

**Statystyka**  
**Lista 3**

**Zad 1.** Zbudować model statystyczny następujących problemów. Czy podane  $T$  jest statystką? Jeśli tak, to sprawdzić, czy jest ona dostateczna.

- a) Liczymy ilość pieszych przechodzących z przez ulicę, aż do momentu, gdy pierwszy z nich zostanie potrącony przez przejeżdżający samochód. Zakładamy, że piesi poruszają się niezależnie. Interesuje nas prawdopodobieństwo potrącenia pojedynczego przechodnia. Niech  $T$  będzie nazwiskiem potrąconego pieszego.
- b) Z partii 500 telewizorów wybieramy 100. Interesuje nas liczba wadliwych sztuk w całej partii. Niech  $T$  odpowiada na pytanie, czy wszystkie telewizory z wybranej partii są zepsute.
- c) Interesuje nas liczba wypadków drogowych w Białymstoku w ciągu tygodnia. Zakładamy, że w przybliżeniu posiada ona rozkład Poissona. W tym celu odnotowujemy liczbę wypadków w ciągu roku (przez 52 tygodnie). Interesuje nas również liczba  $T$  wszystkich wypadków w ciągu roku.
- d) Producent bada partię  $n$  żarówek. Interesuje go czas życia, to jest liczba godzin do przepalenia się żarówki. Zakładamy, że czasy życia  $X_1, \dots, X_n$  badanych żarówek stanowią próbkę z rozkładu wykładniczego. Niech  $T$  będzie średnią arytmetyczną z tej próby.

**Zad 2.** Obserwujemy populację  $n$  ludzi. Dla każdej z osób znamy następujące 4 charakterystyki liczbowe: body fat, age, weight, height. Interesuje nas rozkład ilości tkanki tłuszczowej (body fat). Zakładamy, że body fat zależy afinicznie od pozostałych cech, tzn.

$$\text{bodyfat}_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{weight}_i + \beta_3 \text{height}_i + \text{error}_i \quad \text{dla } i = 1, 2, \dots, n$$

gdzie  $\text{error}_1, \dots, \text{error}_n \text{ i.i.d. } \sim \mathcal{N}(0, \sigma^2)$ , dla pewnego  $\sigma^2 > 0$ . Opisać model statystyczny (przestrzeń próbek, zbiór parametrów oraz rodzinę rozkładów prawdopodobieństwa) oraz wyznaczyć 5-wymiarową statystykę dostateczną dla rozpatrywanego problemu.

**Zad 3.** W populacji kobiet z indiańskiego plemienia Pima mieszkającego w okolicach Phoenix w Arizonie przeprowadzono testy pod kątem cukrzycy zgodnie z kryteriami Światowej Organizacji Zdrowia. Dla każdej z 532 kobiet zebrano następujące informacje: jestem nie lub jestem diabetikiem; wiek w latach; bmi (body mass index); npr (number of pregnancies); bp (diastolic blood pressure); glu (plasma glucose concentration). Zakładamy, że prawdopodobieństwo  $p_i$ , że  $i$ -ta kobieta jest diabetyczką wyraża się wzorem

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{bmi}_i + \beta_3 \text{npr}_i + \beta_4 \text{bp}_i + \beta_5 \text{glu}_i .$$

Opisać model statystyczny dla tego badania oraz wyznaczyć 6-wymiarową statystykę dostateczną dla nieznanych parametrów  $\beta_i$   $i = 0, 1, \dots, 5$ .

**Zad 4.** Niech  $h(x)$  będzie dodatnią funkcją całkowalną na  $\mathbb{R}$  oraz niech  $f(x; \xi, \eta)$  będzie funkcją gęstości postaci

$$f(x; \xi, \eta) = \begin{cases} c(\xi, \eta)h(x) & \text{dla } \xi < x < \eta, \\ 0 & \text{w przeciwnym razie.} \end{cases}$$

Niech  $(X_1, \dots, X_n)$  próbką z rozkładu z gęstością  $f(x; \xi, \eta)$ . Wykazać, że  $(X_{1:n}, X_{n:n})$  jest dostateczną statystyką dla parametrów  $(\xi, \eta)$ .