ZJAZD 2

CKU KROŚNIEWICE 05-07.02.2021

EKA.05 FILIA KROŚNIEWICE

**ELEMENTY STATYSTYKI OPISOWEJ**

Miary tendencji centralnej pozwalają opisać przeciętny poziom badanego zjawiska, jednak nie zwierają informacji o zróżnicowaniu jednostek statystycznych punktu widzenia badanej cechy.

Przykład:

W miejscowości K funkcjonują dwa targowiska. Na każdym z targowisk jest po 11 stoisk. W dniu 18 czerwca 2016r. ceny za 1 kg truskawek na każdym ze stoisk obu targowisk kształtowały się następująco.

Targowisko I:

Stoisko 1 – 3,00 zł. Stoisko 2 – 4,00 zł, stoisko 3 – 5,00 zł, stoisko 4 – 6,00 zł, stoisko 5 – 6,00 zł, stoisko 6 – 6,00 zł, stoisko 7 – 7,00 zł, stoisko 8 – 7,00 zł. Stoisko 9 – 8,00 zł, stoisko 10 – 8 zł, stoisko 11 – 9,00 zł.

Targowisko II:

Stoisko 1 – 6,00 zł. Stoisko 2 – 6,00 zł, stoisko 3 – 6,00 zł, stoisko 4 – 6,00 zł, stoisko 5 – 6,00 zł, stoisko 6 – 6,00 zł, stoisko 7 – 7,00 zł, stoisko 8 – 7,00 zł. Stoisko 9 – 7,00 zł, stoisko 10 – 7 zł, stoisko 11 – 7,00 zł.

Jeżeli dla obu targowisk policzone zostaną miary tendencji centralnej, to okaże się, że mają one taka samą wartość: średnia cena truskawek na każdym z targowisk wynosiła 6,90 zł., najczęściej występującą ceną truskawek (dominantą) była cena 6,00 zł, taką samą wartość (6,00 zł) miała również mediana.

Mim, że miary tendencji centralnej na obu targowiskach były takie same , to łatwo można zauważyć, że oba targowiska były zdecydowanie odmienne z punktu widzenia zróżnicowania cen. Na pierwszym targowisku najtańsze truskawki można było kupić po 3 złote za kg, były tez truskawki po 4,5,6,7,8 i 9 złotych za kilogram. Natomiast na drugim targowisku można było kupić truskawki tylko po 6 lub 7 zł za kg. Osoba, która chciałaby kupić truskawki np. po 4 lub 5 zł za kg, nie znalazłaby na tym targowisku żadnego stoiska, na którym truskawki miałyby taką cenę.

Zbadanie zróżnicowania cechy w badanej zbiorowości statystycznej jest możliwe dzięki przeprowadzeniu analizy rozproszenia. W analizie rozproszenia wykorzystywane są miary rozproszenia, które mogą być nazywane również miarami zmienności, miarami zróżnicowania, miarami rozrzutu lub miarami dyspersji. Miary te można obliczyć tylko wtedy, gdy badana cecha statystyczna jest cecha mierzalną.

Rozproszenie

Co to jest rozproszenie Zmienność lub rozproszenie wyników jest zjawiskiem bardzo prostym. Polega ona na tym, że jednostki populacji według wartości pewnej zmiennej różnią między sobą. Można to zilustrować przykładami zaprezentowanymi w poniższej tabeli.

Tabela 1. Rozproszenie wartości zmiennych

|  |  |  |
| --- | --- | --- |
| Zmienna | Populacja | Wartości |
| Wiek | dyrektorzy szkół gimnazjalnych | 32, 37, 38, 40, 44, 45, 49, 53, 55 |
| Płeć | nauczyciele | 5 mężczyzn i 9 kobiet |
| Stopień wykształcenia | zatrudnieni w WOM | 4 z wykształceniem zawodowym, 6 z wykształceniem średnim, 28 z wyższym |
| Staż pracy | studenci na studiach zaocznych | 1, 2, 5, 6, 10, 11, 16, 21 |
| Liczba programów | Szkoły lotnicze | 4, 7, 2, 4, 5, 9, 3, 5, 5 |

W wszystkich przedstawionych przypadkach pojawia się rozproszenie (zmienność). Zmienności nie byłoby, gdyby na przykład wszyscy dyrektorzy byli w tym samym wieku, wszyscy nauczyciele byli tej samej płci, wszyscy zatrudnieni w WOM posiadali ten sam stopień wykształcenia, wszyscy studenci mieli taki sam staż pracy, a szkoły lotnicze taką samą liczbę programów. Z powyższej tabeli wynika, iż w zmiennej „stopień wykształcenia” nie wszystkie jednostki różnią się od siebie. Pomimo to w grupie istnieje zmienność. Wystarczy bowiem, że chociaż jedna jednostka różni się od pozostałych. Podobnie jest z drugą i piątą zmienną. W pierwszej i czwartej zmiennej natomiast wszystkie jednostki się różnią. Skąd pochodzi rozproszenie Dlaczego jednostki różnią się od siebie?

Dlaczego wszystkie wartości nie są jednakowe? Dlaczego więc pojawia się rozproszenie? Najłatwiejszym jest drugie pytanie, na które odpowiedź brzmi: w pewnej grupie wszystkie jednostki mogą być jednakowe (bardziej precyzyjnie: wartości pewnej zmiennej mogą być u wszystkich jednostek identyczne). Zmienność w takiej grupie jest zerowa, co oznacza, że jej nie ma. Rodzi się, więc pytanie: dlaczego wszystkie przypadki nie są takie same? Analiza o jakiejkolwiek konkretnej zmiennej łatwo prowadzi do poprawnej odpowiedzi. Wszystkie noworodki byłyby tej samej płci, gdyby nic nie miało wpływu na ich płeć. Wszyscy uczniowie posiadaliby taką samą ocenę z pewnego przedmiotu, jeżeli nic nie miałoby wpływu na te oceny. Przyczyną zmienności są więc wpływy. Oczywistym jest fakt, że na przykład na oceny ma wpływ motywacja (bardziej zmotywowani uczniowie na ogół posiadają wyższe oceny), warunki do uczenia się w domu (im lepsze warunki, tym lepsze oceny), czas uczenia się i wiele innych czynników. Zmienność pewnej zmiennej jest zawsze skutkiem wpływu innych zmiennych. Fakt ten otwiera od razu szereg nowych pytań. Pytanie pierwsze: ile jest czynników, które mają wpływ? Na ogół jest to ogromna ilość. Pedagogika nie jest w stanie odkryć wszystkich czynników, a tym bardziej ich zbadać. Nawet w najprostszych zjawiskach edukacyjnych nie da się odkryć wszystkich czynników. Zwykle udaje się wyjaśnić tylko najważniejsze (a nawet i te jedynie częściowo!). Nie oznacza to jednak wcale, iż badanie i wyjaśnianie sedna zjawisk jest bezsensowne i nieprowadzące do nikąd. Każdy nawet częściowo opisany i wyjaśniony wpływ dostarcza nowej wiedzy pedagogicznej. Powstaje kolejne pytanie: czy jeżeli w pewnej populacji nie ma zmienności, to oznacza, iż na tę zmienną nic nie ma wpływu? Odpowiedź brzmi: nie! Być może wpływy istnieją, lecz ich efekty zostały zrównoważone (suma pozytywnych efektów równa się sumie negatywnych efektów). Takie przypadki w praktyce zdarzają się tylko incydentalnie. Oznacza to jednocześnie, że bardzo rzadkie są przypadki, w których pewna populacja posiada zerowe rozproszenie. Zagadnienie to można zilustrować na przykładzie płci: pomimo tego, że profesja nauczycielska jest wyraźnie sfeminizowana, tylko incydentalnie można znaleźć szkołę, w której nie pracuje nawet jeden nauczyciel. W mniejszych populacjach to zdarza się częściej (im mniejsze populacje, tym częściej). Nauczycielami pierwszej klasy szkoły podstawowej często są same kobiety, jednak wśród nauczycieli pierwszych klas w całym województwie można już znaleźć kilku mężczyzn. Wyjaśnia to potrzebę badania zmienności: głównie po to, aby zbadać wpływ innych czynników. Jest to powód, dla którego zmienność lub rozproszenie jest centralnym zjawiskiem statystyki. Zmienność stanowi fundament większości metod statystycznych.

Jak zmierzyć rozproszenie

Rozproszenie to różnice między jednostkami. Wszystkie sposoby pomiaru rozproszenia oparte są na tych różnicach. Ze względu na fakt, że rozproszenie jest środkiem badania współzależności między zjawiskami, prawie nigdy nie oblicza się go bezpośrednio. Najczęściej pomiar rozproszenia jest częścią innych procedur statystycznych. Rozproszenie zmiennych nominalnych Przy zmiennych nominalnych, których wartości (kategorie) nie podlegają stopniowaniu, można ustalić jedynie to. czy jednostki różnią się od siebie, czy też nie. W takich przypadkach nie można dokonywać pomiaru wielkości różnic, lecz tylko ich istnienia. Pomiar rozproszenia dokonuje się poprzez ustalanie ilości (zliczenie) jednostek, które różnią się od siebie. Sposób takiego pomiaru rozproszenia ilustrują poniższe przykłady. Pierwszym z nich jest najprostszy przypadek rozproszenia według płci.

Przykład 1:

W sali wykładowej znajduje się 45 studentów i nie ma żadnej studentki. Jednostki tej populacji nie różnią się między sobą, czyli rozproszenie jest równe zeru.

Przykład 2:

W sali wykładowej znajduje się 44 studentów i jedna studentka. Różni się ona od pozostałych 44 jednostek płcią, co oznacza, że w tej populacji istnieje określone rozproszenie (niewielkie). Identyczne rozproszenie istniałoby w przypadku, gdyby w sali wykładowej znajdowały się 44 studentki i tylko jeden student. W przykładzie nr 2 tylko jedna jednostka różni się od pozostałych. Oznacza to najmniejsze możliwe rozproszenie. Im więcej różnych jednostek, tym większe rozproszenie. W przypadku płci największe możliwe rozproszenie występuje wtedy, gdy połowa jednostek w grupie jest jednej płci, a polowa - drugiej płci. W przypadku zmiennych, które posiadają więcej niż dwie kategorie, sytuacja jest nieco odmienna, lecz w istocie taka sama: im więcej różnych jednostek, tym większe rozproszenie. Nie ma potrzeby obciążać Czytelnika przykładami z trzema, czterema i kolejnymi kategoriami zmiennych. Najprostszą okazuje się sytuacja, w której ilość kategorii jest większa niż ilość jednostek w grupie. Największe możliwe rozproszenie występuje wtedy, gdy wszystkie jednostki różnią się miedzy sobą. Ilustruje to przykład nr 3.

Przykład 3.

W sali wykładowej znajduje się 27 słuchaczy i każdy z nich jest innej narodowości. Jest to przykład największego rozproszenia. Najprostszym sposobem dokonywania porównań pomiędzy populacjami według rozproszenia jest porównywanie procentów poszczególnych kategorii. Nie istnieje jednak specjalna miara rozproszenia zmiennych nominalnych.

Rozproszenie zmiennych porządkowych Wartości zmiennych porządkowych podlegają stopniowaniu. Stopnie tych zmiennych nie narastają równomiernie (przedziały pomiędzy sąsiadującymi stopniami nie są jednakowe), dlatego tez różnice pomiędzy jednostkami są niemożliwe do zmierzenia. Jako przykład można zakładać, że pomiędzy wykształceniem średnim a licencjackim istnieje większa różnica niż pomiędzy wykształceniem licencjackim a magisterskim. Trudno natomiast określić, jak wielkie są te różnice (lub czy w ogóle pierwsza różnica jest większa od drugiej). Dlatego też na ogół w takich przypadkach stosuje się takie samo rozwiązanie, jak przy zmiennych nominalnych - pomiaru rozproszenia dokonuje się poprzez ustalanie ilości jednostek, które różnią się od siebie. Zmienne porządkowe posiadają jednak podstawę ilościową, szkoda zatem zupełnie pominąć wielkości różnic między ich poszczególnymi stopniami. Pomysł uwzględnienia wymienionych różnic jest ciekawy, lecz bardzo trudny do realizacji. Rozwiązanie to prowadziłoby do nie zwykle skomplikowanych miar rozproszenia, a jednocześnie trudnych w interpretacji. Dlatego też w praktyce stosuje się dwie następujące procedury:1. zmienne porządkowe traktuje się najczęściej tak samo, jak zmienne nominalne - rozproszenie mierzy się poprzez ustalenie, ile różnych wartości pojawia się w grupie,2. niekiedy zmienne porządkowe traktuje się jak zmienne przedziałowe - wówczas dla obydwu stosuje się identyczną procedurę obliczania rozproszenia (patrz: następny podrozdział).Identyczne traktowanie zmiennych porządkowych oraz przedziałowych dotyczy tylko procedury obliczenia miary rozproszenia, nigdy natomiast interpretacji otrzymanych rezultatów. Przy interpretacji wyników należy uwzględniać naturę zmiennych porządkowych. Oznacza to tym samym, że w interpretacji rezultatów badacz musi być bardzo ostrożny.

Rozproszenie zmiennych przedziałowych pod rozdział ten traktuje o zmiennych przedziałowych. Jednak wszystkie treści w nim zawarte będą odnosiły się także do zmiennych ilorazowych. W pomiarze rozproszenia zmienne ilorazowe nie są bardziej przydatne niż zmienne przedziałowe. Podstawową właściwością przy pomiarze rozproszenia jest identyczność przedziałów, którą posiadają obydwa rodzaje skal. Dla zmiennych przedziałowych można mierzyć różnice między wartościami (bo istnieje jednostka pomiaru). Dlatego też pomiar rozproszenia zmiennych przedziałowych oparty jest na wielkości różnic, a nie tylko na ich liczebności. Dla zmiennych przedziałowych używa się kilku miar rozproszenia.

Oto najbardziej znane:

•obszar zmienności (rozstęp),

•odchylenie ćwiartkowe,

•odchylenie przeciętne,

•wariancja,

•odchylenie standardowe.