

Zadania ze statystyki dla I roku socjologii

1. Dane poniżej prezentują pomiary dochodu. Uporządkuj poniższe dane w szeregu strukturalnym prostym, od wartości najmniejszych do największych. Następnie utwórz szereg przedziałowy z powyższych danych, przyjmując 5 klas o równych szerokościach przedziałów. Oblicz procentowy udział w poszczególnych klasach i szereg skumulowany.
UWAGA: szerokość przedziału ustalamy obliczając iloraz różnicy wartości maksymalnej i wartości minimalnej, dzieląc go przez liczbę klas (5). Następnie zliczamy wystąpienia obserwacji wpadających do każdej z tych klas.

1234	2890	974
2345	356	2085
5467	3790	5207
3679	3890	3419
4800	19001	4540
1910	7904	1650
987	5100	727
3020	4280	2760
4100	5790	3840
2101	8010	1841

2. Oblicz średnią wartość dochodu w badanej zbiorowości z szeregu prostego oraz szeregu przedziałowego.
3.). Oblicz średni dochód w obrębie 10 najgorzej zarabiających (10 obserwacji o najniższych dochodach) i 10 najlepiej zarabiających (10 obserwacji o najwyższych dochodach). Oblicz stosunek dochodów najlepiej zarabiających do najgorzej zarabiających).
4. Utwórz nowy szereg przedziałowy o dwóch kategoriach (dobrze zarabiający i słabo zarabiający) przyjmując za punkt podziału wartość przeciętną dochodów (z punktu 2, dla całości rozkładu). Następnie oblicz stosunek dochodów dobrze zarabiających do słabo zarabiających.
5. Tabele poniżej prezentują rozkłady częstości wielkości gospodarstw domowych (x_{i1} i x_{i2}) społeczności dwóch miasteczek. Należy obliczyć średnią wielkość gospodarstwa domowego oraz medianę wielkości gospodarstwa domowego w obu tych miasteczkach.
Należy również porównać ze sobą te wartości, wskazując, która zbiorowość charakteryzuje się przeciętnie większym gospodarstwem domowym.

Xi1	ni1
1	14
2	22
3	26
4	30
5	21
6	12
7	9
8	7
Σ	141

Xi2	ni2
1	12
2	30
3	25
4	28
5	21
6	12
7	9
8	7
	144

6. Uporządkowano dochód mieszkańców jednego z tych miasteczek w szeregu przedziałowym. Należy obliczyć przeciętny dochód oraz medianę dochodów mieszkańców tego miasteczka.

dochód	liczebność
0 - 999	20
1000 - 1999	21
2000 - 2999	15
3000 - 3999	11
4000 - 4999	7
Σ	74

7. Dochód mieszkańców drugiego z miasteczek został uporządkowany nieco inaczej, co prezentuje tabela poniżej. Proszę obliczyć średni dochód mieszkańców tego miasteczka, wiedząc, że osoba o najwyższym dochodzie zarabia 8200.

0-2000	40
2001 - 4000	31
4001 - 6000	16
powyżej 6000	4

8. Średni dochód osób w zbiorowości X wynosi 3800 zł a mediana dochodu 2800, zaś w zbiorowości Y średni dochód wynosi 2300 a mediana 2800 zł. Obie zbiorowości są sobie równe ze względu na liczbę mieszkańców. Która z tych zbiorowości posiada mniej zróżnicowane dochody i dlaczego.

PODSUMOWANIE WYKŁADU

Cechy statystyczne dzielimy na mierzalne i niemierzalne:

- **Mierzalne** – dające się wyrazić liczbowo (np. wzrost, wiek, dochód)
- **Niemierzalne (jakościowe)** – nie dające się wyrazić liczbowo (np. pochodzenie narodowe, miejsce zamieszkania, płeć)

Poziomy pomiaru:

1. Poziom nominalny
2. Poziom porządkowy
3. Poziom interwałowy
4. Poziom ilorazowy
5. Poziom absolutny

Poziom nominalny:

- Najniższy poziom pomiaru
- Charakteryzuje się przypisaniem jednostki analizy do określonej kategorii zmiennej (cechy)
- Kategorie opisują właściwości danej zmiennej
- Podział na kategorie musi być rozłączny i wyczerpujący
- Przykład cechy: płeć (kobiety i mężczyźni), narodowości (np. Polacy, Niemcy, Rosjanie)

Poziom porządkowy:

- Wyższy od nominalnego poziom pomiaru.
- Pomiar na tym poziomie polega na przypisaniu jednostek analizy do kategorii zmiennych, które pozostają względem siebie w pewnym porządku (kolejności).
- Kategorie zmiennej tworzą układ hierarchiczny, wyznaczający miejsce danej kategorii w tym układzie
- Miejsce kategorii w układzie oznacza jakiś potencjalny stopień nasycenia cechy (kontinuum natężenia cechy)
- Przykład cechy: wykształcenie (podstawowe, średnie, wyższe)

Poziom interwałowy:

- Wyższy od porządkowego poziom pomiaru
- Każdej jednostce analizy przypisywana jest wartość liczbowa, wyrażająca natężenie danej cechy statystycznej
- Cecha ma stopniowalny charakter, gdzie każdy stopień (interwał) wyraża w sposób liczbowy jej stopień
- Różnice między natężeniem danej cechy można wyrazić liczbowo, jako odległość między nimi i tylko w taki sposób (o ile dane jednostki różnią się od siebie)
- Punkt początkowy (0) wyznacza się arbitralnie
- Przykład cech: Skala temperatury Celsjusza

Poziom ilorazowy i absolutny

- Najwyższy poziom pomiaru
- Każdej jednostce analizy przypisywana jest wartość liczbowa, wyrażająca natężenie danej cechy statystycznej
- Cecha ma stopniowalny charakter, lecz nie sposób w inny niż arbitralny sposób wyznaczyć stopnia natężenia cechy

- Punkt początkowy ma naturalny charakter (0 jako brak natężenia cechy)
- Różnice mogą być wyrażone proporcjami (jednostka różni się od innej o określoną proporcję razy – ileś tam razy)

Mierniki tendencji centralnej:

- Klasyczne: średnia arytmetyczna, średnia arytmetyczna ważona, średnia ważona dla szeregów przedziałowych
- Pozycyjne: mediana, dominanta

Średnia arytmetyczna

Dla szeregu punktowego:

$$\bar{x} = \frac{\sum x_i}{N}$$

Średnia arytmetyczna ważona

Dla szeregu rozdzielczego:

$$\bar{x} = \frac{\sum x_i n_i}{N}$$

Średnia arytmetyczna

Dla szeregu przedziałowego:

$$\bar{x} = \frac{\sum x_i^o n_i}{N}$$

Mediana:

1. Pozycja mediany:

$$poz.Me = \frac{N + 1}{2}$$

2. Mediana dla N nieparzystego:

$$Me = x_n$$

3. Mediana dla N parzystego:

$$Me = \frac{(x_{n-1} + x_{n+1})}{2}$$

4. Mediana dla szeregu przedziałowego:

$$Me = X_0 + \left(\frac{N}{2} - \sum nk \right) \frac{C}{n_0}$$