ANALIZA STRUKTURY

Statystyczna analiza struktury jest opisem grupy badanej pod względem mierzonych cech na podstawie parametrów statystycznych, miar statystycznych. W celu dokładnego zbadania poziomu danej cechy, zmiennej w badanej przez nas grupie korzysta się z szeregu rodzaju miar statystycznych, podzielonych na grupy: miary położenia np. średnia, mediana, modalna

- miary zmienności np. odchylenie standardowe, wariancja
- miary asymetrii np. skośność
- miary położenia np. kurtoza

Głównym celem analizy struktury jest dokładne przedstawienie pod względem statystycznym badanej zbiorowości. Analiza struktury jest tu analizą opisową, jej miary nie służą testowaniu statystycznemu (testy statystyczne), lecz opisie zjawiska w badanej grupie. Za pomocą takich miar jak: średnia arytmetyczna, średnia harmoniczna, średnia geometryczna, modalna, kwartyle (w tym mediana), odchylenie standardowe, odchylenie przeciętne, odchylenie ćwiartkowe, wariancja, współczynnik zmienności, rozstęp, skośność i kurtoza można dokładnie określić badaną grupę pod względem danej cechy (zmiennej). Oczywiście niektóre miary nie mogą być wykorzystane dla danego typu zmiennej, cechy. Takie miary klasyczne jak np. średnia, odchylenie standardowe mogą być wykorzystane gdy zmienna ma charakter ilościowy, w przypadku porządkowego charakteru zmiennej należy skorzystać z takich miar pozycyjnych jak np. mediana, odchylenie ćwiartkowe.

Innymi słowy analiza struktury to charakterystyka badanej grupy osób pod względem danej cechy. Miary statystyczne wykorzystywane w analizie struktury służą do porównywania dwóch grup pod względem danej cechy. Dzięki tym miarom, korzystając z odpowiednich testów można przeprowadzać testowanie statystyczne.

ŚREDNIA ARYTMETYCZNA

Średnia arytmetyczna jest najpopularniejszą statystyką należącą do grupy statystyk opisowych. Jest najbardziej znanym pojęciem statystycznym.

Wzór na średnią arytmetyczną dla dwóch liczb:

m2=a+b2m2=a+b2

Wzór na średnią arytmetyczną dla n liczb:

mn=a1+a2+...+annmn=a1+a2+...+ann

Przykład: średnia liczb 1, 2 i 4

m=1+2+43=73≈2,333m=1+2+43=73≈2,333

Wyjaśnienie symboli:

mm - średnia arytmetyczna

a,b,a1,a2...a,b,a1,a2...- liczby z których liczymy średnią

nn - ilość liczb, z których liczymy średnią

Posiadając zbiór obserwacji, pochodzący np. z odpowiedzi 100 osób nie przedstawiamy wyniku dla każdej z nich z osobna.

Podajemy za to jedną wartość która opisuje niejako całą naszą przebadaną grupę. Tą wartością właśnie jest średnia.

Wartość średnia pochodzi z sumowania poszczególnych wyników i podzielenie tej sumy przez liczbę naszych obserwacji.

Przykład 1:
Zapytano pięć osób, ile razy ciągu ostatniego tygodnia poszukiwali potrzebnych im informacji w internecie. Uzyskano następujące odpowiedzi:

1, 4, 3, 0 i 2 razy w ciągu ostatniego tygodnia.
I tak:
(1 + 4 + 3 + 0 + 2) / 5 = 2

Wiemy zatem, iż w naszej grupie osób w ostatnim tygodniu potrzebnych informacji wyszukiwano średnio 2 razy.

Możemy zatem zauważyć, sumujemy nasze wyniki, dzielimy przez liczbę tych wyników (obserwacji) i otrzymujemy średnią arytmetyczną

Przykład 2:
Sprawdzono ile wynoszą średnie zarobki netto w pewnym dziale firmy X. Pracują tam 4 osoby (n = 4). Ich zarobki wynoszą odpowiednio: 2000, 2500, 3000, 5000. Sumując zarobki wszystkich osób otrzymujemy wynik równy 12500, dzieląc tę liczbę przez 4 otrzymujemy średnią równą 3125.
Średnia arytmetyczna - właściwości:

- średnia obrazuje nam przeciętny wynik w badanej próbie
  - zmiana jakiejkolwiek wartości w zbiorze na inną sprawia, że średnia zmienia swoją wartość
  - średnia arytmetyczna stanowi estymator wartości oczekiwanej w populacji

- średnia arytmetyczna jest wrażliwa na przypadki odstające w zbiorze - przeczytaj więcej: Mediana a średnia arytmetyczna
- dla zmiennych ilościowych stanowi ona podstawową i najważniejszą statystykę opisową w przedstawianym raporcie z badań, informującą o poziomie tej zmiennej dla badanej grupy

Oprócz średniej arytmetycznej występują inne odmiany średniej, np:

- średnia ważona
- średnia geometryczna
- średnia harmoniczna
- średnia ruchoma
- średnia potęgowa

ŚREDNIA HARMONICZNA

Wzór na średnią harmoniczną dla n liczb ma postać:

$h_n=n1a_1+1a_2+....+1a_n$hn=n1a1+1a2+....+1an

Wzór na średnią harmoniczną dla dwóch liczb ma postać:

$h_2=21a+1b$h2=21a+1b

Przykład: średnia liczb 1, 2 i 4:

$h_n=311+12+14=127\approx1,714$hn=311+12+14=127≈1,714

Wyjaśnienie symboli:

hh - średnia harmoniczna

a,b,a1,a2...a,b,a1,a2...- liczby z których liczymy średnią

nn - ilość liczb, z których liczymy średnią

MEDIANA

Mediana jest jedną z najpopularniejszych miar centralnych w statystyce. Mediana zwana jest również jako wartość środkowa zbioru. Mediana dzieli wszystkie nasze obserwacje na dwie równe co do ilości obserwacji grupy (w teorii) - wyniki niższe niż mediana i wyniki wyższe niż mediana. Inaczej mówiąc wartość mediany wskazuje nam, że połowa naszych wyników ma wartość poniżej wartości mediany, a druga połowa ma wartość powyżej wartości mediany.

Przykład:

Uczniowie klasy Vc (15 osób) ustawili się w rzędzie od najniższego do najwyżego. Nauczyciel zmierzył wzrost wszystkim osobom

139, 141, 142, 147, 148, 149, 149, 150, 152, 153, 153, 155, 158, 159, 161

Mógł zatem stwierdzić, że połowa osób ma wzrost nie większy niż 150cm (mediana = 150), a połowa klasy ma wzrost nie mniejszy niż 150 osób. 8 osoba w szeregu stanowiła wartość srodkową zbioru.

(1,2,3,4,5,6,7,**8**,9,10,11,12,13,14,15)

Taka sama liczba osób była poniżej tej wartości, jak również powyżej tej wartości: po 7 osób. W naszym przykładzie 8, osoba z kolei stanowi wartość środkową, czyli medianę. Aby wyliczyć medianę należy wszystkie nasze obserwacje uporządkować od wartości najniższej do najwyższej i wyznaczyć punkt środkowy. Można oczywiście posiłkować się następującym wzorem:

Me (symbol mediany) = (n + 1)/2.

n oznacza liczbę obserwacji

Jednak ten wzór dotyczy TYLKO przypadku, gdy mamy w zbiorze nieparzystą liczbę obserwacji.

Problem pojawia się w sytuacji, gdy mam parzystą liczbę obserwacji. Musimy obliczyć średnią arytmetyczną pomiędzy dwiema środkowymi wartościami.

Przykład:

1, 2, 3, **4, 5,** 6, 7, 8.

W naszym zbiorze nie ma wartości środkowej, albo inaczej rozumując mamy je aż dwie. Dlatego wyliczyć należy średnią arytmetyczną z tych dwóch wartości. W przykładzie będzie to 4,5 ponieważ (4+5)/2 = 4,5.

Kolejnym ważnym aspektem, na który należy zwrócić uwagę to fakt, że gdy mamy nieparzystą liczbę osób powstaje pytanie, czy ta obserwacja, która równa jest medianie należy do pierwszej, czy do drugiej połówki zbioru? I w takim przypadku statystycy umówili się, że taką obserwację będą zaliczać do zbioru poniżej mediany, dodając w opisie tej grupy: wartości niższe bądź równe medianie. Naturalnie tworzy się wtedy minimalna nierównoliczność grup, aczkolwiek jest ona znikoma. Oczywiście odwrotny sposób też jest dopuszczalny, ale należy zastosować inny, odpowiedni komentarz.

MEDIANA

- Mediana równoważna jest drugiemu kwartylowi, są to zamienne nazwy, drugi kwartyl i mediana dzielą zbiór obserwacji na połowy

- Mediana z próby jest zgodnym i nieobciążonym (asymptotycznie) estymatorem wartości oczekiwanej w populacji

- Mediana jest odporna na przypadki odstające w zebranej przez nas próbie - mediana a średnia

- Mediana wykorzystywana jest jako statystyka opisowa dla zmiennych o charakterze porządkowym, jej wartość określa średni poziom danej zmiennej (z racji, że dla zmiennych o takich charakterze średnia arytmetyczna nie występuje)

- **Przykład:**

- Zbadano poziom wykształcenia 20 badanych osób. 3 osoby miały wykształcenie podstawowe, 2 wykształcenie gimnazjalne, 6 wykształcenie zawodowe, 6 wykształcenie średnia a 3 wykształcenie wyższe. Uszeregowano osoby w kolejności od najniższego do najwyższego wykształcenia i sprawdzono jaki poziom wykształcenia mają osoby znajdujące się w "środku". Były to osoby z wykształceniem zawodowym. Można zatem powiedzieć, że mediana wykształcenia badanych osób wyniosła: wykształcenie zawodowe. Przy czym, w opisanym przykładzie nie moglibyśmy obliczyć poziomu średniej, ponieważ nie mamy wartości liczbowych, są tylko wartości porządkowe.

Średnia dotyczy się liczb, a mediana może dotyczyć się kategorii, pod warunkiem, że możemy je uszeregować w jakiejś kolejności.

DOMINATA

Dominanta należy do grupy miar centralnych. Opisuje ona wartość najczęściej występującą w zbiorze danych. Dominantę również nazywa się modalną lub modą, czyli "najmodniejszą wartością".

Przykład:
Uzyskaliśmy następujące dane:

0,1,3,4,4,4,5,5,6,8,9

W tym zbiorze dominanta wynosi 4, dlatego, że pojawia się ona najczęściej dla danej zmiennej. Jest ona najbardziej "popularną" wartością.

Może się jednak pojawić sytuacja następująca:

0,0,2,3,4,4,4,5,5,6,6,6,7

W tej sytuacji mamy dwie najczęściej występujące wartości: 4 i 6. W takim przypadku podaje się dwie wartości, ponieważ zmienna ma dwie dominanty. Czasem jednak można się spotkać z sytuacją, że podaje się wartość o najniższej wartości, informując przy tym, że istnieje więcej wartości modalnych.

Dominanta (modalna) ma szczególne zastosowanie przy zmiennych nominalnych. Jako, że dla tych zmiennych nie można wyznaczyć średniej ani mediany dominanta stanowi jedną z podstawowych miar opisowych.

Przykład:

Ogrodnik po dniu pracy zliczył zebrane owoce: 100 jabłek, 60 śliwek, 20 gruszek i 120 truskawek.

Nie można obliczyć zarówno średniej ani mediany, ale można podać, że najczęściej zbieranym owocem była truskawka.

Czasem można spotkać się z pojęciem zmienna dwumodalna, bądź rozkład dwumodalny.