

Statystyka to nauka, której przedmiotem zainteresowania są metody pozyskiwania i prezentacji, a przede wszystkim analizy danych opisujących zjawiska masowe. Statystyka nie bada pojedynczych przypadków, pojedynczych danych lecz możliwie dużą liczbę informacji, pogrupowanych i zaprezentowanych w taki sposób, by móc wyciągnąć wnioski, które mogą mieć istotne znaczenie dla badanego zjawiska.

W statystyce nie uzyskuje się dobrego rezultatu na podstawie jednej czy nawet kilku informacji. Z punktu widzenia statystycznego nie zagwarantowałoby to, że określona prawidłowość występuje, co oznaczałoby możliwość popełnienia błędów. Ta wiedza numeryczna może być rozpatrywana z dwóch punktów widzenia: opisowego i wnioskowania statystycznego.

Statystyka opisowa:

Zajmuje się opracowaniem danych o zbiorowości czy próbie bez posługiwania się rachunkiem prawdopodobieństwa. Pozwala przy tym na zaprezentowanie danych w sposób uporządkowany, jasny i prosty.

Wnioskowanie statystyczne:

Z punktu widzenia wnioskowania statystycznego statystyka pozwala ustalić prawidłowości i podejmować decyzje na podstawie zredukowanej liczby danych (próby) przy zastosowaniu rachunku prawdopodobieństwa. Dzięki niemu możliwe jest określenie, jaki błąd popełniamy, uogólniając wyniki z próby na całą zbiorowość.

Statystyka jako nauka posługuje się określonymi metodami, które tworzą tzw. badania statystyczne. Celem każdego badania statystycznego jest wyznaczenie wartości wybranych parametrów badanej zbiorowości statystycznej oraz analiza tej zbiorowości. W etapie wstępnym ustalamy przedmiot i zakres badania, tzn. określamy zbiorowość, jednostkę i cechy statystyczne, które będą podlegały badaniu. Następnie wyznaczane są wartości charakterystycznych parametrów i na ich bazie przeprowadza się analizę zbiorowości.

Wybierając metodę badania, uwzględniamy różnorodne czynniki. Najważniejsze z nich to:

- cel badania,
- rodzaj zbiorowości statystycznej,
- tematyka badania,
- stopień szczegółowości badania.

Etapy badania statystycznego:

1. Przygotowanie lub programowanie badania statystycznego (cel, zakres, źródła informacji),
2. Obserwacja statystyczna (m.in. ankieta)
3. Opracowanie surowego materiału statystycznego,

4. Analiza opracowanego materiału statystycznego.

Badanie może mieć charakter:

- pełny - badanie obejmuje całą populację
- częściowy - odbywa się na pewnych (zazwyczaj losowo) wybranych elementach populacji, czyli próbie losowej, zazwyczaj reprezentatywnej dla populacji

Element zbiorowości statystycznej (populacji statystycznej) jest nazywany **jednostką statystyczną**.

Liczba jednostek statystycznych, czyli elementów zbiorowości, jest nazywana **liczebnością zbiorowości**.

Efektem obserwacji statystycznej, która najczęściej przeprowadzana jest za pomocą ankiety, jest surowy materiał statystyczny. Materiał ten należy teraz opracować. Na ten etap badania statystycznego składają się następujące czynności:

- kontrola materiału statystycznego (formalna i merytoryczna),
- porządkowanie i grupowanie materiału statystycznego,
- budowa szeregów i tablic statystycznych (prezentacja tabelaryczna),
- sporządzanie wykresów statystycznych (prezentacja graficzna).

Kwestionariusz ankietowy, czyli podstawowe narzędzie badawcze

Ankieta powinna być tak skonstruowana, aby stwarzała ankietowanemu możliwość formułowania własnych opinii, spostrzeżeń i wniosków. Odpowiedzi ujmowane są częściej werbalnie (opisowo) niż liczbowo. Powinna być w miarę krótka. Jedną z podstawowych zasad stosowanych w badaniach tego rodzaju jest zachowanie anonimowości ankietowanych.

Kwestionariusz składa się z:

- tytułu,
- wyjaśnienia celu badania,
- instrukcji (sposobu wypełnienia ankiety, określenia ilości wyboru odpowiedzi oraz czasu na wypełnienie),
- wyboru takich pytań, aby dokładnie odnaleźć odpowiedź na swoje pytanie,
- metryczki (wiek, płeć, miejsce zamieszkania).

Ze względu na formę w kwestionariuszach ankietowych stosowane są następujące rodzaje pytań:

- pytania otwarte - nie posiadają ustalonego zestawu odpowiedzi. Respondent otrzymuje możliwość swobodnego wypowiedzenia się na dany temat,

- pytania półotwarte – łączą w sobie cechy pytania zamkniętego i otwartego. Respondent dokonuje wyboru tylko jednej odpowiedzi z przewidzianego zestawu. W przypadku gdy żadna z proponowanych odpowiedzi nie oddaje rzeczywistej sytuacji, ma on mimo to możliwość udzielenia odpowiedzi wybierając podpunkt „inne”,
- pytania zamknięte –(pytania skategoryzowane, pytania do wyboru) zawierają pewną liczbę przygotowanych odpowiedzi, spośród których respondent dokonuje wyboru. W pytaniach jednokrotnego wyboru wybiera się tylko jedną odpowiedź, tę, którą uzna za najlepiej oddającą stan jego wiedzy lub sytuację, w której się znajduje. Natomiast pytania wielokrotnego wyboru posiadają ustaloną liczbę wariantów odpowiedzi, z których respondent może wybrać jedną bądź kilka odpowiedzi,
- pytania dychotomiczne proponują respondentowi do wyboru dwie wykluczające się wzajemnie odpowiedzi, zwykle „tak – nie”, bez możliwości udzielenia odpowiedzi neutralnej typu „nie wiem”,
- pytania filtrujące służą autorom badania do podzielenia respondentów na tych, którym na podstawie uzyskanej odpowiedzi będą zadawane określone, dalsze pytania, oraz na tych, których nie dotyczy problematyka poruszana w tych pytaniach.

Statystyka opisowa

I: Prezentacja graficzna opracowanego materiału statystycznego

Dobrze zbudowany szereg lub tablica statystyczna pomaga nam w sposób łatwy i szybki zrozumieć badane zjawisko, jednakże często zdarza się, że to samo zjawisko przedstawione w sposób graficzny staje się bardziej zrozumiałe.

Do prezentacji graficznej opracowanego materiału statystycznego należy:

- wykres słupkowy,
- wykres kołowy,
- wykres obrazkowy (piktogram),
- diagram szeregów czasowych – umieszcza się w układzie współrzędnych. Na osi poziomej (odciętych) umieszcza się skalę czasu, a na osi pionowej (rzędnych) częstości odpowiadające danemu okresowi.
- histogram – umieszcza się również w układzie współrzędnych, oś rzędnych reprezentuje częstości, a oś odciętych reprezentuje różne wartości zmiennej. Zbudowanie histogramu wymaga wykreślenia słupków dla każdego przedziału prezentowanej wartości cechy, których wysokość będzie proporcjonalna do częstości.

- wykresy częstości (diagramy) – polegają na umieszczeniu w układzie współrzędnych częstości na osi rzędnych oraz wartości prezentowanej zmiennej na osi odciętych.
- diagram częstości i procentów skumulowanych – wykreśla się go w układzie współrzędnych na podstawie tablicy częstości skumulowanych, wyznaczając punkt odpowiadający każdej wartości. Następnie łączy się ze sobą punkty. Podobnie można zbudować diagram częstości procentowych. Wtedy na osi pionowej częstości wyrażone będą w procentach.

Kryteria stosowania wykresów:

- by zaprezentować zjawisko z wariantami, czyli zmienną jakościową, często stosuje się wykresy słupkowe, wykresy kołowe i piktogramy (wykresy obrazkowe),
- by przedstawić graficzne szeregi czasowe, stosuje się diagramy szeregów czasowych,
- by zaprezentować zmienne, czyli zjawiska o charakterze ilościowym, wskazane są histogramy i wykresy częstości (diagramy).

Jednakże niezależnie od tych wskazówek, każdy wykres powinien być przejrzysty, tak aby pomógł w zrozumieniu badanego zjawiska. Przejrzystość powinna być najważniejszym kryterium. Każdy wykres powinien zawierać tytuł, legendę (wyjaśnienie użytych znaków czy symboli) oraz źródła informacji.

Statystyka opisowa

II: Miary położenia, zróżnicowania, asymetrii i koncentracji

Najważniejsze miary statystyczne:

- miary położenia, w tym miary tendencji centralnej. Te ostatnie wskazują na średnie wartości pewnej części lub ogółu danych. Ponadto do miar położenia należą percentyle. Ich wartości znajdują zastosowanie w określaniu miar rozproszenia.
- miary rozproszenia (zróżnicowania), które wskazują, czy część lub większość danych jest bardzo skoncentrowana wokół wartości średniej, czy raczej są one bardzo rozproszone.
- miary asymetrii służą do badania stopnia skośności wartości zmiennej
- miary koncentracji.

Miarami tendencji centralnej najczęściej są: średnia arytmetyczna, mediana i dominanta.

Średnia arytmetyczna – inaczej przeciętna. W najprostszy sposób wyznacza się ją, sumując wartości cechy wszystkich obserwacji i dzieląc otrzymaną sumę przez liczbę tych obserwacji.

- średnia harmoniczna (prosta i ważona),

- mediana jest miarą tendencji centralnej. W uporządkowanej z punktu widzenia badanej cechy X zbiorowości (bądź próbie) oznacza ona taką wartość, która jest umieszczona dokładnie w środku. Z tego powodu mediana ma sens wyłącznie wtedy, gdy mamy do czynienia z uporządkowanymi danymi liczbowymi oraz dla przypadku skal największej do najmniejszej lub odwrotnie.
- dominanta (moda) – jest to wartość, która w całej zbiorowości bądź próbie powtarza się najczęściej.

Bezwzględne miary zróżnicowania:

- obszar zmienności - to wartość maksymalna cechy minus wartość minimalna.
- odchylenie przeciętne – średnie odchylenie od średniej arytmetycznej.
- wariancja – średnia arytmetyczna kwadratów różnic wartości cechy i średniej,
- odchylenie standardowe – to pierwiastek kwadratowy z wariancji,
- odchylenie ćwiartkowe.

Względne miary zróżnicowania (rozproszenia)

- współczynnik zmienności.

Miary asymetrii (skośności)

- szereg symetryczny, (asymetria ujemna, asymetria dodatnia),
- współczynnik skośności,
- wszechstronna analiza porównawcza struktury.

Metody indeksowe:

- szereg czasowy,
- średnia chronologiczna,
- przyrost absolutny,
- przyrost względny.

Źródło: Starzyńska Wacława „ Statystyka praktyczna” Wydawnictwo PWN, Warszawa 2000, 2005.

ABC Ankietera Statystyki Publicznej, Urząd Statystyczny w Łodzi, Łódź 2011.