



SIGMA KWADRAT

CZWARTY LUBELSKI KONKURS STATYSTYCZNO-DEMOGRAFICZNY

Pojęcie i metody badań statystycznych

NBP

Narodowy Bank Polski

PROJEKT DOFINANSOWANY ZE ŚRODKÓW
NARODOWEGO BANKU POLSKIEGO



URZĄD STATYSTYCZNY
W LUBLINIE



WYŻSZA SZKOŁA ZARZĄDZANIA
I ADMINISTRACJI W ZAMOŚCIU



POLSKIE TOWARZYSTWO
STATYSTYCZNE

Bibliografia

Całą prezentację opracowano na podstawie kilku podręczników, przy czym w największym zakresie czerpano z:

- M. Sobczyk. Statystyka. Aspekty praktyczne i teoretyczne, Wydawnictwo UMCS w Lublinie, Lublin 2006
- M. Woźniak, Statystyka ogólna, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 1997

oraz:

- J. Z. Holzer, Demografia, PWE, Warszawa, 2003

Geneza statystyki

Termin statystyka wywodzi się od łacińskiego słowa **status** – co oznacza stan, położenie, stosunki a we współczesnym włoskim **stato** oznacza państwo.

W piśmiennictwie słowo statystyka zostało użyte po raz pierwszy w połowie XVIII wieku przez Gottfrieda Achenwalla (1719-1772) profesora uniwersytetów w Marburgu i Getyndze. Pod pojęciem statystyki G. Achenwall rozumiał zbiór szeroko ujmowanych wiadomości o państwie (warunkach fizjograficznych, ludności, ustroju społecznym, gospodarce itp.). Przedmiot zainteresowań statystyki jako nauki o państwie wyłonił się z potrzeb ówczesnej administracji państwowej, na użytek której zaczęto gromadzić dane liczbowe.

W opracowaniach państwoznawczych z tamtego czasu posługiwano się bogatymi danymi liczbowymi, ujmowanymi w postaci zestawień tabelarycznych. Autorów tego rodzaju tablic nazywano tabelarystami. Za twórcę kierunku tabelarystycznego uważa się Duńczyka J. R. Anchersena.

Geneza statystyki cd.

Dynamiczny rozwój statystyki jako narzędzia wykorzystywanego do opisu zjawisk gospodarczych (tablice bilansowe, warunki bytu ludności itp.) datuje się od czasów A. Quetleta (1796-1874). W okresie tym statystykę zaczęto traktować jako naukę, czego wyrazem było powołanie w 1834 r. w Anglii Królewskiego Towarzystwa Statystycznego oraz zwołanie I Międzynarodowego Kongresu Statystycznego w Brukseli (1854 r.).

Znaczenia statystyki

- jako nazwy zbioru informacji liczbowych dotyczących określonych zjawisk i procesów (urodzeń, rolnictwa, wypadków drogowych itp.);
- jako nazwy wszelkich prac związanych z gromadzeniem i opracowywaniem danych liczbowych;
- jako nazwy pewnych charakterystyk liczbowych opisujących właściwości jednostek tworzących zbiorowości próbne (np. średnia arytmetyczna z próby, odchylenie standardowe z próby itp.);
- jako nazwy dyscypliny naukowej zajmującej się metodami analizy zbiorów danych liczbowych, odnoszących się do powtarzalnych zjawisk masowych lub wyników eksperymentów.

Definicja statystyki

Woźniak M., *Statystyka ogólna*,
Akademia Ekonomiczna w Krakowie, Kraków 1997, s. 11:

Nauka o metodach ilościowych badania zjawisk masowych

Statystyczne metody analizy mają charakter uniwersalny i są wykorzystywane niemal we wszystkich dyscyplinach naukowych (psychologii, socjologii, demografii, historii, medycynie, politologii i oczywiście administracji).

Językiem statystyki nieporównanie dokładniej niż słowami (werbalnie) można opisać różnorodne zjawiska-ilościowe

Korzyści wynikające ze statystycznego sposobu myślenia oraz z zastosowania metod statystycznych

Statystyka:

umożliwia dokładniejszy sposób opisu interesującej nas rzeczywistości, zmusza nas do dokładności i śmiałości w działaniu i rozumowaniu, umożliwia formułowanie uogólnień na podstawie uzyskanych wyników analizy, pozwala na przewidywanie rozwoju zjawisk w przyszłości, czyli na budowanie prognoz, dostarcza narzędzi do porządkowania informacji o zjawiskach - a przez to pozwala na budowę ich ogólnego obrazu, dostarcza narzędzi do prowadzenia analizy przyczyn kształtujących badane zjawiska i procesy, a więc umożliwia dokonanie ich klasyfikacji na czynniki systematyczne i przypadkowe.

Funkcje statystyki

- **informacyjna** (dająca pełny i obiektywny obraz badanych zjawisk),
- **analityczna** (dzięki której możliwe jest określenie czynników kształtujących konkretne procesy i zjawiska)
- **prognostyczna** (pozwalająca na przewidywanie kierunku rozwoju analizowanych zjawisk).

Podział statystyki

Statystyka opisowa, która zajmuje się problemami programowania badań statystycznych oraz metodami obserwacji statystycznej, sposobami opracowywania i prezentacji materiału statystycznego oraz syntetyzującym (sumarycznym) opisem właściwości zbioru danych.

Statystyka matematyczna, która zajmuje się metodami wnioskowania (estymacji i weryfikacji hipotez) o całej zbiorowości generalnej na podstawie zbadania wybranej w sposób losowy pewnej jej części zwanej próbą.

Statystyki branżowe (zwane też stosowanymi), które zajmują się badaniem poszczególnych sfer działalności (np. transportu, handlu, usług, produkcji przemysłowej). Wykorzystuje się tu zarówno statystyczne metody opisu, jak również wnioskowania.

Podstawowe pojęcia statystyczne (1)

Mianem **zbiorowości (populacji) statystycznej** określa się zbiory dowolnych elementów (osób, przedmiotów, faktów) podobnych do siebie pod względem określonych właściwości. Kompletny zbiór elementów nosi nazwę **zbiorowości (populacji) generalnej**. **Indywidualne składowe zbiorowości nazywamy jednostkami statystycznymi lub jednostkami badania.**

Zbiorowości można wyodrębniać według różnych kryteriów. I tak, z względu na kryterium czasu wyróżnia się **zbiorowości statyczne** (np. zbiorowość ludności Polski wg stanu na dzień 31.12.2007 r., na 30.6.2008 r. itp.). Zbiorowości statyczne tworzą więc jednostki, które istniały, istnieją lub będą istniały w ściśle określonym momencie.

Zbiorowości dynamiczne składają się z jednostek obserwowanych w pewnym przedziale czasu. Szereg czasowy.

Podstawowe pojęcia statystyczne (2)

Jednostki statystyczne będące elementami składowymi zbiorowości charakteryzują się określonymi właściwościami, które nazywamy **cechami statystycznymi**. Najogólniej rzecz ujmując, cechy statystyczne można podzielić na stałe i zmienne. Cechy stałe (rzeczowe, przestrzenne i czasowe) są wspólne dla wszystkich jednostek zbiorowości. Taką zbiorowość nazywamy wówczas jednorodną.

Cechy stałe nie podlegają badaniu statystycznemu a jedynie umożliwiają zaliczanie jednostek do określonych zbiorowości. Właściwości, które różnicują jednostki statystyczne między sobą, nazywa **cechami zmiennymi**. Przedmiotem badania statystycznego są zbiorowości składające się z jednostek posiadających jedną lub kilka cech wspólnych (stałych) oraz jedną lub wiele cech (zmiennych) je różnicujących.

Podstawowe pojęcia statystyczne (3)

Cechy, których warianty podawane są w sposób opisowy, nazywamy **cechami niemierzalnymi (jakościowymi)**. Cechy, których warianty są wyrażane za pomocą liczb, określa się mianem **cech mierzalnych (ilościowych)**.

Wśród cech ilościowych wyróżnia się cechy skokowe i ciągłe. Pierwsze z nich przyjmują warianty zmieniające się skokowo (bez wartości pośrednich), w postaci liczb całkowitych (np. ilość braków w ciągu zmiany roboczej). Cechy ilościowe ciągłe mogą przyjmować każdą wartość z różną dokładnością (np. wiek, wzrost, waga).

Rodzaje badań statystycznych

Pod pojęciem badania statystycznego rozumie się zespół czynności zmierzających do uzyskania - za pomocą metod statystycznych - informacji charakteryzujących badaną zbiorowość.

Z punktu widzenia liczby jednostek objętych badaniem, wyróżnia się **badania pełne** (kompletne, całkowite, wyczerpujące) i **badania częściowe** (niekompletne, niepełne).

W badaniach częściowych obserwowane są tylko niektóre wybrane lub wylosowane ze zbiorowości generalnej jednostki.

W badaniach częściowych dobór jednostek do próby może być losowy lub celowy.

Stąd też wśród badań częściowych wyróżnia się badania reprezentacyjne, monograficzne oraz ankietowe.

Badanie reprezentacyjne

Polega na tym, że z całej zbiorowości generalnej **pobiera się próbę losową**, przeprowadza się dokładne badanie jednostek wchodzących w jej skład, a następnie **przenosi się wyniki badania próby na całą populację**. Uogólnianie wyników z badania próby na całą zbiorowość generalną odbywa się przy wykorzystaniu rachunku prawdopodobieństwa i nosi nazwę **wnioskowania statystycznego**.

Przykładem badania reprezentacyjnego jest prowadzone od 1992 r. przez Główny Urząd Statystyczny badanie aktywności ekonomicznej ludności (BAEL).

Przykładem badań częściowych z celowym doбором jednostek są badania monograficzne i ankietowe.

Etapy badania statystycznego

W badaniach statystycznych - niezależnie od ich rodzaju - można wyróżnić **cztery etapy**:

- **przygotowanie** (programowanie),
- **obserwację statystyczną**,
- **opracowanie** i prezentację materiału statystycznego,
- **opis** (lub wnioskowanie) statystyczny.

Programowanie i obserwacje statystyczne (1)

1. **Programowanie** oznacza szczegółowo rozpisaną koncepcję zamierzonego badania.
2. **Obserwacja statystyczna** polega na ustaleniu wartości cech ilościowych lub odmian cech jakościowych u jednostek tworzących badaną zbiorowość. Może to odbywać się poprzez bezpośredni pomiar lub zbieranie informacji.

W przypadku badań pełnych można wykorzystać wyniki spisów powszechnych, rejestracji bieżącej czy sprawozdawczości statystycznej. W przypadku badań częściowych korzysta się z rezultatów uzyskanych metodami: reprezentacyjną, monograficzną czy ankietową.

Programowanie i obserwacje statystyczne (2)

Spis powszechny jest jednorazowym lub cyklicznie powtarzanym badaniem statystycznym, którego celem jest ustalenie wielkości i struktury analizowanego zjawiska w określonym momencie czasu (np. powszechne spisy ludności, spisy gospodarstw rolnych).

Rejestracja bieżąca polega na systematycznym ewidencjonowaniu ściśle określonych faktów (np. urodzeń, zgonów, małżeństw, bezrobotnych), które są przedmiotem badania.

W wyniku rejestracji bieżącej powstają rejestry administracyjne (REGON, PESEL, rejestry ZUS itp.). Stanowią one cenne źródło informacji wykorzystywanych w badaniach statystycznych.

Opracowanie i opis materiału statystycznego

Pochodzący z obserwacji statystycznej materiał liczbowy jest z reguły dość obszerny. Utrudnia to dokonanie analizy statystycznej. Materiał ten należy odpowiednio zaprezentować. Celowi temu służą szeregi statystyczne, tablice statystyczne oraz wykresy statystyczne.

Szeregi statystyczne. Szereg prosty

Szeregiem statystycznym

nazywamy ciąg wartości liczbowych badanej cechy, uporządkowanych według określonych kryteriów (np. rosnąco lub malejąco).

1. Szereg prosty (szczegółowy)

przedstawia materiał statystyczny uporządkowany wyłącznie według wartości badanej cechy. Porządkowanie polega tu na wypisywaniu wartości liczbowych w kolejności rosnącej lub malejącej. Szeregi proste zawierają zwykle wartości powtarzające się.

Osoba	Wzrost (cm)
1	158
2	164
3	175
4	175
5	179
6	195
7	200

Szereg czasowy (dynamiczny)

Podstawą grupowania jest
zmiana badanego
zjawiska w czasie

Lata	Zachorowania na nowotwory w Polsce w tys. osób
1989	82,8
1990	83,4
1991	90,1
1992	94,9
1993	95,6
1994	96,1
1995	97,4

3. Szeregi rozdzielcze. Szereg rozdzielczy cechy niemierzalnej

Szereg rozdzielczy (strukturalny) jest to ciąg wartości liczbowych uporządkowanych według wariantów badanej cechy mierzalnej lub niemierzalnej. Poszczególnym wariantom zmiennej przyporządkowane są odpowiadające im liczebności.

3. 1. Szereg rozdzielczy cechy niemierzalnej jest zestawieniem poszczególnych wariantów danej cechy i odpowiadających im liczebności.

Powiat grodzki	Liczba kin
Biała Podlaska	2
Chełm	2
Lublin	6
Zamość	1

3.2. Szeregi rozdzielcze z cechą ilościową (mierzalną)

Szeregi rozdzielcze przedziałowe dotyczą cechy ciągłej. W tym przypadku należy najpierw ustalić liczbę przedziałów klasowych (klas), ich rozpiętości oraz sposób oznaczania granic przedziałów.

Rozpiętość przedziału (zwana też interwałem)

to różnica między górną a dolną wartością przedziału klasowego.

Rozpiętości przedziałów mogą być równe lub różne. Niektórzy

autorzy proponują ustalać liczbę klas według formuły: $k \approx \sqrt{n}$

– liczebność grupy

Szeregi rozdzielcze przedziałowe mogą mieć zamknięte wszystkie przedziały lub też otwarty pierwszy, ostatni lub pierwszy i ostatni przedział.

Otwarte przedziały stosuje się wówczas, gdy w badanej zbiorowości występują wartości skrajne, znacznie różniące się od pozostałych.

3.2.1.1. Prosty punktowy szereg rozdzielczy z cechą mierzalną

Powierzchnia w km ²	Liczba gmin
50-85	2
85-120	4
120-155	8
155-190	12
190-225	16
225-260	7
260-295	3
Razem	52

3.2.1.2 Skumulowany punktowy szereg rozdzielczy z cechą mierzalną

Skumulowane szeregi rozdzielcze

otrzymujemy w drodze łączenia kolejnych przedziałów klasowych i dodawania odpowiadających im liczebności.

Liczebność skumulowana dla ostatniego przedziału klasowego jest wówczas równa ogólnej liczebności badanej zbiorowości.

Kumulacyjne szeregi rozdzielcze informują o tym, ile jest w badanej zbiorowości statystycznej jednostek przyjmujących co najwyżej określoną wartość cechy.

Gminy województwa według powierzchni w km²)

Powierzchnia w km ²	Liczba gmin	Szereg rozdzielczy skumulowany	
		powierzchnia w km ²	skumulowane liczebności
50-85	2	poniżej 85	2
85-120	4	poniżej 120	6
120-155	8	poniżej 155	14
155-190	12	poniżej 190	26
190-225	16	poniżej 225	42
225-260	7	poniżej 260	49
260-295	3	poniżej 295	52
Razem	52	X	X

3.2.2.1.

Prosty przedziałowy szereg rozdzielczy z cechą mierzalną

Waga	Wzrost						
Granice klas	Granice klas						
	161,8– 165,2	165,3– 168,6	168,7– 172,0	172,1– 175,4	175,5– 178,8	178,9– 182,2	182,3– 185,6
49 – 54	4	2	1	1			1
54,1 – 59	2	9	2			2	
59,1 – 64	2	4	8	3	1		
64,1 – 69			4	9	5	1	3
69,1 – 74	1	1			2	9	5
74,1 – 79		1	1				3
79,1 – 84				1		1	1

Tablice

Jeśli grupowanie przeprowadzone jest według dwóch lub większej liczby cech - to wyniki przedstawione są w formie tablicy.

Tablice statystyczne wykorzystywane są do prezentacji danych uporządkowanych według określonego kryterium.

Tablice proste charakteryzują badaną zbiorowość ze względu na jedną cechę, przy czym może to być zarówno cecha ilościowa, jak i jakościowa. Wynika stąd, że tablica prosta może być utożsamiana z szeregiem statystycznym.

Tablice złożone opisują jedną zbiorowość statystyczną według kilku cech (co najmniej dwóch), dwie lub kilka zbiorowości według dwóch lub kilku cech. Te go rodzaju tablice prezentują więc zespół szeregów statystycznych, a stopień ich złożoności zależy od liczby badanych cech lub liczby zbiorowości (badanych zjawisk).

Wykresy

Wykresy statystyczne są narzędziem analizy i formą rejestracji informacji statystycznych. Każdy wykres składa się z części tekstowej oraz pola wykresu.

Cechę ciągłą można zaprezentować za pomocą takich wykresów jak **histogram, diagram** (wielobok liczebności) i krzywą liczebności (ogiwę).

Histogram jest wykresem złożonym z prostokątów, których podstawy (równe interwałom klasowym) spoczywają na osi odciętych, natomiast wysokości są określone na osi rzędnych przez liczebności (absolutne lub względne). Pola prostokątów tworzących histogram są proporcjonalne do liczebności poszczególnych klas, a łączna powierzchnia wszystkich prostokątów reprezentuje ogólną liczebność badanej zbiorowości.

Diagram można otrzymać z histogramu przez połączenie odcinkami kolejnych środków przedziałów klasowych, reprezentujących punkty środkowe górnych boków poszczególnych prostokątów. Otrzymana w ten sposób linia łamana nosi nazwę diagramu (wieloboku liczebności).

System informacyjny statystyki publicznej

Zasady gromadzenia danych i prowadzenia badań statystycznych reguluje Ustawa o statystyce publicznej z 29 czerwca 1995 r. (Dz. U., Nr 88, poz. 439). Całokształt działań związanych ze zbieraniem informacji statystycznych, ich gromadzeniem, przechowywaniem, opracowywaniem, ogłaszaniem, udostępnianiem i rozpowszechnianiem wyników badań statystycznych jako oficjalnych danych tworzy **System Informacyjny Statystyki Publicznej (SISP)**.

Zadania wynikające z funkcjonowania SISP realizowane są przez służby statystyki publicznej, tzn. **Główny Urząd Statystyczny (GUS)** oraz podległe mu urzędy statystyczne. GUS, utworzony w 1918 r., jest centralnym organem administracji państwowej, którego zadaniem są m.in.: prowadzenie i udostępnianie wyników badań statystycznych, międzynarodowa współpraca w zakresie statystyki, rozwijanie metodologii badań statystycznych oraz ustalanie i aktualizacja podstawowych definicji, kodów, klasyfikacji i nomenklatur (tj. zbioru nazw i terminów używanych w statystyce).

Zadania statystyki publicznej

Do zadań służb statystyki publicznej należy także prowadzenie, w skali krajowej, **dwóch rejestrów**: urzędowego (zwanego też administracyjnym) oraz podmiotów gospodarki narodowej (REGON).

Rejestr urzędowy jest wykazem podmiotów informującym o ich działalności. Jest prowadzony przez sądy i organy administracji publicznej na podstawie aktów wykonawczych do ustaw (np. rejestry podatników, działalności gospodarczej, udzielonych zezwoleń i koncesji, gruntów, budynków).

REGON

REGON to powszechny system identyfikacji i kodowania jednostek gospodarki narodowej.

W rejestrze tym ujmowane są dane o wszystkich osobach prawnych, jednostkach nieposiadających osobowości prawnej, indywidualnych gospodarstwach rolnych oraz osobach fizycznych prowadzących działalność gospodarczą.

Dowodem wpisu do REGON-u jest nadanie określonemu podmiotowi numeru identyfikacyjnego, zawierającego zakodowane informacje o jego cechach.