Projekt realizowany z Narodowym Bankiem Polskim w ramach programu edukacji ekonomicznej

SIGMA KWADRAT

LUBELSKI KONKURS STATYSTYCZNO-DEMOGRAFICZNY

Projekt realizowany
z Narodowym Bankiem Polskim
w ramach programu edukacji ekonomicznej

Opisowa analiza struktury zjawisk masowych

OPIS STATYSTYCZNY,

czyli stwierdzenie jak jest, dotyczyć może tak zbiorowości generalnych jak i próbnych. Wykorzystujemy wtedy numeryczne lub graficzne metody opisu struktury, dynamiki i korelacji badanych zjawisk.

Należy pamiętać, że w wielu przypadkach poddanie badaniu wszystkich jednostek statystycznych jest *niemożliwe* (np. gdy zbiór jednostek jest nieskończony, zbiór ryb w morzu) lub *niecelowe* (np. ze względu na koszty).

Jeśli obserwacja statystyczna jest częściowa, ważną informacją jest stwierdzenie faktu, jaką drogą (*losową lub nielosową*) została pobrana próba z całej zbiorowości, bowiem tylko w przypadku losowego wyboru próby można w merytorycznie prawidłowy sposób stosować statystyczne metody wnioskowania z próby losowej o populacji generalnej.

W przypadku próby nielosowej należy poprzestać na opisie statystycznym, ponieważ przenoszenie wyników z prób nielosowych na zbiorowości generalne jest nieuzasadnione.

Opis statystyczny

Opis statystyczny ma sumaryczny charakter i dotyczy tylko i wyłącznie danej zbiorowości statystycznej. Przedmiotem opisu statystycznego są zatem obserwacje będące rezultatem badania pełnego, a jego narzędziem są odpowiednie miary, zwane **charakterystykami** lub **parametrami opisowymi**.

Parametrami opisowymi zbiorowości statystycznej nazywa się takie liczby, które z przyjętego punktu widzenia pozwalają na sumaryczną charakterystykę tej zbiorowości.

Do obliczania parametrów wykorzystuje się wzory, które za pomocą cyfr, liter i znaków algebraicznych wyrażają relacje między danymi wielkościami szeregu statystycznego.

Opis statystyczny

Analiza danych statystycznych powinna doprowadzić do zwięzłego przedstawienia wyników badań za pomocą odpowiednich charakterystyk liczbowych zwanych parametrami statystycznymi.

Parametry statystyczne to liczby służące do syntetycznego opisu struktury zbiorowości statystycznej, do nich zalicza się: miary położenia, zmienności, asymetrii, koncentracji.

Im mniej parametrów wykorzystujemy do opisu, tym więcej tracimy informacji o zbiorowości.

Opis statystyczny

W badaniach przekrojowych obliczane parametry opisowe dotyczą struktury zjawisk masowych. Analizę struktury przeprowadza się za pomocą charakterystyk zaliczanych do czterech grup:

- **miary średnie** (klasyczne i pozycyjne),
- **miary zmienności** (klasyczne i pozycyjne, względne i bezwzględne),
- **miary asymetrii** (skośności),
- **miary koncentracji i spłaszczenia** (kurtozy).

Opis statystyczny

Tabela: Podstawowe metody analizy statystycznej

<table>
<tr><th>Metody analizy</th><th>Opis statystyczny</th><th colspan="2">Wnioskowanie statystyczne</th></tr>
<tr><td rowspan="4">Struktury zjawisk</td><td>Miary położenia</td><td rowspan="9">Estymacja parametrów populacji (punktowa i przedziałowa)</td><td rowspan="9">Weryfikacja hipotez statystycznych (parametrycznych i nieparametrycznych)</td></tr>
<tr><td>Miary zmienności</td></tr>
<tr><td>Miary asymetrii</td></tr>
<tr><td>Miary koncentracji</td></tr>
<tr><td rowspan="3">Współzależności zjawisk</td><td>Rachunek korelacyjny</td></tr>
<tr><td>Rachunek regresyjny</td></tr>
<tr><td>Rachunek wariancyjny</td></tr>
<tr><td rowspan="2">Dynamiki zjawisk</td><td>Model wahań w czasie</td></tr>
<tr><td>Indeksy dynamiki</td></tr>
</table>

Źródło: opracowanie na podstawie A. Luszniewicz, T. Słaby, Statystyka z pakietem komputerowym STATISTICA™PL, Wydawnictwo C.H.Beck, Warszawa 2001, s. 16

Opis statystyczny

Analiza danych statystycznych powinna *doprowadzić* do *zwięzłego* przedstawienia *wyników badań* za *pomocą* odpowiednich charakterystyk liczbowych zwanych parametrami statystycznymi.

<u>Parametry statystyczne</u> to liczby *służące* do syntetycznego opisu struktury *zbiorowości* statystycznej, do nich zalicza *się*: miary *położenia*, *zmienności*, asymetrii, koncentracji.

Im mniej *parametrów* wykorzystujemy do opisu, tym *więcej* tracimy informacji o *zbiorowości*.

Opis statystyczny

Tabela: Podstawowe metody analizy statystycznej

Metody analizy	Opis statystyczny	Wnioskowanie statystyczne	
Struktury zjawisk	Miary położenia	Estymacja parametrów populacji (punktowa i przedziałowa)	Weryfikacja hipotez statystycznych (parametrycznych i nieparametrycznych)
	Miary zmienności		
	Miary asymetrii		
	Miary koncentracji		
Współzależności zjawisk	Rachunek korelacyjny		
	Rachunek regresyjny		
	Rachunek wariancyjny		
Dynamiki zjawisk	Model wahań w czasie		
	Indeksy dynamiki		

Źródło: opracowanie na podstawie A. Luszniewicz, T. Słaby, Statystyka z pakietem komputerowym STATISTICA^{TM}PL, Wydawnictwo C.H.Beck, Warszawa 2001, s. 16

Opis statystyczny

<u>ROZKŁAD EMPIRYCZNY</u> **to rozkład cechy, w którym poszczególnym wariantom (wartościom) cechy podanym punktowo lub przedziałowo przyporządkowane są liczebności z jakimi dane warianty cechy w badanej zbiorowości występują.**

ROZKŁAD EMPIRYCZNY **otrzymuje się w wyniku badania statystycznego i najczęściej prezentuje się w formie szeregu statystycznego, tablicy statystycznej lub wykresu statystycznego.**

Opis statystyczny

Do oceny rozkładu empirycznego badanej zbiorowości wykorzystuje się szereg różnych charakterystyk liczbowych (mierników statystycznych), które mogą mieć charakter liczb absolutnych albo względnych.

Liczby absolutne (bezwzględne) są to wielkości, które otrzymuje się w wyniku mierzenia lub sumowania jednostek badanej zbiorowości. Mają one charakter wielkości pierwotnych, są liczbami mianowanymi i wyraża się je w sztukach, litrach, metrach, kilogramach, hektarach, złotówkach itp.., itd.

Liczby względne są to wielkości uzyskane w drodze dzielenia przez siebie dwóch liczb absolutnych. Przedstawiają one stosunki pomiędzy liczbami absolutnymi.

Opis statystyczny

Tabela: Przegląd podstawowych wybranych charakterystyk liczbowych rozkładów empirycznych stosowanych w analizie struktury

Wyszczególnienie	**Charakterystyki liczbowe**	
	klasyczne	**pozycyjne**
udział	***wskaźnik struktury***	
natężenie	***wskaźnik natężenia***	
poziom przeciętny	***średnia arytmetyczna***	***mediana, dominanta***
zmienność	***wariancja, odchylenie standardowe, współczynnik zmienności***	***odchylenie ćwiartkowe, pozycyjny współczynnik zmienności***
asymetria	***współczynnik asymetrii***	***pozycyjny współczynnik asymetrii***

Źródło: opracowanie własne.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH STOSOWANYCH W ANALIZIE STRUKTURY

1. **Wskaźnik struktury i natężenia**
2. **Miary położenia**

- klasyczne: **średnia arytmetyczna**
średnia harmoniczna
średnia geometryczna

- pozycyjne:
mediana i kwartyle
dominanta

3. Miary zmienności

- klasyczne: **wariancja**
odchylenie standardowe
odchylenie przeciętne
współczynnik zmienności

- pozycyjne:
odchylenie ćwiartkowe
współczynnik zmienności

4. Miary asymetrii i kurtozy
wskaźnik i współczynnik asymetrii
współczynnik kurtozy
współczynnik koncentracji Lorenza

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 1 Wskaźnik struktury i natężenia

Wskaźnik struktury **jest stosunkiem części badanej zbiorowości statystycznej do całości tej zbiorowości. Wskaźnik ten określa udział (częstość) części zbiorowości w całej zbiorowości, stąd często nazywa się go udziałem (częstością) względnym lub stosunkiem względnym. Analitycznie wskaźnik struktury zapisuje się następująco:**

- *wskaźnik struktury*:

$$\mathrm{w}_{\mathrm{i}} = \frac{\mathrm{n}_{\mathrm{i}}}{\mathrm{n}}, \quad \textbf{przy czym: } \sum_{i=1}^{k} n_i = n, \quad \textbf{a} \quad \sum_{i=1}^{k} w_i = 1\textbf{, czyli } \mathbf{0 \le w_i \le 1.}$$

gdzie: **n_i – liczba jednostek posiadających cechę i,**
n - ogólna liczba jednostek badanej zbiorowości statystycznej.

Wskaźnik struktury można wyrazić w procentach mnożąc powyższą formułę przez 100. Należy pamiętać, że:

$$\sum_{i=1}^{k} n_i = n, \quad \mathrm{a} \quad \sum_{i=1}^{k} w_i = 1 \ \mathrm{lub}\, 100$$

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH STOSOWANYCH W ANALIZIE STRUKTURY

Ad. 1 Wskaźnik struktury i natężenia

Wskaźnik natężenia **określa stosunek dwóch wielkości pozostających ze sobą w logicznym związku. Wskaźnik natężenia pozwala ustalić ile jednej wielkości przypada na jednostkę drugiej wielkości i oblicza się według wzoru:**

- ***wskaźnik natężenia*:** $W_n = \frac{w_1}{w_2}$,

gdzie: w_1 – wielkość pierwsza,
w_2 - wielkość druga logicznie powiązana z wielkością w_1.

Przykładowe współczynniki natężenia:
stopa bezrobocia - stosunek liczby bezrobotnych do liczby ludności czynnej zawodowo,
gęstość zaludnienia - liczba ludności przypadająca na 1 km^2 powierzchni,
wskaźnik umieralności - liczba zmarłych do średniej liczby ludności,
wskaźnik rozwoju gospodarczego - produkt krajowy brutto (netto) do liczby ludności kraju,
wskaźnik wydajności pracy - wielkość produkcji do czasu pracy,
wskaźnik spożycia i usług - wielkość spożycia i usług do liczby ludności,
wskaźnik rentowności - zysk do wielkości sprzedaży,
wskaźnik efektywności - zysk do zaangażowanego kapitału,
wskaźnik produktywności - sprzedaż do do zaangażowanego kapitału

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 2 Miary położenia

- klasyczne

***<u>Średnia arytmetyczna</u>* ($\overline{x}$) jest definiowana jako suma wszystkich wariantów cechy w zbiorowości statystycznej (w rozkładzie empirycznym) podzielona przez jej liczebność. Technika jej wyznaczania zależy od rodzaju szeregu statystycznego:**

- szereg szczegółowy:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

- szereg rozdzielczy punktowy:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + ... + x_k n_k}{n} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i,$$

- szereg rozdzielczy przedziałowy:

$$\bar{x} = \frac{\hat{x}_1 n_i + \hat{x}_2 n_2 + ... + \hat{x}_k n_k}{n} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i.$$

Średnia arytmetyczna dla szeregu rozdzielczego nosi również nazwę średniej ważonej ze względu na fakt, że wariantom cechy x_i w tych szeregach odpowiadają określone liczebności n_i zwane „wagami”.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 2 Miary położenia

- klasyczne

<u>Średnia harmoniczna</u> **jest odwrotnością średniej arytmetycznej z odwrotności wariantów cechy. Technika obliczania średniej harmonicznej zależy od rodzaju szeregu statystycznego.**

Szereg szczegółowy:

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}},$$

Szereg rozdzielczy punktowy:

$$H = \frac{n}{\sum_{i=1}^{k} \frac{1}{x_i} n_i},$$

Szereg rozdzielczy przedziałowy:

$$H = \frac{n}{\sum_{i=1}^{k} \frac{1}{x_i'} n_i}.$$

Średnią harmoniczną **stosuje się wtedy, gdy wartości cechy podane są w liczbach względnych, tzn. w km/h, kg/osobę, zł/szt. itp., a wagi, czyli n_i podane są w jednostkach licznika tych liczb.**

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 2 Miary położenia
- *klasyczne*

Średnia geometryczna (średnie tempo zmian)

<u>Średnia geometryczna</u> **jest pierwiastkiem n-tego stopnia z iloczynu n indeksów dynamiki łańcuchowych. Zgodnie z tym określeniem wzór na średnią geometryczną ma następującą postać:**

$$\bar{i} = \sqrt[n]{i_1 i_2 \ldots i_n} = \sqrt[n]{\frac{y_1}{y_0} \cdot \frac{y_2}{y_1} \cdot \ldots \cdot \frac{y_n}{y_{n-1}}} = \sqrt[n]{\frac{y_n}{y_0}}$$

Średnia geometryczna **mierzy średnie tempo zmian zjawiska w dłuższym przedziale czasowym, ale także może być wykorzystana do celów prognozowania krótkookresowego przy założeniu, że średnie tempo zmian zjawiska w okresie prognozowanym nie ulegnie zmianie.**

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 2 Miary położenia

- pozycyjne: **MEDIANA I KWARTYLE**

Mediana **Me (wartość środkowa) jest to taka wartość cechy w rozkładzie empirycznym uporządkowanym rosnąco, jaką posiada jednostka badanej zbiorowości, która dzieli wszystkie jednostki badanej zbiorowości na dwie jednakowo liczne części, z których pierwsza, stanowiąca połowę posiada wartość cechy nie mniejszą niż mediana, a druga również stanowiąca połowę posiada wartość cechy nie większą niż mediana.**

Technika wyznaczania mediany zależy od rodzaju szeregu statystycznego.

- szereg szczegółowy i rozdzielczy punktowy:

a) gdy n jest nieparzyste: $$M_e = X_{\frac{n+1}{2}}$$

b) gdy n jest parzyste: $$M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 2 Miary położenia

-pozycyjne:

MEDIANA **cd.**

- szereg rozdzielczy przedziałowy:

$$M_e = x_{Me} + \frac{\frac{N}{2} - n_{cum-1}}{n_{Me}} h_{Me}$$

gdzie: x_{Me} **– dolna granica przedziału mediany,**

N **- liczebność badanej zbiorowości,**

n_{Me} **– liczebność przedziału mediany,**

n_{cum-1} **- suma liczebności przedziałów poprzedzających przedział mediany,**

h_{Me} **– rozpiętość przedziału mediany.**

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 2 Miary położenia

-pozycyjne: ***KWARTYLE***

Obok mediany, jako miary pozycyjne rozkładu zmiennej losowej można stosować kwartale, decyle i percentyle.

Kwartylem pierwszym Q_1 przyjęto nazywać taką wartość zmiennej w rozkładzie empirycznym, poniżej której znajduje się ¼ jednostek badanej zbiorowości. Kwartylem trzecim Q_3 nazywamy wartość, poniżej której znajduje się ¾ jednostek zbiorowości.

Wartości Q_1 i Q_3 obliczamy w następujący sposób:

$$Q_1 = x_0 + \frac{\frac{N}{4} - n_{cum-1}}{n_0} h_0 \qquad Q_3 = x_0 + \frac{\frac{3N}{4} - n_{cum-1}}{n_0} h_0$$

x_0 – **dolna granica przedziału, który zawiera dany kwartyl,**

n_{cum-1} – **skumulowana liczebność przedziałów poprzedzających przedział danego kwartyla,**

n_0 – **liczebność przedziału kwartyla,**

h_0 – **rozpiętość przedziału kwartyla.**

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 2 Miary położenia

-pozycyjne: ***DOMINANTA***

Dominanta **jest to taki wariant cechy, który w badanym rozkładzie występuje najczęściej. W szeregach szczegółowych i rozdzielczych punktowych dominantę można wskazać bezpośrednio, gdyż jest ona obserwowalna bezpośrednio bez żadnych obliczeń. W rozkładzie rozdzielczym przedziałowym dominantę wyznacza się w przybliżeniu ze wzoru:**

$$D \approx x_D + \frac{n_D - n_{D-1}}{(n_D - n_{D-1}) + (n_D - n_{D+1})} h_D$$

gdzie: x_D **- dolna granica klasy najczęstszej,**

n_D **- liczebność klasy najczęstszej,**

n_{D-1} **- liczebność klasy poprzedzającej klasę najliczniejszą,**

n_{D+1} **- liczebność klasy następnej po najliczniejszej,**

h_D **- rozpiętość klasy najliczniejszej.**

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

W rozkładach umiarkowanie asymetrycznych między miarami tendencji centralnej zachodzi następująca relacja (wzór Pearsona):

$$\bar{x} - \mathrm{D} = 3(\bar{x} - \mathrm{Me})$$

Ad. 3. <u>Miary zmienności</u>

Obok oceny poziomu przeciętnego cechy w badanej zbiorowości statystycznej istnieje konieczność zmierzenia zmienności (rozproszenia, dyspersji) wariantów cechy czyli różnic, jakie między nimi występują. Do mierzenia zmienności wariantów cechy w zbiorowości statystycznej służą miary zmienności (rozproszenia, dyspersji), spośród których należy zwrócić uwagę na:

- **wariancję,**
- **odchylenie standardowe,**
- **współczynnik zmienności,**
- **odchylenie ćwiartkowe i**
- **pozycyjny współczynnik zmienności.**

Punktem odniesienia dla liczbowego wyznaczenia miar zmienności jest albo średnia arytmetyczna (klasyczne miary zmienności), albo mediana (pozycyjne miary zmienności).

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 3. Miary zmienności

Klasyczne: **WARIANCJA**

Wariancja **jest to średnia arytmetyczna kwadratów odchyleń wariantów cechy od średniej arytmetycznej tej cechy. Sposób wyznaczania wariancji zależy od rodzaju szeregu statystycznego. Wyznacza się ją z następującego wzoru:**

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$ **dla szeregu wyliczającego,**

$$S^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i$$ **dla szeregu rozdzielczego punktowego,**

$$S^2 = \frac{1}{n}\sum_{i=1}^{k}(\hat{x}_i - \overline{x})^2 n_i$$ **dla szeregu rozdzielczego przedziałowego.**

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 3. Miary zmienności

Klasyczne: **WARIANCJA**

Należy pamiętać, ze we wszystkich pakietach komputerowych oblicza się wariancję nieobciążoną $\hat{S}^2$, tzn. że w mianowniku znajduje się nie n lecz n-1. Wzajemną relację między $\hat{S}^2$, a S^2 zapisuje się:

$$S^2 = \frac{n-1}{n}\hat{S}^2$$

Wariancja jest miarą nieinterpretowalną (jest miarą, której się nie interpretuje ze względu na kwadrat odchylenia wartości cechy od poziomu przeciętnego). W celach interpretacyjnych stosuje się pierwiastek z wariancji czyli

odchylenie standardowe S.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 3. Miary zmienności

Klasyczne: ***ODCHYLENIE STANDARDOWE***

Sens interpretacyjny posiada pierwiastek kwadratowy z wariancji zwany odchyleniem standardowym:

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

Odchylenie standardowe określa średnie zróżnicowanie wariantów cechy w badanej zbiorowości od średniej arytmetycznej tej cechy. Traktuje się je za najdokładniejszą miarę zróżnicowania wartości cechy i szeroko wykorzystuje w badaniach statystycznych.

POPRAWKA SHEPPARDA

Obliczając odchylenie standardowe dla szeregu rozdzielczego przedziałowego przyjmujemy założenie, że rozkład liczebności w przedziałach jest równomierny i proporcjonalny do wzrostu wartości zmiennej, czego wyrazem jest przyjmowany do obliczeń środek przedziału klasowego. Jednak nie zawsze rozkład liczebności w przedziałach jest równomierny, ani symetryczny. Błąd spowodowany tym, że zamiast faktycznych średnich arytmetycznych przyjmujemy do obliczeń środki przedziałów klasowych jest tym większy, im większa jest rozpiętość przedziałów klasowych (i im mniejsza liczba przedziałów). Błąd ten zwykle powoduje podwyższenie wartości odchylenia standardowego.

$$s_{skr} = \sqrt{s^2 - \frac{h^2}{12}}$$

gdzie: h – rozpiętość przedziałów klasowych.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 3. Miary zmienności

***Klasyczne*:** **ODCHYLENIE PRZECIĘTNE**

Odchylenie przeciętne (d) jest to średnia arytmetyczna z bezwzględnych odchyleń wartości zmiennej X od średniej arytmetycznej.

$$d_x = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \overline{x}\right|$$

dla szeregu wyliczającego,

$$d_x = \frac{1}{n}\sum_{i=1}^{k}\left|x_i - \overline{x}\right| n_i$$

dla szeregu rozdzielczego punktowego,

$$d_x = \frac{1}{n}\sum_{i=1}^{k}\left|\hat{x}_i - \overline{x}\right| n_i$$

dla szeregu rozdzielczego przedziałowego.

Ad. 3. Miary zmienności

Klasyczne: ***WSPÓŁCZYNNIK ZMIENNOŚCI***

Współczynnik zmienności wykorzystuje się wówczas, gdy zachodzi potrzeba porównania zmienności wartości cechy w różnych zbiorowościach, bądź gdy chcemy porównać zmienność tej samej zbiorowości statystycznej, ale badanej pod względem różnych cech. Jest miarą względną.

Współczynnik zmienności wyznacza się korzystając z następującego wzoru:

$$V = \frac{S}{\bar{x}}100$$

Współczynnik zmienności wyraża się zwykle w procentach.

Określa udział odchylenia standardowego w średniej arytmetycznej i pozwala ocenić siłę zróżnicowania, a przede wszystkim porównać zróżnicowanie różnych cech lub różnych zbiorowości.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 3. Miary zmienności

<u>Pozycyjne:</u>

Poza klasycznymi miarami zmienności korzysta się także z pozycyjnych miar zmienności. Najważniejszymi z nich jest odchylenie ćwiartkowe i pozycyjny współczynnik zmienności. Pozycyjne miary zmienności mierzą przeciętną zmienność wariantów cechy w badanej zbiorowości nie względem średniej arytmetycznej, lecz mediany.

ODCHYLENIE ĆWIARTKOWE

Odchylenie ćwiartkowe **jest to ½ obszaru zmienności, a tym samym 50% środkowych jednostek zbiorowości:**

$$Q_c = \frac{Q_3 - Q_1}{2}$$

gdzie: Q_1 – ***kwartyl pierwszy,***
Q_3 – ***kwartyl trzeci*****.**

Badanie zmienności 50% środkowych jednostek zbiorowości uzasadnione jest dążeniem do wyeliminowania wpływu na tę miarę rozproszenia wartości skrajnych umiejscowionych na krańcach rozkładu w szeregu uporządkowanym, tj. w I i IV ćwiartce zbiorowości statystycznej.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 3. Miary zmienności

Pozycyjne: **WSPÓŁCZYNNIK ZMIENNOŚCI**

Gdy średni poziom określamy za pomocą mediany, współczynnik zmienności obliczamy dzieląc odchylenie ćwiartkowe Q przez medianę:

Pozycyjny współczynnik zmienności:

$$V_Q = \frac{Q}{Q_2}\,100 = \frac{Q}{M_e}\,100$$

Pozycyjne miary zmienności wykorzystuje się wówczas, gdy w rozkładzie empirycznym obserwuje się nadmierne zróżnicowanie (np. gdy V jest znacznie większy niż 50%) oraz wtedy gdy rozkład jest bardzo silnie asymetryczny i posiada więcej niż jedną dominantę.

Zarówno klasyczny, jak i pozycyjny współczynnik zmienności stanowią przykład mierników statystycznych względnych, gdyż zgodnie z ich z analitycznymi formułami, określają udział odpowiedniej miary zmienności w określonym mierniku poziomu przeciętnego cechy, tj. średniej arytmetycznej bądź medianie.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 4. Miary asymetrii i kurtozy

WSKAŹNIK I WSPÓŁCZYNNIK ASYMETRII

Szeregiem symetrycznym nazywamy taki szereg statystyczny, w którym liczebności są rozłożone równomiernie (proporcjonalnie) po obu stronach średniej arytmetycznej, wtedy miary poziomu przeciętnego kształtują się następująco:

$$\bar{x} = \mathrm{Me} = \mathrm{D}$$

Szereg, który tego warunku nie spełnia nazywa się *szeregiem asymetrycznym* lub skośnym. Szereg może być *asymetryczny prawostronnie* lub *lewostronnie*. Do oceny kierunku asymetrii wykorzystuje się różnicę między średnią arytmetyczną i dominantą, tzw. miernik skośności:

$\bar{x} - \mathrm{D} = 0$ **szereg symetryczny,**

$\bar{x} - \mathrm{D} < 0$ **szereg o asymetrii ujemnej (lewostronnej),**

$\bar{x} - \mathrm{D} > 0$ **szereg o asymetrii dodatniej (prawostronnej).**

Prawostronność w rozkładzie empirycznym oznacza, że więcej jednostek w badanej zbiorowości statystycznej posiada wariant cechy mniejszy niż wariant średni. Lewostronność w rozkładzie empirycznym oznacza natomiast, że większość jednostek wchodzących w skład badanej zbiorowości statystycznej posiada wariant cechy większy niż średnia arytmetyczna.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 4. Miary asymetrii i kurtozy

WSKAŹNIK I WSPÓŁCZYNNIK ASYMETRII (c.d.)

Współczynnik asymetrii

Z uwagi na małą przydatność bezwzględnych miar skośności, szczególnie przy porównywaniu cech zbiorowości mierzonych za pomocą odmiennych jednostek, do określenia siły i kierunku asymetrii najczęściej wykorzystuje się tzw. współczynnik asymetrii:

$$A_s = \frac{\bar{x} - D}{S} \qquad A_s = \frac{(Q_3 - Me) - (Me - Q_1)}{2Q} = \frac{Q_3 - 2M_e + Q_1}{Q_3 - Q_1}$$

Współczynnik asymetrii najczęściej przyjmuje wartości z przedziału <-1; 1>, jedynie w przypadkach bardzo silnej asymetrii (skrajnej) przekracza te granice.

Znak („+” lub „-”) współczynnika asymetrii świadczy o kierunku asymetrii, a jego wartość bezwzględna o jej natężeniu (sile). Im współczynnik asymetrii jest bliższy (bezwzględnie) jedności, tym silniejsza asymetria. Należy jednak podkreślić, że pierwszy z tych współczynników asymetrii mierzy ją w całym przedziale zmienności wartości cechy, natomiast drugi objaśnia jedynie 50% środkowych jednostek badanej zbiorowości. Fakt ten decyduje o większych walorach poznawczych A_s, który opiera się na średniej arytmetycznej, dominancie i odchyleniu standardowym.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Do klasycznych współczynników asymetrii należy także (bardziej pracochłonny i dlatego rzadziej stosowany) współczynnik asymetrii A oparty na 3-cim momencie centralnym:

$$A = \frac{m_3}{S^3}$$

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3 \qquad m_3 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^3 n_i \qquad m_3 = \frac{1}{n}\sum_{i=1}^{k}(\hat{x}_i - \bar{x})^3 n_i$$

Zadanie

W poniższej tabeli przedstawiono strukturę płac pracowników trzech zakładów produkcyjnych:

Stawka godzinowa (PLN)	Odsetek pracowników		
	Zakład I	Zakład II	Zakład III
10-20	10	5	10
20-30	20	35	25
30-40	40	25	25
40-50	20	25	35
50-60	10	10	5
Razem	100	100	100
$\bar{x}$	35	35	35
S^2 (S)	120 (10,95)	120 (10,95)	120 (10,95)

D_1= 35; D_2= 27,5; D_3= 42,5

Me_1 = 35; Me_2 = 34; Me_3 = 36.

$$A_{s1} = \frac{35-35}{10,95} = 0; A_{s2} = \frac{35-27,5}{10,95} =$$

$$A_{s3} = \frac{35-42,5}{10,95} =$$

W zakładzie II większa część pracowników otrzymuje stawki poniżej przeciętnej, natomiast w zakładzie trzecim większa część pracowników ma stawki powyżej przeciętnej.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Ad. 4. Miary asymetrii i kurtozy

WSPÓŁCZYNNIK KURTOZY

<u>Koncentracja zbiorowości wokół średniej, tzw. kurtoza</u>

Obok miar tendencji centralnej, rozproszenia i skośności, zbiorowość statystyczna może być badana pod kątem koncentracji (skupienia) poszczególnych wartości zmiennej wokół średniej.

Skupienie poszczególnych wartości zmiennej wokół średniej zależy od rozproszenia – im większe jest rozproszenie (zróżnicowanie), tym mniejsza koncentracja, i odwrotnie. *Jednak dwa szeregi charakteryzujące się takim samym lub bardzo podobnym odchyleniem standardowym, czyli o tym samym lub podobnym rozproszeniu, mogą różnić się pod względem koncentracji. Ma to miejsce wówczas, gdy obszar zmienności szeregów jest różny.*

$$K = \frac{m_4}{s^4} \qquad m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4 \qquad m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4 n_i \qquad m_4 = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - \bar{x})^4 n_i$$

Im wyższa wartość współczynnika K, tym bardziej wysmukła jest krzywa liczebności, a zatem większa koncentracja wartości cech wokół średniej. Małe wartości wskazują natomiast na spłaszczenie rozkładu zbiorowości względem badanej cechy.

PRZEGLĄD PODSTAWOWYCH WYBRANYCH CHARAKTERYSTYK LICZBOWYCH ROZKŁADÓW EMPIRYCZNYCH

Przyjmuje się, ze jeśli zbiorowość ma rozkład normalny, to K = 3, bardziej spłaszczony rozkład od normalnego ma K < 3, a bardziej wysmukły K > 3. Z tego względu współczynnik koncentracji podaje się w postaci:

$$K' = \frac{m_4}{s^4} - 3$$