Statystyka Stosowana *Lista 5*

1. Suma opadów (w mm) w Warszawie w lipcu w kolejnych latach poczynając od roku 1811 do roku 1960 wynosiła: 35, 82, 48, 75, 77, 123, 117, 75, 92, 101, 116, 113, 42, 44, 36, 71, 9, 74, 114, 49, 83, 94, 223, 28, 57, 46, 33, 86, 85, 74, 72, 104, 37, 229, 41, 50, 73, 40, 76, 100, 171, 41, 160, 120, 144, 46, 143, 105, 29, 92, 138, 44, 26, 80, 50, 84, 78, 74, 53, 51, 76, 30, 48, 6, 54, 63, 20, 74, 81, 45, 50, 174, 82, 18, 139, 31, 47, 78, 173, 71, 72, 20, 85, 19, 35, 39, 120, 92, 172, 98, 37, 77, 143, 26, 96, 13, 132, 109, 116, 132, 37, 32, 91, 101, 77, 87, 99, 181, 166, 68, 5, 122, 33, 84, 66, 64, 149, 23, 20, 115, 71, 108, 55, 166, 124, 115, 53, 71, 49, 73, 93, 76, 113, 53, 77, 37, 78, 124, 84, 44, 68, 26, 65, 136, 154, 82, 88, 38, 80, 159. Obliczyć średnią, wariancję, medianę i rozstęp międzykwartylowy. Sporządzić histogram i wykres ramkowy. Wyznaczyć średnią obciętą odrzucając po 15% skrajnych wyników. Ocenić własności rozkładu sumy opadów (jednomodalność, skośność, spłaszczenie). (Koronacki i Mielniczuk, zad. 1.3, str. 57)

2. Badanie długości pracy T (w godzinach) 200 lamp dało następujące wyniki:

Nr grupy i	Granice grupy $[x_{i-1}, x_i)$	Liczebność grupy n_i
1	[0, 300)	53
2	[300, 600)	41
3	[600, 900)	30
4	[900, 1200)	22
5	[1200, 1500)	16
6	[1500, 1800)	12
7	[1800, 2100)	9
8	[2100, 2400)	7
9	[2400, 2700)	5
10	[2700, 3000)	3
11	[3000, 3300)	2
12	[3300, A)	0

(a) Narysować histogram. (b) Porównać go z wykresem gęstości rozkładu wykładniczego z parametrem $\lambda = 0,00115$. (c) Dla kilku wybranych przedziałów porównać części wypadania obserwacji do przedziału z prawdopodobieństwem liczonym przy założeniu gęstości teoretycznej. (d) Obliczyć empiryczną wartość oczekiwaną i wariancję i porównać z wartościami teoretycznymi.

3. Roczny dochód Z pewnego przedsiębiorstwa ma rozkład Pareto o gęstości $f(x) = 3(x+1)^{-4}$ dla $x > 0$ i zero poza tym (w mln. zł). Obliczyć $P(Z > \frac{1}{2})$ oraz EZ i $VarZ$. Wygenerować próbę 30 - elementową z tego rozkładu, obliczyć częstość zdarzenia $\{Z > \frac{1}{2}\}$, średnią oraz wariancję z próby i porównać wyniki z wartościami dokładnymi (teoretycznymi).

4. Wygenerować 400 liczb z rozkładu jednostajnego na odcinku (0, 1). (a) Narysować histogram i obliczyć średnią, medianę, wariancję oraz kwartyl dolny i górny z próby oraz porównać je z rzeczywistymi wartościami tych parametrów. (b) Następnie pogrupować wygenerowane liczby na 40 ciągów 10 - elementowych i obliczyć w każdej grupie średnią. Powtórzyć punkt (a) dla czterdziestu uzyskanych średnich. Do gęstości jakiego rozkładu (na mocy centralnego twierdzenia granicznego) będzie bliski histogram dla średnich?

5. 64% studentów zdaje egzaminy w pierwszym terminie. Obliczyć prawdopodobieństwo tego, że w grupie 5000 studentów w pierwszym terminie zda od 3000 do 3500 studentów. Użyć centralnego twierdzenia granicznego.

6. W pewnej loterii wygrana W wynosi 1, 2 i 10 złotych z prawdopodobieństwem $p_1 = 0.4$, $p_2 = 0.2$ oraz $p_{10} = 0.1$. W pozostałych wypadkach los jest pusty. Podać symulowane wartości wygranych dla 36 losów oraz częstość wygrania 1 zł. Stosując centralne twierdzenie graniczne, obliczyć prawdopodobieństwo tego, że łączna wygrana 36 losów znajdzie się w granicach 50 do 80 zł. Jaka jest łączna wygrana dla wygenerowanej próby?

7. Tygodniowe wypłaty z pewnego funduszu są niezależnymi zmiennymi losowymi o rozkładzie wykładniczym z tym samym parametrem $\lambda = \frac{1}{1000 \text{ zł}}$. Obliczyć prawdopodobieństwo, że łączna wypłata z tego funduszu w okresie roku, tzn. 52 tygodni, przekroczy 70 000 zł.

8. W celu oszacowania dokładności wskazań pewnego przyrządu zmierzono wielkość wzorcową 1000 razy. Niech X_i, i = 1, ..., 1000 oznaczają błędy wskazań. Jako oszacowania σ^2 przyjęto $\overline{\sigma}^2 = \frac{1}{1000}(X_1^2 + ... + X_{1000}^2)$. Obliczyć prawdopodobieństwo tego, że błąd oszacowania tj. $|\overline{\sigma}^2 - \sigma^2|$ nie przekroczy 5% rzeczywistej wartości σ^2. Przyjąć $EX_1^4 = 2\sigma^4$.