absolută cumulată** este: $\sum_{j=1}^{k} n_j, k \leq m-1$.
- **Frecvența relativă cumulată** $\frac{\sum_{j=1}^{k} n_j}{n}, k \leq m-1$.

Gruparea pe clase

- Dacă avem un șir $x_1, x_2, ..., x_n$ de n date numerice obținute în urma măsurării unor mărimi fizice, tehnice, economice de care depinde evoluția unor procese sau fenomene, notăm cu $m = \min_i x_i$ și $M = \max_i x_i$.

Gruparea pe clase

- Dacă avem un șir $x_1, x_2, ..., x_n$ de n date numerice obținute în urma măsurării unor mărimi fizice, tehnice, economice de care depinde evoluția unor procese sau fenomene, notăm cu $m = \min_i x_i$ și $M = \max_i x_i$.
- Uneori este util să organizăm datele selecției în grupe sau subintervale obținute divizând intervalul $[m, M]$ într-un număr de subintervale congruente (de aceeași lungime).

Gruparea pe clase

- Dacă avem un şir $x_1, x_2, ..., x_n$ de n date numerice obţinute în urma măsurării unor mărimi fizice, tehnice, economice de care depinde evoluţia unor procese sau fenomene, notăm cu $m = \min_i x_i$ şi $M = \max_i x_i$.
- Uneori este util să organizăm datele selecţiei în grupe sau subintervale obţinute divizând intervalul $[m, M]$ într-un număr de subintervale congruente (de aceeaşi lungime).

Definition 2

Se numeşte **pasul de histogramă** numărul real pozitiv

$$h = \frac{M - m}{[1 + 3.322 \log_{10} n]}.$$

Se consideră intervalele

$$I_1 = [m, m + h), I_1 = [m + h, m + 2h), ..., I_r = [m + (r - 1)h, M];$$

numărul r are formula $r = [1 + 3.322 \log_{10} n]$.

Gruparea pe clase

- Formula: $h = \dfrac{M - m}{[1 + 3.322 \log_{10} n]}$ a lui **H. A. Sturges** (1926) dă aproximativ mărimea intervalului de grupare.

Gruparea pe clase

- Formula: $h = \dfrac{M - m}{[1 + 3.322 \log_{10} n]}$ a lui **H. A. Sturges** (1926) dă aproximativ mărimea intervalului de grupare.

Remark 1

Deoarece $\log_{10} x = \frac{\log_2 x}{\log_2 10}$ *și* $\log_2 10 = 3.3219 \approx 3.322$ *formula este echivalentă cu* $h = \dfrac{M - m}{[1 + \log_2 n]}$.

Gruparea pe clase

- Formula: $h = \dfrac{M - m}{[1 + 3.322 \log_{10} n]}$ a lui **H. A. Sturges** (1926) dă aproximativ mărimea intervalului de grupare.

Remark 1

Deoarece $\log_{10} x = \frac{\log_2 x}{\log_2 10}$ *şi* $\log_2 10 = 3.3219 \approx 3.322$ *formula este echivalentă cu* $h = \dfrac{M - m}{[1 + \log_2 n]}$.

- Notăm cu n_1 numărul acelor date dintre x_i care sunt situate în intervalul I_1, n_2 numărul acelor date dintre x_i care sunt situate în intervalul I_2, ..., n_r numărul acelor date dintre x_i care sunt situate în intervalul I_r.

Gruparea pe clase

- Formula: $h = \dfrac{M - m}{[1 + 3.322 \log_{10} n]}$ a lui **H. A. Sturges** (1926) dă aproximativ mărimea intervalului de grupare.

Remark 1

Deoarece $\log_{10} x = \frac{\log_2 x}{\log_2 10}$ *și* $\log_2 10 = 3.3219 \approx 3.322$ *formula este echivalentă cu* $h = \dfrac{M - m}{[1 + \log_2 n]}$.

- Notăm cu n_1 numărul acelor date dintre x_i care sunt situate în intervalul I_1, n_2 numărul acelor date dintre x_i care sunt situate în intervalul $I_2, ..., n_r$ numărul acelor date dintre x_i care sunt situate în intervalul I_r.
- **Histograma** asociată datelor $x_1, x_2, ..., x_m$ este reuniunea dreptunghiurilor $D_k = I_k \times [0, n_k]$ (produs cartezian), pentru $1 \leq k \leq r$.

Gruparea pe clase

- Formula: $h = \dfrac{M - m}{[1 + 3.322 \log_{10} n]}$ a lui **H. A. Sturges** (1926) dă aproximativ mărimea intervalului de grupare.

Remark 1

Deoarece $\log_{10} x = \frac{\log_2 x}{\log_2 10}$ *și* $\log_2 10 = 3.3219 \approx 3.322$ *formula este echivalentă cu* $h = \dfrac{M - m}{[1 + \log_2 n]}$.

- Notăm cu n_1 numărul acelor date dintre x_i care sunt situate în intervalul I_1, n_2 numărul acelor date dintre x_i care sunt situate în intervalul $I_2, \ldots, n_r$ numărul acelor date dintre x_i care sunt situate în intervalul I_r.
- **Histograma** asociată datelor $x_1, x_2, \ldots, x_m$ este reuniunea dreptunghiurilor $D_k = I_k \times [0, n_k]$ (produs cartezian), pentru $1 \leq k \leq r$.
- **Moda** selecției este subgrupa (intervalul I_k) ce corespunde dreptunghiului cel mai înalt.

Gruparea pe clase

- Formula: $h = \dfrac{M - m}{[1 + 3.322 \log_{10} n]}$ a lui **H. A. Sturges** (1926) dă aproximativ mărimea intervalului de grupare.

Remark 1

Deoarece $\log_{10} x = \frac{\log_2 x}{\log_2 10}$ *și* $\log_2 10 = 3.3219 \approx 3.322$ *formula este echivalentă cu* $h = \dfrac{M - m}{[1 + \log_2 n]}$.

- Notăm cu n_1 numărul acelor date dintre x_i care sunt situate în intervalul I_1, n_2 numărul acelor date dintre x_i care sunt situate în intervalul $I_2, \ldots, n_r$ numărul acelor date dintre x_i care sunt situate în intervalul I_r.
- **Histograma** asociată datelor $x_1, x_2, \ldots, x_m$ este reuniunea dreptunghiurilor $D_k = I_k \times [0, n_k]$ (produs cartezian), pentru $1 \leq k \leq r$.
- **Moda** selecției este subgrupa (intervalul I_k) ce corespunde dreptunghiului cel mai înalt.

Algoritmul pentru construcția histogramelor

Date inițiale: $x_1, x_2, ..., x_n$

Algoritmul pentru construcția histogramelor

Date inițiale: $x_1, x_2, ..., x_n$

Pasul I. Se calculează $m, M, r = [1 + 3.322 \log_{10} n]\,, h = \dfrac{M - m}{r}$.

Algoritmul pentru construcția histogramelor

Date inițiale: $x_1, x_2, \dots, x_n$

Pasul I. Se calculează $m, M, r = [1 + 3.322 \log_{10} n]\,, h = \dfrac{M - m}{r}$.

Pasul II. Se determină intervalele I_k și numerele naturale $n_k, 1 \leq k \leq r$.

Algoritmul pentru construcția histogramelor

Date inițiale: $x_1, x_2, ..., x_n$

Pasul I. Se calculează $m, M, r = [1 + 3.322 \log_{10} n]\, , h = \dfrac{M - m}{r}$.

Pasul II. Se determină intervalele I_k și numerele naturale $n_k, 1 \leq k \leq r$.

Pasul III. Histograma este reuniunea tuturor dreptunghiurilor $D_k = I_k \times [0, n_k]\, , 1 \leq k \leq r$.

Algoritmul pentru construcția histogramelor

Date inițiale: $x_1, x_2, ..., x_n$

Pasul I. Se calculează $m, M, r = [1 + 3.322 \log_{10} n]\,, h = \dfrac{M-m}{r}$.

Pasul II. Se determină intervalele I_k și numerele naturale $n_k, 1 \leq k \leq r$.

Pasul III. Histograma este reuniunea tuturor dreptunghiurilor $D_k = I_k \times [0, n_k]\,, 1 \leq k \leq r$.

Example 3

Realizați histograma în cazul Exemplului 1.

Algoritmul pentru construcția histogramelor

Date inițiale: $x_1, x_2, ..., x_n$

Pasul I. Se calculează $m, M, r = [1 + 3.322 \log_{10} n], h = \dfrac{M-m}{r}$.
Pasul II. Se determină intervalele I_k și numerele naturale $n_k, 1 \leq k \leq r$.
Pasul III. Histograma este reuniunea tuturor dreptunghiurilor $D_k = I_k \times [0, n_k], 1 \leq k \leq r$.

Example 3

Realizați histograma în cazul Exemplului 1.

Example 4

Presupunem că într-un atelier lucrează 225 de muncitori și că s-a analizat într-o perioadă de timp modul în care își realizează sarcinile. S-a constatat că 10 dintre ei realizează 101%, etc conform tabelului de mai jos

10	15	40	23	17	18	36	45	8	5	8
101%	89%	85%	106%	127%	95%	105%	112%	117%	96%	80%

Realizați histograma.

Caracteristici statistice ale datelor experimentale

- Dacă avem un şir $x_1, x_2, ..., x_n$ de n date numerice definim **media de selecţie** ca fiind

$$\overline{x} = \frac{1}{n}\left(x_1 + x_2 + ... + x_n\right),$$

adică media aritmetică a celor n valori.

Caracteristici statistice ale datelor experimentale

- Dacă avem un șir $x_1, x_2, ..., x_n$ de n date numerice definim **media de selecție** ca fiind
$$\overline{x} = \frac{1}{n}\left(x_1 + x_2 + ... + x_n\right),$$
adică media aritmetică a celor n valori.
- Dacă datele sunt organizate sub forma unui tabel (1) atunci **media selecției** este
$$\overline{x} = \frac{1}{\sum\limits_{i=1}^{k} n_i} \cdot \sum_{i=1}^{k} n_i x_i$$
unde n_i reprezintă frecvența absolută de apariție a valorii x_i a variabilei în selecția considerată.

Caracteristici statistice ale datelor experimentale

- Dacă avem un șir $x_1, x_2, \ldots, x_n$ de n date numerice definim **media de selecție** ca fiind
$$\overline{x} = \frac{1}{n}\left(x_1 + x_2 + \ldots + x_n\right),$$
adică media aritmetică a celor n valori.
- Dacă datele sunt organizate sub forma unui tabel (1) atunci **media selecției** este
$$\overline{x} = \frac{1}{\sum\limits_{i=1}^{k} n_i} \cdot \sum_{i=1}^{k} n_i x_i$$
unde n_i reprezintă frecvența absolută de apariție a valorii x_i a variabilei în selecția considerată.
- În cazul datelor grupate x_i se înlocuiește cu x_i^* care este valoarea din mijloc a intervalului $[x_i, x_{i+1})$ deci
$$\overline{x} = \frac{1}{\sum\limits_{i=1}^{r} n_i} \cdot \sum_{i=1}^{r} n_i x_i^*,$$
unde r este numărul intervalelor.

Caracteristici statistice ale datelor experimentale

- Dacă avem un şir $x_1, x_2, ..., x_n$ de n date numerice definim **media de selecţie** ca fiind
$$\overline{x} = \frac{1}{n}\left(x_1 + x_2 + ... + x_n\right),$$
adică media aritmetică a celor n valori.
- Dacă datele sunt organizate sub forma unui tabel (1) atunci **media selecţiei** este
$$\overline{x} = \frac{1}{\sum\limits_{i=1}^{k} n_i} \cdot \sum_{i=1}^{k} n_i x_i$$
unde n_i reprezintă frecvenţa absolută de apariţie a valorii x_i a variabilei în selecţia considerată.
- În cazul datelor grupate x_i se înlocuieşte cu x_i^* care este valoarea din mijloc a intervalului $[x_i, x_{i+1})$ deci
$$\overline{x} = \frac{1}{\sum\limits_{i=1}^{r} n_i} \cdot \sum_{i=1}^{r} n_i x_i^*,$$
unde r este numărul intervalelor.

Caracteristici statistice ale datelor experimentale

- **Mediana** împarte șirul ordonat de date în două părți egale. Dacă șirul are $2k+1$ unități, atunci mediana coincide cu unitatea de ordin $k+1$, dacă șirul are $2k$ unități, mediana este media aritmetică a unităților de ordin k și $k+1$.

Caracteristici statistice ale datelor experimentale

- **Mediana** împarte șirul ordonat de date în două părți egale. Dacă șirul are $2k+1$ unități, atunci mediana coincide cu unitatea de ordin $k+1$, dacă șirul are $2k$ unități, mediana este media aritmetică a unităților de ordin k și $k+1$.

Example 5

Pentru șirul de date: 2.5; 3.7; 1.4; 0.2; 5.4; 8.9; 4.2 șirul ordonat este 0.2; 1.4; 2.5; 3.7; 4.2; 5.4; 8.9. $m = 3$ și unitatea de ordin $m+1 = 4$ este 3.7 care este mediana.

Dacă $x_{me} = \overline{x}$ atunci repartiția este simetrică.

Caracteristici statistice ale datelor experimentale

- **Mediana** împarte șirul ordonat de date în două părți egale. Dacă șirul are $2k+1$ unități, atunci mediana coincide cu unitatea de ordin $k+1$, dacă șirul are $2k$ unități, mediana este media aritmetică a unităților de ordin k și $k+1$.

Example 5

Pentru șirul de date: 2.5; 3.7; 1.4; 0.2; 5.4; 8.9; 4.2 șirul ordonat este 0.2; 1.4; 2.5; 3.7; 4.2; 5.4; 8.9. $m=3$ și unitatea de ordin $m+1=4$ este 3.7 care este mediana.

Dacă $x_{me}=\overline{x}$ atunci repartiția este simetrică.

- **Modul.** Pentru datele negrupate este valoarea observată care are frecvența maximă. Pentru date grupate este o valoare din clasa cu cel mai mare număr de observații.

Caracteristici statistice ale datelor experimentale

- **Mediana** împarte şirul ordonat de date în două părţi egale. Dacă şirul are $2k+1$ unităţi, atunci mediana coincide cu unitatea de ordin $k+1$, dacă şirul are $2k$ unităţi, mediana este media aritmetică a unităţilor de ordin k şi $k+1$.

Example 5

Pentru şirul de date: 2.5; 3.7; 1.4; 0.2; 5.4; 8.9; 4.2 şirul ordonat este 0.2; 1.4; 2.5; 3.7; 4.2; 5.4; 8.9. $m=3$ şi unitatea de ordin $m+1=4$ este 3.7 care este mediana.

Dacă $x_{me}=\overline{x}$ atunci repartiţia este simetrică.

- **Modul.** Pentru datele negrupate este valoarea observată care are frecvenţa maximă. Pentru date grupate este o valoare din clasa cu cel mai mare număr de observaţii.

Theorem 1

Au loc relaţiile:

$$\overline{x+c}=\overline{x}+c, \overline{kx}=k\overline{x}, \overline{x+y}=\overline{x}+\overline{y}.$$

Caracteristici statistice ale datelor experimentale

- **Dispersia de selecție** este

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

Observăm că dacă s^2 este mic atunci datele diferă unele de altele foarte puțin, dacă s^2 este mare, datele diferă foarte mult.

Caracteristici statistice ale datelor experimentale

- **Dispersia de selecție** este

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

Observăm că dacă s^2 este mic atunci datele diferă unele de altele foarte puțin, dacă s^2 este mare, datele diferă foarte mult.

- $\overline{x}$ se mai numește **speranța matematică** a selecției, dispersia de selecție se mai numește **varianța** selecției.

Caracteristici statistice ale datelor experimentale

- **Dispersia de selecție** este

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

Observăm că dacă s^2 este mic atunci datele diferă unele de altele foarte puțin, dacă s^2 este mare, datele diferă foarte mult.

- $\overline{x}$ se mai numește **speranța matematică** a selecției, dispersia de selecție se mai numește **varianța** selecției.
- **Amplitudinea** unei serii statistice este diferența dintre cea mai mare și cea mai mică valoare a variabilei $\mathcal{A} = x_{\max} - x_{\min}$.

În cazul v. continue, amplitudinea este diferența dintre limita superioară a ultimului interval și limita inferioară a primului.

Caracteristici statistice ale datelor experimentale

- **Dispersia de selecție** este

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

Observăm că dacă s^2 este mic atunci datele diferă unele de altele foarte puțin, dacă s^2 este mare, datele diferă foarte mult.

- $\overline{x}$ se mai numește **speranța matematică** a selecției, dispersia de selecție se mai numește **varianța** selecției.
- **Amplitudinea** unei serii statistice este diferența dintre cea mai mare și cea mai mică valoare a variabilei $\mathcal{A} = x_{\max} - x_{\min}$.

În cazul v. continue, amplitudinea este diferența dintre limita superioară a ultimului interval și limita inferioară a primului.

- **Abaterea medie de selecție** este $s = \sqrt{s^2}$.

Caracteristici statistice ale datelor experimentale

- **Dispersia de selecţie** este

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

Observăm că dacă s^2 este mic atunci datele diferă unele de altele foarte puţin, dacă s^2 este mare, datele diferă foarte mult.

- $\overline{x}$ se mai numeşte **speranţa matematică** a selecţiei, dispersia de selecţie se mai numeşte **varianţa** selecţiei.
- **Amplitudinea** unei serii statistice este diferenţa dintre cea mai mare şi cea mai mică valoare a variabilei $\mathcal{A} = x_{\max} - x_{\min}$.

În cazul v. continue, amplitudinea este diferenţa dintre limita superioară a ultimului interval şi limita inferioară a primului.

- **Abaterea medie de selecţie** este $s = \sqrt{s^2}$.
- **Coeficientul de variaţie** este $v = \dfrac{\sqrt{s^2}}{\overline{x}}$. O selecţie de valori pozitive se consideră omogenă dacă $v < \frac{1}{3}$.

Caracteristici statistice ale datelor experimentale

- **Dispersia de selecție** este

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

Observăm că dacă s^2 este mic atunci datele diferă unele de altele foarte puțin, dacă s^2 este mare, datele diferă foarte mult.

- $\overline{x}$ se mai numește **speranța matematică** a selecției, dispersia de selecție se mai numește **varianța** selecției.
- **Amplitudinea** unei serii statistice este diferența dintre cea mai mare și cea mai mică valoare a variabilei $\mathcal{A} = x_{\max} - x_{\min}$.

În cazul v. continue, amplitudinea este diferența dintre limita superioară a ultimului interval și limita inferioară a primului.

- **Abaterea medie de selecție** este $s = \sqrt{s^2}$.
- **Coeficientul de variație** este $v = \dfrac{\sqrt{s^2}}{\overline{x}}$. O selecție de valori pozitive se consideră omogenă dacă $v < \frac{1}{3}$.

Caracteristici statistice ale datelor experimentale

Pentru exemplul 4 vrem să calculăm media şi dispersia. Calculele se organizează în felul următor:

x_i^*	n_i	$x_i^* n_i$	$\frac{x_i^* n_i}{n}$	$(x_i^* - \overline{x})^2$	$\frac{n_i}{n-1}(x_i^* - \overline{x})^2$
82.7	48	3969.6	17.643	309.76	66.377
88.1	15	1321.5	5.8733	148.84	9.9670
93.5	23	2150.5	9.5578	46.24	4.7479
98.9	10	989.0	4.3956	1.96	8.7500×10^{-2}
104.3	59	6153.7	27.350	16.0	4.2143
109.7	45	4936.5	21.94	88.36	17.751
115.1	8	920.8	4.0924	219.04	7.8229
120.5	0	0	0	408.04	0
125.1	17	2126.7	9.452	615.04	46.677
			$\overline{x} = 100.3$	1853.3	$s^2 = 157.64$

Caracteristici statistice ale datelor experimentale

Obţinem

$$\begin{aligned}\overline{x} &= \frac{\sum_{i=1}^{r} n_i x_i^*}{\sum_{i=1}^{r} n_i} \\ &= 17.643 + 5.873\,3 + 9.557\,8 + 4.95\,6 + 27.50 + 21.94 + 4.092\,4 + 9.452 \\ &= 100.3\end{aligned}$$

$$s^2 = 157.64 \Rightarrow s = \sqrt{157.64} = 12.555.$$