7. TESTOWANIE ZGODNOŚCI

ZADANIE 7.1 Losową próbę studentów spytano o ich ulubiony przedmiot. Otrzymano następujące odpowiedzi:

Przedmiot	Fizyka	WF	Mechanika	Statystyka
Liczba studentów, którzy najbardziej lubią ten przedmiot	380	340	380	500

Na poziomie istotności 0,05 sprawdzić hipotezę, że rozkład preferencji jest równomierny.

ZADANIE 7.2 Zbadano grupę krwi 100 osób. Grupę 0 miało 36 osób, A - 42 osoby, B - 14 osób i grupę AB - 8 osób. Zweryfikować hipotezę, że prawdopodobieństwa wystąpienia grup krwi 0, A, B, AB w populacji są równe odpowiednio: 0.4; 0.4; 0.1; 0.1. Przyjąć poziom istotności 0,05.

ZADANIE 7.3 Aby zaliczyć programowanie Maciek musi napisać program generujący liczby losowe z rozkładu dwumianowego o parametrach 3 i 0,5. Co więcej, Maciek musi wykazać, że jego program pracuje prawidłowo. W tym celu nasz bohater wygenerował 200 liczb i otrzymał następujące wyniki:

Wygenerowana liczba losowa	0	1	2	3
Liczba uzyskanych wyników	24	73	77	26

Czy na poziomie istotności 0,05 można stwierdzić, że generator Maćka działa prawidłowo?
WSKAZÓWKA: Jeśli X ma rozkład dwumianowy $bin(n, p)$, gdzie $n \in \{0, 1, \ldots\}$, $p \in (0, 1)$, to

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ dla } k = 0, 1, \ldots, n.$$

Ponadto funkcja `dbinom(x=k,size=n,prob=p)` podaje prawdopodobieństwo $P(X = k)$.

ZADANIE 7.4 Naukowiec chce sprawdzić czy liczba cząstek emitowanych przez pewną substancję promieniotwórczą w ciągu 10-ciu sekund jest zmienną losową o rozkładzie Poissona. W tym celu zbadał liczbę cząstek, które zostały wyemitowane przez tą substancję w ciągu dziesięciosekundowych odcinków czasu i zebrane dane zapisał w poniższej tabelce:

liczba wyemitowanych cząstek w ciągu 10-ciu sekund	0	1	2	3	4	5
liczba przypadków, kiedy zostało wyemitowanych tyle cząstek	140	280	235	200	100	45

.

Jakie wnioski wyciągnie naukowiec na poziomie istotności 0,1?

ZADANIE 7.5 Posługując się pakietem R i ustalając ziarno generatora równe 4411

```
> set.seed(4411)
```

wygenerować 200 liczb z rozkładu wykładniczego o parametrze $\lambda = 2$ (użyć funkcji `rexp()`). Następnie, na poziomie istotności 0,05, sprawdzić, używając testu zgodności chi-kwadrat, czy liczby te rzeczywiście pochodzą z rozkładu wykładniczego.
WSKAZÓWKA: Aby przeprowadzić test zgodności chi-kwadrat musimy najpierw dane zdyskretyzować, rozdzielając je do odpowiedniej liczby klas. Wiemy, że pożądanym jest aby prawdopodobieństwa klas p_j^0 były przynajmniej w przybliżeniu równe i by był spełniony warunek, że wszystkie $np_j^0 \geqslant 5$. Zdecydujmy się na równe prawdopodobieństwa wszystkich klas i wynoszące $\frac{1}{20}$; zagwarantuje to, że $np_j^0 = 10 \geqslant 5$. Zatem chcemy mieć 20 klas. Jeśli za końce klas (skoro chcemy 20 klas, to potrzebujemy 21 punktów końcowych) przyjmiemy odpowiednie kwantyle rozkładu wykładniczego z parametrem λ, wyszacowanym metotą największej wiarygodności,

```
> konce.przedzialow=qexp(seq(0,1,length.out=21),estymator.lambdy)
```

to dla każdej klasy rzeczywiście będziemy mieć $p_j^0 = \frac{1}{20}$. Pozostaje zliczyć ile obserwacji wpadło do poszczególnych klas. Możemy to zrobić używając funkcji `cut()` i `table()`:

```
> licznosci.klas=table(cut(x=probka,breaks=konce.przedzialow))
```

ZADANIE 7.6 Używając testu Kołmogorowa-Smirnowa, sprawdzić czy liczby wygenerowane w zadaniu 7.5 rzeczywiście pochodzą z rozkładu wykładniczego o parametrze $\lambda = 2$. Przyjąć poziom istotności $\alpha = 0,01$.

ZADANIE 7.7 W kolumnie *czas* w pliku *infolinia.txt* zapisano czasy oczekiwania (w min.) na połączenia z pewną infolinią. Używając testu Kołmogorowa-Smirnowa, sprawdzić czy można uznać, że prezentowane czasy pochodzą z rozkładu gamma $Gamma(a, \beta)$ z parametrem kształtu $a = 4,5$ i drugim parametrem $\beta = 4$. Przyjąć poziom istotności $\alpha = 0,05$.

ZADANIE 7.8 Używając testu zgodności chi-kwadrat Pearsona sprawdzić czy czasy oczekiwania, o których mowa w poprzednim zadaniu, pochodzą z

(a) rozkładu gamma,

(b) rozkładu gamma $Gamma(a, \beta)$ z parametrem kształtu $a = 4,5$ i drugim parametrem $\beta = 4$.

Przyjąć poziom istotności 0,05.

ZADANIE 7.9
(a) Wygenerować po $N = 1000$ liczb z następujących rozkładów:
(a1) normalnego o średniej = 20 i odchyleniu standardowym = 5,
(a2) jednostajnego na przedziale $(-1, 1)$,
(a3) wykładniczego o średniej = 5,
(a4) Poissona o średniej = 3.
WSKAZÓWKA. Użyć funkcji: (a1) `rnorm` (a2) `runif` (a3) `rexp` (a4) `rpois`

(b) Dla każdej wygenerowanej próbki sporządzić wykres normalności i wykresy te wyświetlić w jednym oknie. Przeanalizować ich kształt.

(c) Dla każdej wygenerowanej próbki sporządzić wykres skrzynkowy i wykresy te wyświetlić w jednym oknie. Przeanalizować ich kształt.

(d) Dla każdej wygenerowanej próbki sporządzić histogram częstości i nanieść na niego jądrowy estymator gęstości. Przeanalizować kształty tych wykresów.

(e) Dla danych wygenerowanych w pkt. (a1) i (a2) przeprowadzić, na poziomie istotności $0,05$, test normalności Shapiro-Wilka.

(f) Punkty (a) - (e) powtórzyć z $N = 100$ i z $N = 10$.