

Statystyka Opisowa z Demografią oraz Biostatystyka

Badania Statystyczne

Aleksander Denisiuk

denisjuk@euh-e.edu.pl

Elbląska Uczelnia Humanistyczno-Ekonomiczna

ul. Lotnicza 2

82-300 Elbląg

Badania Statystyczne

Najnowsza wersja tego dokumentu dostępna jest pod adresem

<http://denisjuk.euh-e.edu.pl/>

Przedmiot statystyki

- nauka o ilościowych metodach zjawisk masowych
 - prawidłowości statystyczne

Podstawowe pojęcia statystyczne

- zbiorowość statystyczna
- jednostki statystyczne
- cechy statystyczne
 - stałe
 - zmienne
 - jakościowe
 - ilościowe
 - skokowe
 - ciągłe
 - pseudo ciągłe
- cechy dwudzielne (dychotomiczne, zero-jedynkowe)

Zbiorowości statystyczne

- jednowymiarowe (jednocechowe)
- wielowymiarowe (wielocechowe)

Pomiar

- przypisanie cechom jednostek symboli
- pomiar bezpośredni
- skale pomiarowe, skalowanie
 - niemetryczne (jakościowe)
 - nominalna
 - porządkowa (rangowa)
 - metryczne
 - przedziałowa (interwałowa)
 - stosunkowa (ilorazowa)

Skala nominalna

- liczby — umowne symbole
- działania: zaliczanie jednostki do kategorii i obliczenie proporcji, odsetek
- porównywane: równe ($=$) lub nie równe (\neq)

Skala porządkowa

- porządkowanie jednostek pod względem natężenia badanej cechy (bez miary odległości)
- uporządkowanie słabe: \leq lub \geq
- uporządkowanie mocne: $<$ lub $>$
- nie wiadomo o ile

Skala przedziałowa

- określenie odległości między jednostkami
- działania: odejmowanie (o ile większy)
- nie można dzielić (ile razy większy)

Skala stosunkowa

- posiadanie naturalnego punktu zerowego (brak cechy)
- można dzielić (ile razy większy)

Skale

- mają kumulatywny charakter
- skale mocniejsze
- transformacja skal

Badanie statystyczne

- dotyczy zbiorowości
- określa prawidłowości dotyczące całej zbiorowości, a nie poszczególnych jednostek
- prawidłowości powinny dotyczyć zmiennych cech

Metody badań

- każde badanie wymaga ustalenia metody
 - badanie *pełne (całkowite, wyczerpujące)*
 - badanie *niepełne (częściowe)*
 - liczna zbiorowość
 - badanie ma charakter niszczący
 - wyniki orientacyjne
 - szacunki *interpolacyjne i ekstrapolacyjne*

Badania

- ciągłe
- okresowe
- doraźne

Badania pełne

- spis statystyczny
- rejestracja bieżąca

Badania częściowe

- badania ankietowe
- badania monograficzne
 - szczegółowy opis i analiza jednostki lub niewielkiego zespołu jednostek
 - wybierana jednostka powinna być typowa
- metoda reprezentacyjna
 - badanie próby, pobranej ze zbiorowości generalnej losowo
 - możliwość oszacowania popełnionego błędu

Szacunek statystyczny

- ustalenie pewnych wielkości (właściwości) nieznannej zbiorowości na podstawie zbiorowości znanej, z nią związanej
 - interpolacja
 - ekstrapolacja

Interpolacja i ekstrapolacja

- liniowa
- nieliniowa

Interpolacja i ekstrapolacja liniowa

- przedsiębiorstwo X w 1992 roku zatrudniało 5000 osób, w 2002 — 8200
- oszacować zatrudnienie w latach 1995 (5960) oraz 2003 (85200)

Interpolacja i ekstrapolacja nieliniowa

- funkcja wykładnicza
- niech L_0 będzie liczbą ludności w okresie wyjściowym, L_n — po upływie n lat, p — roczną stopą przyrostu
 - $L_1 = L_0 + L_0p = L_0(1 + p)$
 - $L_2 = L_1 + L_1p = L_1(1 + p) = L_0(1 + p)(1 + p) = L_0(1 + p)^2$
 - $L_n = L_0(1 + p)^n$
 - $p = \sqrt[n]{\frac{L_n}{L_0}} - 1$

Interpolacja i ekstrapolacja nieliniowa. Przykład

- liczba ludności województwa wynosiła 100 000 osób
- zakłada się, że w okresie dwudziestoletnim średnie roczne tempo wzrostu liczby ludności wyniesie 2%
- określić liczbę ludności w końcu ostatniego roku
 - $L_{20} = 100\,000(1 + 0,02)^{20} = 100\,000(1,02)^{20} = 148595$

Organizacja badań statystycznych

- przygotowanie badania
- obserwacja statystyczna
- opracowanie i prezentacja materiału statystycznego
- opis lub wioskowanie statystyczne

Przygotowanie badania statystycznego

- ustalenie celu i metody badania
- określenie zbiorowości statystycznej i badanych cech
- określenie jednostki statystycznej i sprawozdawczej

Obserwacja statystyczna

- ustalenie wartości cech
- wynik: materiał statystyczny
 - materiał pierwotny
 - materiał wtórny

Materiał statystyczny

- surowy materiał statystyczny
- błędy:
 - systematyczne
 - przypadkowe
- kontrola:
 - formalna (kompletność, pełność, zupełność)
 - merytoryczna:
 - logiczna
 - arytmetyczna

Opracowanie materiału statystycznego

- grupowanie
- zliczanie

Grupowanie

- wyodrębnienie jednorodnych części zbiorowości statystycznej
- przejście od informacji o właściwościach poszczególnych jednostek do informacji o całej zbiorowości (próby)
- zalecane kryteria podziału:
 - jednostki grupy nie powinny być zbyt zróżnicowane
 - liczba grup nie powinna być zbyt duża

Grupowanie. Klasyfikacja

- w badaniach doraźnych grupowanie jest odmienne w każdym badaniu
- w badaniach ciągłych i periodycznych — jednolity system grupowania, *klasyfikacja*

Grupowanie proste i złożone

- podział według jednej cechy — grupowanie *proste*
- podział według kilka niepowiązanych cech — grupowanie *złożone*

Grupowanie typologiczne i wariacyjne

- typologiczne:
 - na podstawie cech jakościowych
- wariacyjne:
 - na podstawie cechy ilościowej

Zliczanie danych

- ręczne
- komputerowe

Szeregi statystyczne

- zbiór wyników obserwacji według pewnej cechy
- nieuporządkowany
- uporządkowany

Szereg szczegółowy

- uporządkowany według wartości badanej cechy
- zazwyczaj według kolejności rosnącej

Szereg rozdzielczy

- uporządkowany według wariantów badanej cechy
- poszczególnym wariantom przyporządkowane odpowiadające im liczebności
- dla cechy mierzalnej warianty określone są
 - punktowo
 - przedziałowo:
 - rozłącznie
 - wyczerpująco

Szereg rozdzielczy przedziałowy

- liczba przedziałów zależy od obszaru zmienności cechy, od liczebności zbiorowości, od celu badania
- niektóre zalecenia:
 - 15–20 przedziałów
 - od 10 do 20
 - $k \leq 5 \log N$
 - $k = 1 + 3,322 \log N$
- rozpiętość przedziału (długość, interwał, rozstęp przedziałowy):
 - $i \approx \frac{x_{\max} - x_{\min}}{k}$ lub $i \approx \frac{x_{\max} - x_{\min}}{1 + 3,322 \log N}$
 - zazwyczaj interwały powinny być jednakowe
 - czasami uzasadnione są nierówne interwały

Szereg rozdzielczy przedziałowy

- dwa sposoby oznaczenia granic interwałów:
 1. 2–4, 4–6, 6–8, etc
 2. 2–4, 5–7, 8–9, etc
- przedziały otwarte:
 - mniej niż 2, 2–4, etc
 - 2–4, 5–7, 8 i więcej

Szeregi przestrzenne (geograficzne)

- według jednostek administracyjnych

Szeregi dynamiczne (czasowe, chronologiczne)

- prezentują rozwój jednostek w czasie

Tablice statystyczne

- prezentują dane uporządkowane według określonego kryterium
 - tytuł, słowne oznaczenia kolumn i wierszy, źródło danych, dodatkowe objaśnienia
- znaki umowne
 - zjawisko nie występuje
 - 0 zjawisko występuje w ilościach mniejszych od liczb, które mogą być wyrażone uwzględnionymi znakami cyfrowymi
 - . zupełny brak informacji (brak informacji wiarygodnych)
 - × wypełnienie pozycji jest niemożliwe lub niecelowe
 - w tym** oznacza, że nie podaje się wszystkich składników sumy
- tablice proste
- tablice złożone

Wykresy

- rodzaje wykresów

Opis statystyczny

- czwarty etap
- statystyka opisowa
- statystyka matematyczna