Programowanie narzędzi analitycznych – Przykładowa kartkówka – Z08

Zadanie 1

Zbiór danych `contest_data.csv` pochodzi ze strony `https://www.kaggle.com/jaysobel/kcbs-bbq`. This data set is the aggregate of 1,559 KCBS competitions from July 2013 through December 2016. The Kansas City Barbeque Society (KCBS) is "world's largest organization of barbeque and grilling enthusiasts with over 20,000 members worldwide." The data set was constructed by scraping the KCBS events page.

Wykorzystując niezerowe obserwacje dla zmiennej *prize* oszacować parametry rozkładu tej zmiennej - μ oraz σ metodą największej wiarygodności (zakładając rozkład normalny zmiennej *prize*) oraz sprawdzić statystyczną istotność parametru μ za pomocą statystyki z.

$$\ln L(\mu, \sigma; x) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \tag{1}$$

Zadanie 2

Z rozkładu $N(\mu, 2^2)$ wygenerowanych zostało 40 obserwacji:
0.62,-2.52,-0.31,-0.73,2.54,-1.52,-1.18,2.06,2.53,2.52,0.66,0.02,-0.93,-0.09,0.81 -1.60,2.70,0.52,1.75,-0.79, 3.66,-1.05,-1.32,-2.42,0.41,-2.09,2.67,1.36,0.94,0.58,-0.40,1.91,0.18,1.41,4.56,-0.01,-1.60,-0.07,1.79,2.23, 0.52,-2.81,-1.74,0.71,2.09,2.25,1.33,0.37,-2.04,2.29.
Oszacować parametr μ metodą największej wiarygodności. Przetestować statystyczną istotność parametru μ.

Zadanie 3

Rozważmy ponownie zbiór danych o nagrodach w konkursach barbecue. Nie ograniczając zbioru danych wyznaczyć parametry rozkładu normalnego nieprzyjmującego wartości ujemnych (w tym przypadku rozwiązania brzegowe).

$$X \sim N(\mu, \sigma^2) \Rightarrow f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\sum_{i=1}^{n}(x-\mu)^2}{2\sigma^2}\right)$$

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad \mathbb{P}(X < x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \Rightarrow \quad \mathbb{P}(X < 0) = \Phi\left(\frac{0-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{\mu}{\sigma}\right)$$

Rozkład normalny nieprzyjmujący wartości ujemnych dla rozwiązań brzegowych ma funkcję gęstości:

$$f(x_i; \mu, \sigma) = \begin{cases} 1 - \Phi\left(\frac{\mu}{\sigma}\right) & \text{dla} \quad y_i = 0 \\ \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) & \text{dla} \quad y_i > 0 \end{cases} \tag{2}$$

Funkcja $\Phi(.)$ jest dystrybuantą rozkładu standardowego normalnego i w programie R może być zapisana poleceniem `pnorm()`. W programowaniu funkcji wiarygodności:
`pnorm(mu/sigma, mean=0, sd=1)`.

Uwaga: Moglibyśmy rozważyć jeszcze inny rozkład: rozkład normalny obcięty w zerze, tj. taki, który określony jest tylko dla $x > 0$. Wówczas funkcja gęstości miałaby postać:

$$f(x; \mu, \sigma) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{1 - \Phi(0)} \tag{3}$$