

literatura:

A.D. Aczel – „Statystyka w zarządzaniu”

H. Kassyk –Rokicka – „Statystyka – zbiór zadań”

M. Sobczyk – „Statystyka”

G. Trzpiot, K. Kończak – „Statystyka dla studentów ekonomii”

G. Kończak, G. Trzpiot – „Metody statystyczne z wykorzystaniem programów komputerowych”

K. Zajac – „Zarys metod statystycznych”

Statystyka obejmuje metody zbierania, prezentacji i analizy danych dotyczących zjawisk masowych.

Zadaniem statystyki jest badanie prawidłowości zachodzących w zjawiskach masowych na podstawie badań.

Badania statystyczne:

- planowanie badania;
- obserwacja statystyczna
- opracowanie i prezentacja zebranego materiału statystycznego;
- Opis i wnioskowanie statystyczne.

1. Planowanie badania

Należy określić:

- cel badania
- przedmiot badania
- zakres badania

Przedmiotem badania jest populacja generalna (zbiorowość statystyczna), czyli zbiór przedmiotów lub osób (fizycznych lub prawnych) posiadających wiele cech oraz przynajmniej jedną cechę pozwalającą na rozróżnienie elementów tego zbioru pomiędzy sobą.

Dowolny element populacji nazywamy jednostką statystyczną.

Cele badań:

- 1) Celem badań jest diagnoza kondycji finansowej gospodarstw domowych na terenie Katowic.
- 2) Znajomość języków obcych studentów GWSH.

W pierwszym przykładzie populację stanowi zbiór gospodarstw domowych, w drugim studenci są populacją jak również jednostką statystyczną.

Zakres badań.

Wszystkie badania dzielimy na pełne i częściowe. Pełne, to takie badania, w których badaniu podlega cała populacja. Częściowe – do badania wybieramy podzbiór – próbę statystyczną.

Badania pełne:

- rejestracje

- spisy



duży koszt badań, nie zawsze możliwe do wykonania

Badania częściowe:

- próba statystyczna.

Spisy – badania prowadzone na koszt państwa, określona jest data początku badania, czas trwania, data końca (1988, 2002, planowany 2011)

Rejestracje bieżące:

- na koszt państwa;
- w trybie ciągłym;
- przez odpowiednie urzędy.

Sposoby losowania próby

- 1.) Dobór celowy (sytuacje badawcze, które opisują nowe zjawisko – nowe jednostki statystyczne, metoda monograficzna badań).
- 2.) Dobór losowy podstawowe populacje dzieli się na zespoły (wg cech) potem kolejny podział (wiek, płeć itp.)

Próba reprezentatywna: taka której jednostki statystyczne odzwierciedlają strukturę jednostek populacji. Takie podejście nazywa się metodą reprezentatywną.

Obserwacje statystyczne

- cechy statystyczne
- skale statystyczne
- rodzaj ankiety
- kontrola zebranego materiału

Cechy statystyczne

- cechy stałe: rzeczowe, czasowe, przestrzenne;
- cechy zmienne: jakościowe, ilościowe (skokowe ciągłe).

Cechy stałe są to cechy, które nie podlegają badaniu, są tu wymienione ponieważ służą identyfikacji jednostki statystycznej populacją (pytamy o nie w pierwszej kolejności)

- rzeczowe (kto, co)
- czasowe (kiedy)
- przestrzenne (gdzie)

Cechy zmienne (podlegają badaniu):

- jakościowe - wartości tych cech nie wyrażamy liczbowo (płeć, wykształcenie, miejsce zamieszkania),
- ilościowe – wartości tych cech są liczbami (wiek, wzrost, wielkość obciążeń firmy)
- skokowe – zbiór dyskretu (naturalne)
- ciągłe – liczby rzeczywiste

Skale statystyczne:

- nominalne – stosowane dla cech jakościowych, pozwala na klasyfikację wartości badanych cech statystycznych

*klasyfikacje: nr, wzrost, wiek, kolor oczu, nr dowodu osobistego

- porządkowe – możliwość uporządkowania oraz klasyfikacji wartości badanych cech np. wykształcenie (podstawowe, zawodowe, średnie)

- przedziałowa (interwałowa), przedziały mają określoną długość, brak stałego zera

- stosunkowa (proporcjonalne, ilorazowe), zero absolutne, przedziały mają ustaloną długość. (zero absolutne - 0 długu, zero umowne – 0°)

Rodzaj ankiety. Czy mamy pytania zamknięte, półotwarte, otwarte w ankiecie.

- zamknięte – respondent wybiera jedną z podanych odpowiedzi (nie pozyskujemy nowej wiedzy);

- dla pozyskania nowej informacji, wprowadzanie pytań otwartych, półotwartych.

Kontrola zebranego materiału:

1.) formalne – dotyczy cech stałych

2.) merytoryczne – sprawdzamy odpowiedzi związane cechami zmiennymi

Prezentacje zebranego materiału statystycznego.

- szeregi statystyczne
- tablice statystyczne
- wykresy statystyczne

Szeregi statystyczne

1.) wyliczające/szczegółowe

2.) rozdzielcze/strukturalne

- dla cechy mierzalnej
 - punktowe
 - przedziałowe
- dla cechy niemierzalnej

3.) przestrzenne

4.) czasowe

- momentów
- okresów

=====

Szereg wyliczający czasu dojazdu studentów GWSH w dniu 23.02.2008 (w minutach)
(cechy stałe – kto, gdzie, kiedy; cechy zmienne – czas)

23, 34, 45, 23, 34, 23, 45

Źródło: dane umowne

Sze

reg rozdzielczy punktowy

Wartość cechy

inf. Ile razy dana wartość się powtarza(liczebność) n_i

X_j	n_i
23	3
34	2
45	2

Suma – ile elementów w próbie

Szereg rozdzielczy punktowy

Dane są pogrupowane nie dają dokładnej informacji o wartości

$< x_j , x_{j+1})$	n_i
4 – 6	12
6 – 8	10
8 -10	8
10 -12	6

suma → ilość badań

szereg czasokresów

t_j	y_j
1998	24
1999	25
2000	23
2001	22
2002	23

TABLICA STATYSTYCZNA to więcej niż jeden szereg statystyczny

Wykresy wykonujemy poza układem współrzędnych lub układ współrzędnych.

- diagnoza
- wykres słupkowy

Wykresy statystyczne:

- histogram
- wielobok liczebności
- krzywa liczebności

Opis statystyczny to analizy wyników przeprowadzonych badań

Grupy analiz: analiza zbiorowości, analiza współzależności, analiza zmian zjawisk w czasie.

Dla badań cząstkowych dodatkowo do opisu statystycznego przeprowadzamy wnioskowanie statystyczne .

Wnioskowanie statystyczne to przenoszenie wyników analiz z próby statystycznej na populację z wykorzystaniem elementów rachunku prawdopodobieństwa.

ANALIZA STRUKTURY

Parametry statystyczne są to liczby służące do opisu zbiorowości statystycznej.

Stosowane w analizach parametry dzielimy na grupy miar:

- 1.) miary przeciętne
- 2.) miary zmienności

- 3.) miary asymetrii
- 4.) miary koncentracji

1.) Miary przeciętne

Miary poziomu przeciętnego:

- a) Klasyczne
 - średnia arytmetyczna
 - średnia geometryczna
 - średnia harmoniczna
- b) pozycyjne
 - dominanta (moda)
 - kwantyle
 - kwantyl pierwszy
 - mediana
 - kwantyl trzeci

Średnia arytmetyczna

$$\sum_{i=1}^n x_i$$

Średnia arytmetyczna w szeregu wyliczającym

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

n – wielkość próby
 x_i – wartość badanej cechy w próbie

Średnia arytmetyczna ważona – w szeregu rozdzielczym punktowym

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i n_i$$

X _i	W _i
3	0,05
3,5	0,10
4	0,15
4,5	0,40
5	0,30

$$W_i = \frac{n_i}{n}$$

$$\sum_{i=1}^n w_i = \sum_{i=1}^n \frac{n_i}{n} = \frac{1}{n} \sum_{i=1}^n n_i = \frac{1}{n} n = 1$$

W_i - częstość względna

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k x_i n_i = \sum_{i=1}^k x_i w_i$$

Średnia arytmetyczna ważona – w szeregu rozdzielczym przedziałowym

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k \hat{x}_i n_i$$

n – suma liczebności n_i

k – liczba klas (wierszy) w szeregu

x_i – środek i -tego przedziału

n_i – liczebność i -tego przedziału

Własności średniej arytmetycznej

- 1.) jest to wypadkowa wszystkich obserwacji z próby
- 2.) suma odchyłeń wartości cechy od średniej jest równa zero
- 3.) powiększenie wszystkich wartości w próbie o pewną stałą powiększy średnią arytmetyczną o tę wartość
- 4.) suma wartości zmiennych równa jest iloczynowi średniej arytmetycznej i liczebności próby
- 5.) na poziom średniej silny wpływ mają wartości ekstremalne

ad.2.) $\sum (x_i - \bar{x}) = 0$ - właściwość nr 2 w postaci wzoru

Ad. 4.) ?

DOMINANTA jest to wartość cechy, która w danej próbie występuje najczęściej

23, 34, 45, 23

$\Rightarrow D = 23$

x_i	n_i
11	2
12	5
13	8 D=13
14	3
15	6
x_i	n_i
11	2
12	5
13	4

14	5
15	4

$D_1=12$

szereg bimodalny (dwie dominanty)

- $D_2=14$

x_i	n_i
11	0,2
12	0,2
13	0,2
14	0,2
15	0,2

szereg antymodalny (brak dominanty)

Dominanta w szeregu rozdzielczym przedziałowym

1.) wskazujemy przedział, w którym jest dominanta

2.) wyznaczamy dominantę wykorzystując wzór

$$D = x_k + \frac{n_k - n_{k-1}}{(n_k - n_{k-1}) + (n_k - n_{k+1})} \Delta x_k$$

x – wartość cechy

n– liczebność

k – przedział dominanty

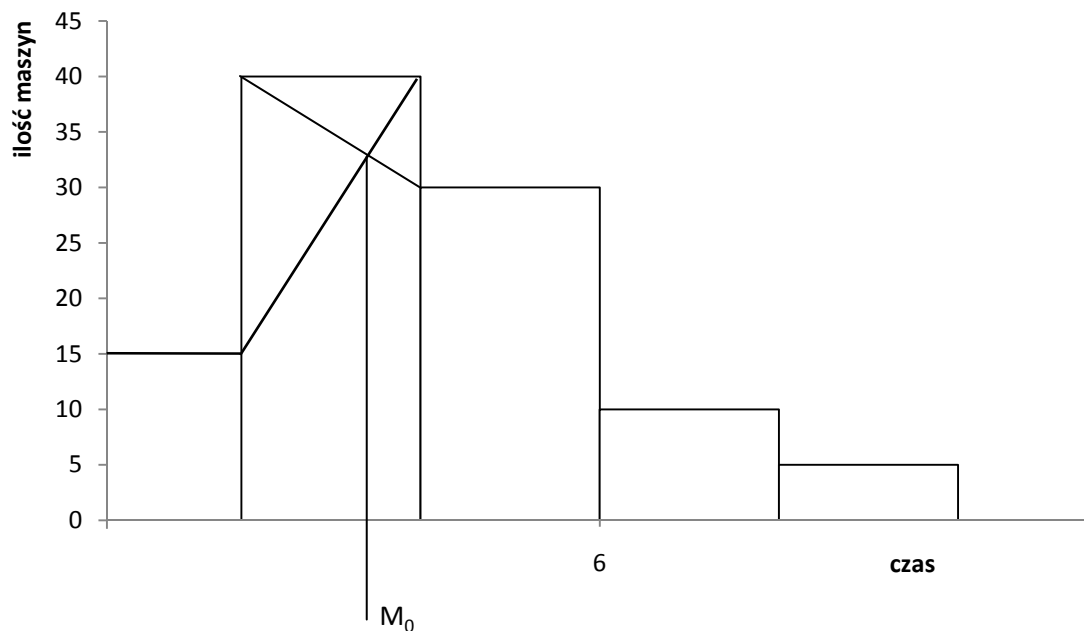
Wyznaczyć dominujący czas eksploatacji maszyn.

Czas (godz.)	liczba maszyn	
<0,2)	15	
<2,4)	40	
<4,6)	30	$D \in <2,4)$
<6,8)	10	
<8,10)	5	

$$D = 2 + \frac{40 - 15}{(40 - 15) + (40 - 30)} \cdot 2 = 3,42[\text{godz}]$$

W badanej próbie czas eksploatacji najczęściej wynosił 3,42 godziny

Graficzna metoda wyznaczania dominanty wykorzystuje histogram rozkładu badanej cechy -



Dominanta w szeregu rozdzielczym przedziałowym

Gdy przedziały w szeregu nie są jednakowej długości celem wyznaczenia dominanty musimy wyznaczyć gęstości

$$g_i = \frac{n_i}{\Delta x_i}$$

n_i –liczebność i-tego przedziału

Δx_i – długość i-tego przedziału

Wyznaczyć wartości dominanty

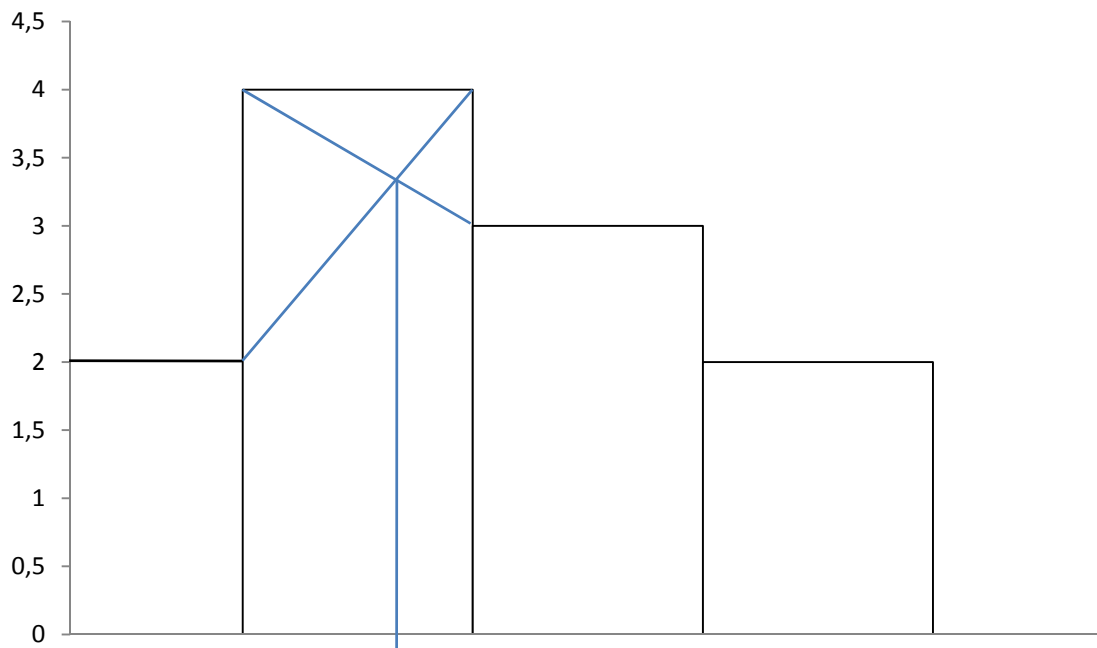
x_i	n_i
0-5	10
5-10	20
10-20	30
20-40	40

$$D = x_k + \frac{g_k + g_{k-1}}{(g_k - g_{k-1}) + (g_k - g_{k+1})} \cdot \Delta x_k$$

x_i	n_i	g_i
0-5	10	2
5-10	20	4
10-20	30	3
20-40	40	2

$$g_i = \frac{n_i}{\Delta x_i}$$

$$D = 5 + \frac{4-2}{(4-2)+(4-3)} \cdot 5 = 8,33$$



W badanej próbie dominowała wartość 8,33

MEDIANA – wartość cechy, która dzieli próbę na dwie części tak, że połowa wartości jest nie większa oraz połowa wartości jest nie mniejsza od mediany (wartość środkowa w próbie).

$$43,56,76,84,12 \Rightarrow Me=76$$

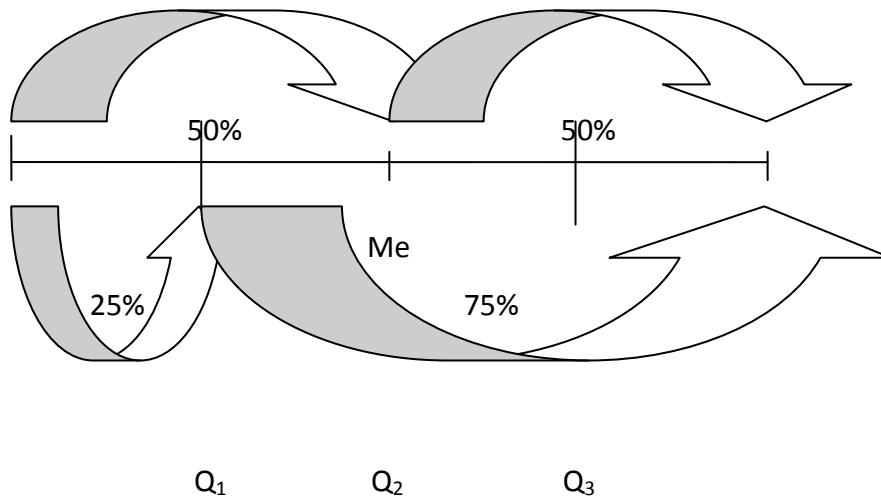
$$43,56,76,84 \Rightarrow Me=(56+76)/2=66$$

$$\text{Próba} \Rightarrow n=123$$

$$Me = \frac{n+1}{2} = \frac{123+1}{2} = 62$$

$$Me = \frac{1}{2} \left(\frac{x_n}{2} + \frac{x_{n+1}}{2} + 1 \right)$$

KWARTYLE



Wartość cechy, która dzieli próbę na dwie części tak, że ¼ wartości jest

Szereg rozdzielczy przedziałowy:

- 1) wyznaczyć szereg, w którym jest kwartył
- 2) obliczamy kwartył

$$Me = x_k + \frac{0,5n - \sum_{i=1}^{k-1} n_i}{n_k} \cdot \Delta x_k$$

x- wartość cechy n- liczebność k- przedział mediany

Q₁ – 0,25 Q₂ – 0,75

Wyznaczyć kwartyle czasu eksploatacji maszyn

czas(godz)	liczba maszyn	n	w _i	
<0,2)	15	15	0,15	0,15
<2,4)	40	55	0,40	0,55
<4,6)	30	85	0,30	0,85
<6,8)	10	95	0,10	0,93
<8,10)	5	100	0,05	1,00

$$Q_1 \in <2,4)$$

$$Q_2 \in <2,4)$$

$$Q_3 \in <4,6)$$

$$Q_b = x_k + \frac{Bn - \sum_{i=1}^{k-1} n_i}{n_k} \cdot \Delta x_k$$

$$Q_1 = 2 + \frac{25-15}{40} \cdot 2 = 2,5n$$

w badanej próbie 25% maszyn nie przekraczało 2,5h

$$Q_2 = 2 + \frac{50-15}{40} \cdot 2 = 3,75$$

$$Q_3 = 4 + \frac{75-55}{30} \cdot 2 = 5,3$$