Lista 3. Twierdzenia graniczne, statystyka opisowa

Zad. 1 Zmienne losowe X_1, X_2,…, X_{100} są niezależne o jednakowym rozkładzie Poissona z parametrem λ=2. Obliczyć przybliżoną wartość wyrażenia $Pr\left(\sum_{k=1}^{100} x_k \geq 190\right)$.

Zad.2. W centrali telefonicznej znajduje się *n* linii działających niezależnie. Prawdopodobieństwo, że dowolna linia jest zajęta, jest równe 0,1. Jakie powinno być *n*, aby prawdopodobieństwo tego, że co najmniej 7 **procent** linii jest zajętych było równe 0,9?

Zad.3. Prawdopodobieństwo, że w czasie *T* przestanie świecić jedna żarówka jest równe 0,1. Obliczyć prawdopodobieństwo, że w czasie *T* spośród 100 przestanie świecić od 7 do 15 żarówek przy założeniu, że żarówki przepalają się niezależnie.

Zad. 4. Zbadano pojemność elektryczną 20płyt z ceramiki tytanianu baru i otrzymano następujące wyniki (w pF)

11,0	9,2	9,9	12,0	8,0	8,7	7,1	11,8	11,7	10,3
11,2	8,1	9,5	11,5	11,6	9,6	10,2	11,4	8,6	10,0

Narysować dystrybuantę **empiryczną**

Zad.5. W IV klasie szkoły podstawowej nr 123 dokonano pomiaru poziomu inteligencji uczniów. Wyniki pomiarów były następujące: 127, 129, 130, 132, 134, 135, 135, 135, 135, 140, 142, 143, 143, 144, 145, 145, 149, 152, 152, 153. Zanalizuj szereg przedstawiający poziom inteligencji dzieci. W tym celu wyznacz następujące wielkości:

a. wartość średniej arytmetycznej,
b. wartość dominanty i mediany,
c. wartość rozstępu, wariancji i odchylenia standardowego,
d. wartość współczynnika zmienności oceniając stopień zróżnicowania poziomu inteligencji dzieci w klasie,
e. wartość odchylenia przeciętnego jako alternatywnej miary, na podstawie której można ocenić stopień zróżnicowania populacji.

Zad.6. W pewnym mieście zebrano następujące informacje o wielkości zatrudnienia w pewnej liczbie małych firm prywatnych: 1, 1, 2, 2, 3, 4, 5, 5, 6, 6, 6, 7, 7, 7, 7, 8, 9, 9, 10, 10, 12, 12, 12, 12, 13, 14, 14, 14, 15, 15, 15, 15, 15, 15, 16, 18, 20, 23, 25, 30. Dokonaj agregacji danych przez zbudowanie szeregu rozdzielczego. Proszę przyjąć następujące granice klas: (0,5], (5,10], (10, 15], (15, 20], (20,25], (25, 30].

Zad.7. Dokonaj statystycznej analizy szeregu rozdzielczego obrazującego rozkład zatrudnienia z zadania 6. W tym celu:

f. ustal wielkość przeciętnego zatrudnienia w firmie,
g. wyznacz dominantę i medianę dla tej zbiorowości,
h. określ stopień zróżnicowania poszczególnych elementów w całej zbiorowości,
i. sprawdź asymetrię rozkładu,
j. pokaż, jak można zanalizować stopień skoncentrowania rozkładu.

Zad. 8. Narysować obok siebie histogramy oraz wyznaczyć średnią arytmetyczną i cztery pierwsze momenty centralne dla czterech szeregów rozdzielczych. Dodatkowo obliczyć współczynniki: asymetrii (współczynnik skośności) $g_a = \frac{C_3}{s^3}$, koncentracji (współczynnik skupienia zwany też kurtozą) $K = \frac{C_4}{s^4}$, ekces (współczynnik spłaszczenia) $g_e = K - 3$, współczynnik zmienności $v = \frac{S}{\overline{X}} 100\%$.

Środki klas $\bar{x}_i$	Liczności n_i			
	szereg I	szereg II	szereg III	szereg IV
1	0	2	0	2
2	6	2	2	4
3	12	10	20	10
4	14	22	12	12
5	12	10	10	20
6	6	2	4	2
7	0	2	2	0

Zad.9. Urząd statystyczny dostarczył informacji o 100 firmach, które płacą roczne karty za zanieczyszczenie środowiska.

Wysokość płaconych kar na rzecz miasta	**Liczb firm płacących karę w przedziale**
poniżej 10 000	12
[10 000, 30 000)	13
[30 000, 50 000)	13
[50 000, 80 000)	20
[80 000, 120 000)	34
[120 000, 150 000)	6
powyżej 150 000	2
Razem	**100**

Dokonaj statystycznej analizę szeregu rozdzielczego przedstawiającą strukturę firm według wysokości płaconych kar. Celem tej analizy powinno być:

k. wyznaczenie mediany wysokości kar,
l. wyznaczenie kwartyli – pierwszego i trzeciego,
m. obliczenie odchylenia ćwiartkowego,
n. obliczenie wartości współczynnika zmienności (pozycyjnego),
o. określenie asymetrii rozkładu.

Zad. 10. Kierowca z Krakowa odwoził swoją rodzinę na wczasy nad morze pokonując trasę 600 kilometrów z prędkością 75 km/h. Po przybyciu nad morze natychmiast wyruszył w drogę powrotną do Krakowa. Ze względu na zmęczenie w drodze powrotnej osiągnął prędkość 37,5 km/h. Oblicz przeciętna prędkość kierowcy w ciągu całej podróży.

Zad. 11. Pięć kul o promieniach r_1=1,2 cm, r_2=1,4 cm, r_3=1,7 cm, r_4=2,1 cm, r_5=2,7 cm zastąp pięcioma jednakowymi kulami o promieniu r w ten sposób, by suma objętości wszystkich kul była taka sama w obu przypadkach. Przedstaw ogólną zależność *r* od r_i.

Zad. 12 W pewnym mieście zebrano następujące informacje o wielkości zatrudnienia w pewnej liczbie małych firm prywatnych: 1, 1, 2, 2, 3, 4, 5, 5, 6, 6, 6, 7, 7, 7, 7, 8, 9, 9, 10, 10, 12, 12, 12, 12, 13, 14, 14, 14, 15, 15, 15, 15, 15, 15, 16, 18, 20, 23, 25, 30. Dokonaj agregacji danych przez zbudowanie szeregu rozdzielczego. Narysuj histogram.

Wskazówki
W zad. 1. należy skorzystać z tw. Lindberga-Levy'ego.
W zad. 2-3 skorzystać z tw. Moivre'a Laplace'a
Elementy statytsyku opisowej są dobrze wyjaśnione w drugiej części książki
[Krysicki] W.Krysicki, J.Bartos, W.Dyczka, K.Królikowska, M.Wasilewska, „Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach", część II, PWN, Warszawa
(w książce tej są stosowane inne oznaczenia niż w poniższych wzorach).

Zadania 10-12 zostaną ewentualnie rozwiązane, jeżeli wystarczy czasu

Tab. 1 Momenty

	Moment rzędu k (zwykły)	Moment centralny rzędu k
Momenty teoretyczne	$m_k = EX^k$	$c_k = (X - EX)^k$
Momenty empiryczne	$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$	$C_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$, gdzie $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
Momenty empiryczne (dane pogrupowane w klasy) U - liczba klas, $\mathring{n}_i$ - liczność i-tej klasy, $\mathring{X}_i$ - środek i-tej klasy.	$M_k = \frac{1}{n}\sum_{i=1}^{U} \mathring{X}_i^k n_i$	$C_k = \frac{1}{n}\sum_{i=1}^{U}(\mathring{X}_i - \overline{X})^k n_i$ gdzie $\overline{X} = \frac{1}{n}\sum_{i=1}^{U} \mathring{X}_i n_i$.

$X_{1,} X_{2,} \dots, X_n$ dane (z próby losowej prostej)

Wariancja $VarX = E(X - EX)^2 = EX^2 - (EX)^2$

Wariancja empiryczna $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^{2,}$, gdzie $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

Wariancja empiryczna poprawiona $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^{2,}$ (w [Krysicki] oznaczona S z gwiazdką)

Wariancja empiryczna (dane pogrupowane w klasy) $S^2 = \frac{1}{n}\sum_{i=1}^{U}(\mathring{X}_i - \overline{X})^2 n_i$, gdzie $\overline{X} = \frac{1}{n}\sum_{i=1}^{U} \mathring{X}_i n_i$.

Dystrybuanta empiryczna jest funkcją określoną wzorem:

$$F_n(x) = \frac{1}{n}\left|\{i : X_i < x\}\right|,$$

gdzie:
n - liczba danych
$|\{i : X_i < x\}|$ - liczba danych X_i mniejszych niż x (linie $|\ |$ oznaczają liczność zbioru).

Poniższe wzory należy wykorzystać w zadaniach 7 i 9 (proszę się ich nie uczyć na pamięć)
Mediana dla szeregu rozdzielczego:

$$m_e = x_{0m} + \frac{b}{n_m}\left(\frac{n}{2} - \sum_{i=1}^{m-1} n_i\right),$$

gdzie:
b - długość klasy,
m - numer klasy zawierającej medianę,
n - liczba wszystkich danych
n_i - liczność i-tej klasy,
n_m - liczność klasy zawierającej medianę,
x_{0m} - lewy koniec klasy zawierającej medianę.

Moda (wartość modalna, dominanta, wartość najczęstsza) w szeregu rozdzielczym:

$$m_O = x_{0d} + \frac{n_d - n_{d-1}}{(n_d - n_{d-1}) + (n_d - n_{d+1})} b,$$

gdzie:
b - długość klasy,
d - numer klasy zawierającej medianę,
n_d - liczność klasy zawierającej modalnej,
n_{d-1}, n_{d+1} - liczności sąsiednich klas,
x_{0d} - lewy koniec klasy zawierającej dominantę.

Pierwszy i trzeci kwartyl dla szeregu rozdzielczego:

$$Q_1 = x_{01} + \frac{b_{Q1}}{n_{q1}}\left(\frac{n}{4} - \sum_{i=1}^{q_1-1} n_i\right),$$

$$Q_3 = x_{03} + \frac{b_{Q3}}{n_{q3}}\left(\frac{3n}{4} - \sum_{i=1}^{q_3-1} n_i\right),$$

gdzie:
$b_{Q1,} b_{Q3}$ - szerokość klas z I i III kw. (kw. oznacza kwartyl dla danych niezagregowanych w klasy),
$q_{1,} q_3$ - numery klas zawierającej I i III kw.,
n - liczba wszystkich danych
n_i - liczność i-tej klasy,
$n_{q1,} n_{q3}$ - liczność klas zawierających I i III kw.,
x_{01} - lewy koniec klasy z I kw.