Wykład 3. 12.03.2014r.

Przykład 1.

Dzienna wydajność pracy:

Dzienna wydajność pracy w szt. (x_i)	Liczba osób (n_i)
2	6
3	11
4	24
5	15
6	4
	+ 60.

↓

$$\bar{X} = \frac{2\cdot 6 + 3\cdot 11 + 4\cdot 24 + 5\cdot 15 + 4\cdot 6}{60} = 4.$$

Przykład 2.

Dzienne obroty sklepu.

Obroty sklepu	Liczba dni „n_i"	środek przedziału „$\dot{x}$"	$n \cdot x$
1-3	1	2	$1\cdot 2$
3-5	5	4	$4\cdot 5$
5-7	$7 = n_{d-1}$	6	$6\cdot 7$
7-9	$11 = n_d$	8	$8\cdot 11$
9-11	$8 = n_{d+1}$	10	$10\cdot 8$
11-13	6	12	$12\cdot 6$
13-15	3	14	$14\cdot 3$
	+ 41.		

$9-7=2 = h_d = x - x_{od}$

$$\bar{X} = \frac{1}{41}\sum_{i=1}^{7} \dot{x}_i \cdot n_i = \frac{2+20+42+88+80+72+\ldots}{}$$

1. PARAMETRY ROZKŁADU (miary statystyczne).

- Po wstępnej analizie danych (po porządkowaniu, grupowaniu, przedstawianiu ich w postaci tabelarycznej i graficznej) należy przejść do wyznaczania charakterystyk liczbowych badanych cech.
- Syntetyczne własności rozkładu badanej cechy opisane są przez parametry zwane też CHARAKTERYSTYKAMI ROZKŁADU albo MIARAMI STATYSTYCZNYMI.

● MIARY OPISU ROZKŁADU dzielimy na:

1) miary położenia
2) miary zróżnicowania
3) miary asymetrii
4) miary koncentracji.

2. MIARY POŁOŻENIA:

a) klasyczne (średnia arytmetyczna, średnia geometryczna, średnia harmon

b) pozycyjne (wartość modalna, mediana, kwantyle).

- Miary położenia, które służą do wyznaczania takiej wartości cechy, wokół której grupują się dane, nazywają się MIARAMI TENDENCJI CENTRALNEJ BADANEJ CECHY.

* Miarami tendencji centralnej cechy są:
- średnia
- mediana
- moda.

Np. Badając zarobki osób mówiąc, że średnia wynosi 3800 zł.
Oznacza to, że większość osób będzie miała zarobki zbliżone do tej śr. chociaż będzie dużo osób o zarobkach niższych i wyższych.

$$\frac{x_1 + x_2 + x_3}{3} = x_1 \cdot \underbrace{\left(\frac{1}{3}\right)}_{w_1} + x_2 \cdot \underbrace{\left(\frac{1}{3}\right)}_{w_2} + x_3 \cdot \underbrace{\left(\frac{1}{3}\right)}_{w_3}$$

3. ŚREDNIA ARYTMETYCZNA:

$$\boxed{\overline{X} = \frac{x_1 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

- dla danych niepogrupowanych $x_1 + x_2, \ldots x_n$.

- Gdy dane są pogrupowane w szereg rozdzielczy punktowy to:

$$\boxed{\overline{X} = \frac{n_1 x_1 + \ldots + n_k x_k}{n} = x_1 w_1 + \ldots + x_k w_k}$$

- Gdy dane są pogrupowane w szereg rozdzielczy przedziałowy to:

$$\overline{X} = \frac{n_1 \mathring{x}_1 + \ldots + n_k \mathring{x}_k}{n} = \mathring{x}_1 w_1 + \ldots + \mathring{x}_k w_k$$

- Średnia arytmetyczna ($\bar{x}$) jest wypadkową wszystkich zaobserwowanych wartości cechy.
- W przypadku danych ~~cechy~~ pogrupowanych, średnią arytmetyczną ($\bar{x}$) nazywa się również ŚREDNIĄ WAŻONĄ.

4. ŚREDNIA GEOMETRYCZNA: ($\bar{x}_G$) - liczy się ją tylko dla cechy przyjmującej wartości dodatnie. Dla szeregu szczegółowego:

$$\boxed{\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}}$$

gdzie:

x_i = wielkość (dynamika) zjawiska

n = liczba obserwowanych elementów (zjawisk).

- Średnią geometryczną stosuje się przy badaniu średniego tempa zmian (wzrost lub spadek) zjawisk w czasie (analizy dynamiki zjawiska).

Również stosuje się ją, gdy wartości cechy są mocno zróżnicowane, jeśli chodzi o ich częstotliwość występowania.

5. ŚREDNIA HARMONICZNA: - jest odwrotnością średniej arytmetycznej z odwrotności wartości cechy.

- Dla szeregów szczegółowych, średnia harmoniczna:

$$\boxed{\bar{x}_H = \frac{n}{\frac{1}{x_1} + \ldots \frac{1}{x_n}}}$$

- Dla szeregów rozdzielczych punktowych:

$$\boxed{\bar{x}_H = \frac{n}{\frac{n_1}{x_1} + \ldots + \frac{n_k}{x_k}}}$$

- Dla szeregów rozdzielczych przedziałowych $\bar{x}_H$ wyznacza się jak wyżej, tylko zamiast x_i wstawia się wartości $\mathring{x}_i$.
- Średnią harmoniczną stosuje się, gdy wartości cechy podane są w jednostkach względnych

6. WARTOŚĆ MODALNA (moda, dominanta).

- Wartość modalna - Mo - jest wartością cechy pojawiającą się najczęściej.
- Gdy dane są pogrupowane, wartość Mo wyznacza się wzorem:

$$\boxed{Mo = x_{0d} + \frac{n_d - n_{d-1}}{(n_d - n_{d-1}) + (n_d - n_{d+1})} \cdot h_d}$$

gdzie:

x_{0d} - dolna granica przedziału, w którym występuje Mo

h_d - rozpiętość tego przedziału

n_d - liczebność przedziału, w którym występuje Mo

n_{d-1} - -||- -||- przedziału poprzedniego

n_{d+1} - -||- -||- -||- -||- następnego.

7. MEDIANA.

- MEDIANA - Me - jest to taki punkt, że co najmniej połowa danych jest od n większa, a co najmniej połowa jest niemniejsza.
- Dla szeregu szczegółowego mediana określa się wzorem:

$$\boxed{Me = \begin{cases} x_{(n+1)/2} & \text{gdy } n \text{ jest nieparzyste} \\ \dfrac{x_{n/2} + x_{n/2+1}}{2} & \text{gdy } n \text{ jest parzyste.} \end{cases}}$$

8. KWARTYLE:

- KWARTYL DOLNY - Q_1 - jest to taki punkt, że co najmniej jedna czwarta dan jest od niego niewiększa, a co najmniej trzy czwarte jest niemniejsze.
- Kwartylem Q_2 - jest mediana (Me).
- Kwartyl Górny - Q_3 - jest to taki punkt, że co najmniej jedna czwarta dan jest od niego niemniejsza, a co najmniej trzy czwarte jest niewiększe.

- WZORY NA KWARTYLE, gdy dane są w postaci szeregu rozdzielczego przedziałowego:

$$Q_1 = XQ_1 + \left(0{,}25n - \sum_{i=1}^{m_1-1} n_i\right) \frac{hQ_1}{nQ_1}$$

→ Wzory na kwartyle Q_2 i Q_3:

$$Me = Q_2 = XQ_2 + \left(0{,}5n - \sum_{i=1}^{m_2-1} n_i\right) \frac{hQ_2}{nQ_2}.$$

$$Q_3 = XQ_3 + \left(0{,}75n - \sum_{i=1}^{m_3-1} n_i\right) \frac{hQ_3}{nQ_3}$$

XQ_1, XQ_2, XQ_3 - dolne granice przedziałów, w których znajdują się odpowiednie Q_1; Q_2; Q_3

nQ_1, nQ_2, nQ_3 - liczebności przedziałów, w których znajdują się odpowiedni Q_1; Q_2; Q_3.

hQ_1; hQ_2; hQ_3 - rozpiętości przedziałów, w których znajdują się odpowiedni Q_1, Q_2; Q_3.

m_1; m_2; m_3 - numery przedziałów, w których znajdują się odpowiedn Q_1, Q_2, Q_3.

Wykład 4. 19.03.2014r.

Temat: Miary zmienności.

1. KLASYFIKACJA MIAR ZMIENNOŚCI:

- MIARY nazywane są MIARAMI ZRÓŻNICOWANIA lub ROZPROSZENIA.
- Miary te pozwalają sprawdzić, czy informacje o wartościach badanej cechy są bardzo rozproszone czy skoncentrowane.

* Klasyfikacja miar rozproszenia:

a) BEZWZGLĘDNE

→ klasyczne (wariancję; odchylenie standardowe)

→ pozycyjne (odchylenie ćwiartkowe).

b) WZGLĘDNE

→ klasyczne (współczynnik zmienności oparty na miarach klasycznych)

→ pozycyjne (współczynnik zmienności oparty na miarach pozycyjnych).

2. WARIANCJA - informuje o stopniu rozproszenia wartości badanej cechy wokół jej średniej arytmetycznej.

- Dla danych niepogrupowanych wariancję empiryczną określa się wzorem:

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

np.

x_i	n_i
2	6
3	11
4	24
5	15
6	4
$\overline{x}=4.$	$\Sigma=60$

$$s^2 = \frac{1}{60}\left[(2-4)^2\cdot 6 + (3-4)^2\cdot 11 + (4-4)^2\cdot 24 + (5-4)^2\cdot 15 + (6-4)\right.$$

$$= \frac{90}{60} = \frac{3}{2} = 1{,}5$$

• Wariancja dla danych pogrupowanych w szereg rozdzielczy punktowy:

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

• Wariancja dla danych pogrupowanych w szereg rozdzielczy przedziałowy:

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}(\mathring{x}_i-\overline{x})^2 n_i$$

*Zawsze $s^2 > 0$.

Przykład 2.

$x_i - x_n$	n_i	$\mathring{x}$
1-3	1	2
3-5	5	4
5-7	7	6
7-9	11	8
9-11	8	10
11-13	6	12
13-15	3	14

$\overline{x} = 8,44.$

$$s^2 = \frac{1}{41}\left[(2-8,44)^2\cdot 1 + (4-8,44)^2\cdot 5 + (6-8,44)^2\cdot 7 + (8-8,44)^2\cdot 11 + (10-8,44)^2\cdot 8 + (12-8,44)^2\cdot 6 + (14-8,44)^2\cdot 3\right] = 9,09.$$

* Równoważne wzory na wariancję:

→ dla danych niepogrupowanych:

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

→ dla danych pogrupowanych w szereg rozdzielczy punktowy:

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}x_i^2 n_i - \overline{x}^2$$

→ dla danych pogrupowanych w szereg rozdzielczy przedziałowy:

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}\mathring{x}_i^2 n_i - \overline{x}^2$$

- Im zbiorowość jest bardziej zróżnicowana, tym większa jest WARIANCJA. „Odchylenie standardowe empiryczne" jest pierwiastkiem z wariancji.

3. ODCHYLENIE STANDARDOWE:

- WARIANCJA jest podawana w jednostkach cechy podniesionych do kwadratu np. w $zł^2$; cm^2. Aby tego uniknąć, stosuje się jako ODCHYLENIE STANDARDOWE.

→ ODCHYLENIE STANDARDOWE EMPIRYCZNE - jest pierwiastkiem kwadratowym z wariancji empirycznej.

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

- Im wyższe odchylenie standardowe, tym rozproszenie cechy większe. Obserwacje różniące się od średniej $\overline{x}$ nie więcej niż s, nazywamy obserwacjami typowymi, czyli należą do klasycznego obszaru zmienności.
- Klasyczny obszar zmienności:

$$\overline{x} - s < x_{typ} < \overline{x} + s.$$

4. BEZWZGLĘDNA POZYCYJNA MIARA ROZPROSZENIA:
- Pozycyjną miarą rozproszenia jest odchylenie ćwiartkowe.

$$Q = \frac{Q_3 - Q_1}{2}.$$

- Interpretacja Q: wartości cechy różnią się od mediany przeciętnie o ale dotyczy tylko środkowych 50% obserwacji.
- Pomiędzy odchyleniami ćwiartkowym i standardowym zachodzi nierówność:

$$Q \leq s.$$

- Obserwacje różniące się od mediany (Me) nie więcej niż Q, nazywamy OBSERWACJAMI TYPOWYMI, czyli należą do pozycyjnego obszaru zmienności.

- Pozycyjny obszar zmienności:

$$Me - Q < X_{typ} < Me + Q$$

5. WZGLĘDNE MIARY ROZPROSZENIA - współczynniki zmienności:

- Klasyczny współczynnik zmienności:

$$V_S = \frac{S}{\bar{x}}$$

- Pozycyjny współczynnik zmienności:

$$V_Q = \frac{Q}{Me}$$

- Współczynniki zmienności informują o sile rozproszenia wartości badanej cechy. Duże ich wartości świadczą o niejednorodności badanej zbiorowości.
- Współczynnik zmienności jest stosowany do porównywania i określenia, który zbiór informacji jest bardziej zróżnicowany w stosunku do swojej wartości średniej.