

APLICAȚII BAZELE STATISTICII – RECAPITULARE

1. Un eșantion de angajați ai unei firme a răspuns la câteva întrebări cu caracter personal și profesional. Precizați natura următoarelor variabile:

- + numărul de animale al familiei – var numerică discretă
- + locul nașterii – val calitativă, nominală
- + vârsta în ani împliniți - var numerică continuă
- + nivelul ultimei diplome de studii obținută (liceu, facultate, master, doctorat, post-doctorat) - variabilă ordinală
- + tipul de mașină deținut: autohton sau de import – var nominală
- + distanța, în km, de acasă până la locul de muncă - var numerică continuă
- + poziția ierarhică ocupată în firmă (funcționar, middle management, top management) - variabilă ordinală
- + poziția privind politica firmei de a reduce salariile în locul reducerii personalului existent (acord total, acord, indiferent, dezacord, dezacord total). – variabilă ordinală

2. Distribuția notelor obținute de un grup de studenți la examenul bazele statisticii se prezintă astfel:

Nota obținută	6	7	8	9	10
Număr studenți	4	8	12	14	5

Să se precizeze volumul eșantionului, unitatea statistică - studentul, variabila-notele și tipul variabilei. Să se identifice și să se interpreteze media, mediana și modul.

Rezolvare

Nota obținută (x_i)	Număr studenți (n_i)	$N_i \downarrow$
6	4	3
7	8	12
8	12	24
9	14	38
10	5	43
Total	43	-

Media aritmetică

$$\bar{x} = \frac{\sum x_i * n_i}{\sum n_i} = \frac{6 * 4 + 7 * 8 + 8 * 12 + 9 * 14 + 10 * 5}{43} = 8.18$$

Interpretare: Studenții au obținut în medie, nota 8.18 la examenul bazele statisticii.

Mediana

1. Se calculează unitatea mediană: $U^{Me} = (n + 1) / 2 = (43 + 1) / 2 = 22$
2. Se calculează $N_i \downarrow$
3. Se află prima valoare $N_i \downarrow \geq U^{Me}$: **24**
4. Valoarea x_i corespunzătoare acesteia este Me: **8**

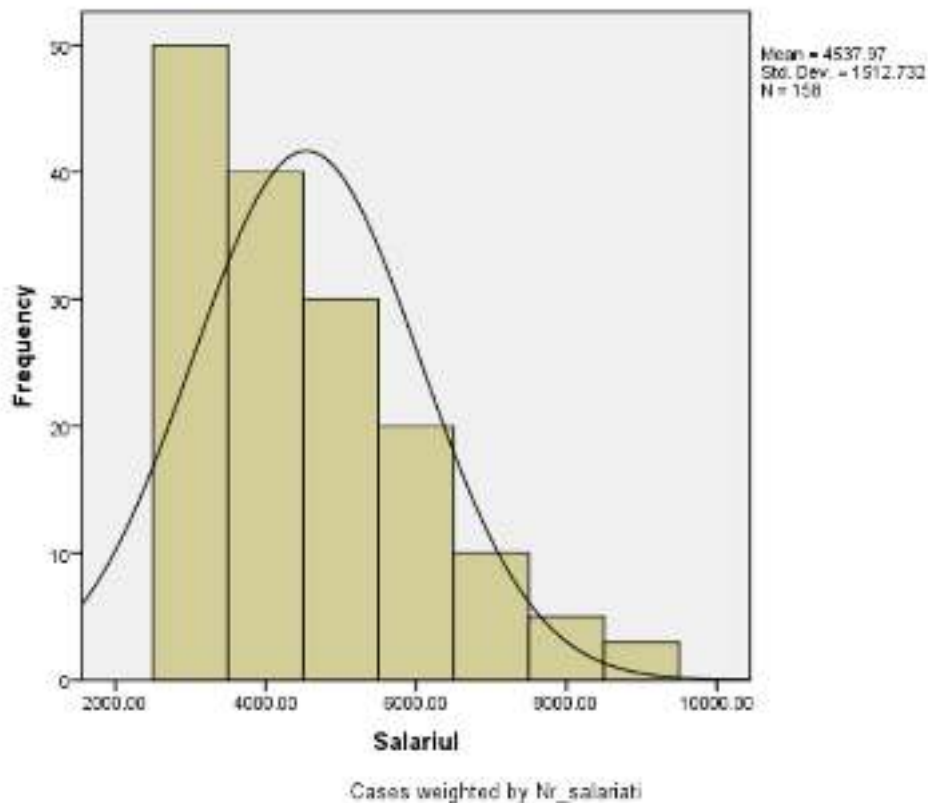
Interpretare: Jumătate dintre studenți au obținut până la nota 8 la examenul bazele statisticii, iar jumătate au obținut peste 8.

Modul

$n_{i \max} = 14 \Rightarrow$ pentru $x_i = 9 \Rightarrow Mo = 9$

Interpretare: Cei mai mulți studenți au obținut nota 9 la examenul bazele statisticii.

3. Distribuția unui eșantion de angajați în funcție de salariul obținut, se prezintă astfel:



Să se caracterizeze distribuția. Media obținută este semnificativă?

Rezolvare

Distribuția analizată este asimetrică la dreapta (pozitivă).

Pentru a identifica daca media este reprezentativa, este necesar sa calculam coeficientul de variatie.

$$v = \frac{s}{\bar{x}} \cdot 100$$

$$v = \frac{1512.73}{5437.97} * 100 = 27.82\%$$

Interpretare: Media esantionului analizat este reprezentativa (coeficientul de variatie este mai mic de 50%).

4. In anul 2020, pentru un esantion de familii, media venitului pe gospodarie era de 4000 lei, mediana de 4500, iar modul de 4300. Sa sa precizeze forma distributiei.

Rezolvare

Avand in vedere ca media are valoarea cea mai mica dintre cei 3 indicatori medii, distribuția este asimetrică la stânga (asimetrie negativă).

5. Se considera distributia unui esantion de studenti in functie de numarul de ore de voluntariat realizat la un camin de copii saptamanal.

Numarul de ore de voluntariat (xi)	Număr studenți (ni)
2	10
3	19
4	80
5	25
6	10
Total	144

Se cere sa se determine:

- + Cati studenti fac 3 ore de voluntariat saptamanal?
- + Cati studenti face cel mult 4 ore de voluntariat saptamanal?
- + Cat % dintre studenti fac cel putin 4 ore de voluntariat saptamanal?
- + Cat % dintre studenti fac maxim 5 ore de de voluntariat saptamanal?
- + Cat % dintre studenti fac 6 ore de de voluntariat saptamanal?
- + Cati studenti fac minim 3 ore de voluntariat saptamanal?
- + Media esantionului este reprezentativa? Dar populatia, este omogena?

Rezolvare

Numarul de ore de voluntariat (xi)	Număr studenți (ni)	Ni ↓ crescator	Ni ↑ decrecator	fi	Fi ↓ (%)	Fi ↑ (%)
2	10	10	144	6.9%	6.9%	100.0%
3	19	29	134	13.2%	20.1%	93.1%
4	80	109	115	55.6%	75.7%	79.9%
5	25	134	35	17.4%	93.1%	24.3%
6	10	144	10	6.9%	100.0%	6.9%
Total	144	-	-	100.0%	-	-

✚ Cati studenți fac 3 ore de voluntariat saptamanal?

19 studenți fac 3 ore de voluntariat saptamanal.

✚ Cati studenți face cel mult 4 ore de voluntariat saptamanal?

$N_3 \downarrow = 10 + 19 + 80 = 109$ studenți

✚ Cat % dintre studenți fac cel puțin 4 ore de voluntariat saptamanal?

$F_3 \uparrow = (80 + 25 + 10)/144 * 100 = 79.9\%$

✚ Cat % dintre studenți fac maxim 5 ore de de voluntariat saptamanal?

$F_4 \downarrow = (10 + 19 + 80 + 25)/144 * 100 = 93.1\%$

✚ Cat % dintre studenți fac 6 ore de de voluntariat saptamanal?

$f_5 = (10/144)*100 = 6.9\%$

✚ Cati studenți fac minim 3 ore de voluntariat saptamanal?

$N_2 \uparrow = 19 + 80 + 25 + 10 = 134$

✚ Media esantionului este reprezentativa? Dar populatia, este omogena?

Este necesar sa calculam coeficientul de variatie.

$$v = \frac{s}{\bar{x}} \cdot 100$$

Media

$$\bar{x} = \frac{\sum x_i * n_i}{\sum n_i} = \frac{2 * 10 + 3 * 19 + 4 * 80 + 5 * 25 + 6 * 10}{144} = 4.04$$

Abaterea standard

$$s = \sqrt{s^2}$$

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2 \cdot n_i}{\sum_i n_i}$$

$$s = \sqrt{\frac{(2 - 4.04)^2 * 10 + (3 - 4.04)^2 * 19 + (4 - 4.04)^2 * 80 + (5 - 4.04)^2 * 25 + (6 - 4.04)^2 * 10}{144}} = 0.93$$

$$v = \frac{0.93}{4.04} * 100 = 23.01\%$$

Coeficientul de variatie este mai mic de 50%, deci **media este reprezentativa**, iar **distributia este omogena**.

6. O variabila ia urmatoarele valori: 14, 5, 8, 2, 12. Sa se identifice si sa se interpreteze media, mediana si abaterea standard.

Rezolvare

Media aritmetica

$$\bar{x} = \frac{\sum_i x_i}{n} = \frac{14 + 5 + 8 + 2 + 12}{5} = 8.2$$

Interpretare: Media variabilei analizate este de 8.2.

Mediana

- Se ordoneaza crescator seria: 2, 5, **8**, 12, 14 => Me = **8**

Interpretare: Jumatate dintre valorile variabilei sunt sub valoarea 8, iar cealalta jumatate peste 8.

Abaterea standard

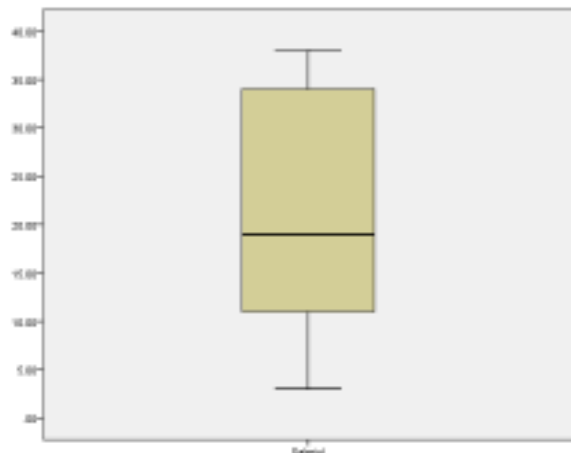
$$s = \sqrt{s^2}$$

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

$$s = \sqrt{\frac{(2 - 8.2)^2 + (5 - 8.2)^2 + (8 - 8.2)^2 + (12 - 8.2)^2 + (14 - 8.2)^2}{5}} = 4.4$$

Interpretare: Valorile variabilei analizate se abat, in medie, fata de nivelul mediu cu 4.4, in ambele sensuri, pozitiv si negativ.

7. In urma analizei statistice privind salariul per familie (exprimat in mii lei), s-a obtinut urmatoarea diagrama Box-plot.

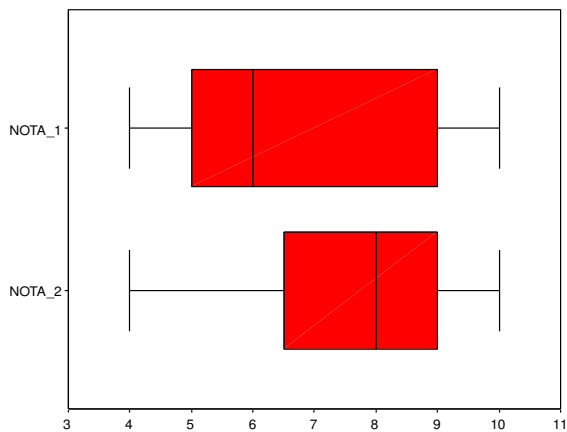


Sa se interpreteze acest rezultat.

Rezolvare

Distributia analizata este asimetrica la dreapta.

8. Pentru doua serii de studenti s-au reprezentat grafic distributiile notelor obtinute la un test. Sa se precizeze sensul asimetriei. Sa se specifice care serie a obtinut note mai mari.



Distributia primei serii este asimetrica la dreapta, distributia celei de-a doua serii este asimetrica la stanga.

Prin urmare, in urma analizarii tuturor acestor informatii, a doua serie a obtinut note mai mari decat prima.

9. Se considera distribuția unui eșantion firme din județul Iasi, observate după valoarea profitului acestora (milioane lei). Datele sunt prezentate in tabelul urmator.

Profitul $[x_{i-1}, x_{i+1})$	Număr firme (n_i)
20-30	10
30-40	20
40-50	40
50-60	35
60-70	7
Total	112

Sa se calculeze si sa se interpreteze media aritmetica

Rezolvare

Profitul $[x_{i-1}, x_{i+1})$	Număr firme (n_i)	x'_i	$x'_i * n_i$
20-30	10	25	250
30-40	20	35	700
40-50	40	45	1800
50-60	35	45	1575
60-70	7	65	455
Total	112	-	4780

Pentru a calcula indicatorii descriptive in cazul unei serii continue, pentru care valorile acesteia sunt grupate pe intervale de variatie, se va proceda mai intai la “discretizarea” seriei, adica la **folosirea mijloacelor intervalelor de variatie a variabilei**, ca valoare considerate reprezentativa pentru tot intervalul.

$$\bar{x} = \frac{\sum x'_i \cdot n_i}{\sum n_i}$$

$$\bar{x} = \frac{4780}{112} = 42.68$$

Interpretare: In medie, o firma realizeaza un profit de 42.68 milioane lei.

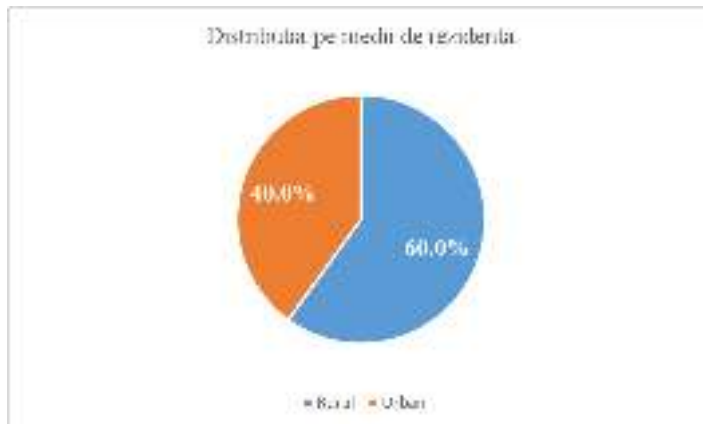
10. Se considera distribuția unui eșantion de masini, observate după pretul acestora (mii euro).
Sa se interpreteze indicatorii prezentati in tabelul de mai jos.

Statistics

N	Valid	155
	Missing	2
Mean		27.39
Median		22.79
Mode		12.64
Std. Deviation		14.35
Variance		205.97
Skewness		1.76
Kurtosis		3.63
Percentiles	25	17.89
	50	22.79
	75	31.96

- **N** este volumul esantionului – 155 masini.
- **Mean** este media aritmetica – pretul mediu al unei masini este de 27.39 mii euro.
- **Median** este mediana - 50% dintre masini au un pret de cel mult 22.79 mii euro, iar 50% au un pret de peste 22.79 mii euro.
- **Mode** este modul - Cele mai multe masini au un pret de 12.64 mii lei.
- **Std. deviation** este abaterea standard sau abaterea medie patratica - In medie, pretul unei masini se abate de la profitul mediu al grupului cu 14.35 mii euro.
- **Variance** este varianța - Nu se interpreteaza.
- **Skewness** este coeficientul de asimetrie Fisher - Distributia este asimetrica la dreapta.
- **Kurtosis** este coeficientul de boltire Fisher - Distributia este leptocurtica.
- **Percentile 25** este Q1 - 25% dintre masini costa cel mult 17.89 mii euro, iar 75% costa peste 17.89 mii euro.
- **Percentile 50** este mediana sau Q2 sau D5 - 50% dintre masini costa cel mult 22.79 mii euro, iar 50% costa peste 22.79 mii euro.
- **Percentile 75** este Q3 - 75% dintre masini costa cel mult 31.96 mii euro, iar 25% costa peste 31.96 mii euro.

11. Reprezentarea grafică a unei populații pe medii de rezidență este realizată în figura de mai jos:



Să se interpreteze valoarea modului.

Rezolvare:

Interpretare: Modul este reprezentat de categoria “mediul rural”.

12. Pentru un esantion de angajați, distribuiți după variabila educație (în ani) și poziția în firmă, s-au obținut următoarele rezultate:

	Junior	Middle	Senior
Media	24	20	36
Varianta	64	16	36
Numar angajati	526	46	92

- ✚ Să se precizeze variabila X și variabila Y;
- ✚ Să se determine care din cele trei grupe de persoane este mai omogenă în raport cu nivelul de educație;
- ✚ Factorul de grupare influențează semnificativ variația variabilei numerice?

Rezolvare:

- ✚ Să se precizeze variabila X și variabila Y;

Variabila X este reprezentată de numărul de ani de educație, iar variabila Y este reprezentată de poziția în firmă.

- ✚ Să se determine care din cele trei grupe de persoane este mai omogenă în raport cu nivelul de educație;

Se calculează coeficienții de variație:

$$v = \frac{s}{\bar{x}} \cdot 100$$

$$v_{junior} = \frac{8}{24} * 100 = 33.3\%$$

$$v_{middle} = \frac{4}{20} * 100 = 20.0\%$$

$$v_{senior} = \frac{6}{36} * 100 = 16.7\%$$

Interpretare: Cea mai omogena grupa este reprezentata de grupa „senior”.

Factorul de grupare influențează semnificativ variația variabilei numerice?

$$k_1 = \frac{s_{\bar{x}_j}^2}{s_X^2} \cdot 100$$

$$k_2 = \frac{\bar{s}^2}{s_X^2} \cdot 100$$

	Junior	Middle	Senior
Media	24	20	36
Varianta	64	16	36
Numar angajati	526	46	92

Etape :

Calculam media pe total

$$\bar{x} = \frac{\sum_{j=1}^p \bar{x}_j \cdot n_{\bullet j}}{\sum_{j=1}^p n_{\bullet j}}$$

$$\bar{x} = \frac{24 * 526 + 20 * 46 + 36 * 92}{664} = 25.38$$

Calculam media varianțelor de grupă (varianța intra-grupe)

$$\bar{s}^2 = \frac{\sum_j s_j^2 \cdot n_{\bullet j}}{\sum_j n_{\bullet j}}$$

$$\overline{s^2} = \frac{64 * 526 + 16 * 46 + 36 * 92}{664} = 56.79$$

Calculam varianța între grupe sau varianța mediilor de grupă (varianța intergrupe)

$$s_{\bar{x}_j}^2 = \frac{\sum_{j=1}^p (\bar{x}_j - \bar{x})^2 \cdot n_{\bullet j}}{\sum_{j=1}^p n_{\bullet j}}$$

$$s_{\bar{x}_j}^2 = \frac{(24 - 25.38)^2 * 526 + (20 - 25.38)^2 * 46 + (36 - 25.38)^2 * 92}{664} = 19.14$$

Calculam varianța generală

$$s_X^2 = \bar{s}^2 + s_{\bar{x}_j}^2$$

$$s_X^2 = 56.79 + 19.14 = 75.93$$

Calculam coeficientul de influență a factorului de grupare și coeficientul de influență a factorilor aleatori/întâmplători

$$k_1 = \frac{s_{x_j}^2}{s_X^2} \cdot 100 = \frac{19.14}{75.93} \cdot 100 = 25.2\%$$

$$k_2 = \frac{\bar{s}^2}{s_X^2} \cdot 100 = \frac{56.79}{75.93} \cdot 100 = 74.8\%$$

Interpretare: Deoarece coeficientul factorului de grupare k_1 este mai mic decât coeficientul influenței factorilor întâmplători k_2 , factorul de grupare, poziția în firmă nu are influență asupra nivelului de educație (variabila numerică exprimată în ani).

În acest caz, varianța din interiorul grupelor este mai mare decât varianța intergrupe. Există diferențe mai mari de comportament între unitățile statistice din aceeași grupă decât între unitățile statistice din grupuri diferite.