# Modele Linowe

**Lista 2**

W tym zadaniu zostaną wykorzystane dane z pliku **ch01pr20.txt**. Druga kolumna zawiera liczbę kopiarek, a pierwsza kolumna zawiera czas (w godzinach) potrzebny na utrzymanie tych kopiarek.

1. Przedstaw dane na wykresie. Czy zależność jest w przybliżeniu liniowa?
2. Wyznacz regresję liniową z $y = \textit{czas obsługi}$ i $x = \textit{liczba obsługiwanych maszyn}$. Give the estimated regression equation. Oblicz slope i intersept za pomocą wzorów teoretycznych oraz poleceń wbudowanych w $R$. Do wykresu dodaj prostą regresji liniowej.
3. Wyznacz 95% przedział ufności dla slope'a i intersepta: za pomocą wzorów teoretycznych oraz poleceń wbudowanych w $R$.
4. Przedstaw wyniki testu istotności slope'a i intersepta. Podaj testowane hipotezy, statystyki testowe z liczbą stopni swobody, p-wartości i własne wnioski. Oblicz statystyki testowe oraz p-wartości za pomocą wzorów teoretycznych oraz poleceń wbudowanych w $R$.
5. Podaj estymator wartości oczekiwanej czasu obsługi, której można oczekiwać, gdyby serwisowanych było $k$ maszyn oraz 95% przedział ufności dla tej wartości, $k \in \{1, 5, 8, 11, 25, 100\}$. Oblicz przedział ufności za pomocą wzorów teoretycznych oraz poleceń wbudowanych w $R$. Jak długość przedziału ufności zależy od miejsca punktu względem danych?
6. Podaj przewidywany czas obsługi, którego można oczekiwać, jeśli $k$ maszyn było serwisowanych oraz 95% przedział predykcyjny dla tego czasu, $k \in \{1, 5, 8, 11, 25, 100\}$. Oblicz przedział predykcyjny za pomocą wzorów teoretycznych oraz poleceń wbudowanych w $R$. Jak długość przedziału predykcyjnego zależy od miejsca punktu względem danych?
7. Dodaj do wykresu z danymi 95% przedziały ufności oraz przedziały predykcyjne dla poszczególnych obserwacji. Wyjaśnij, dlaczego przedziały ufności są zawsze mniejsze niż przedziały predykcyjne.
8. Załóżmy, że $n = 40$, $\sigma^2 = 70$, $\mathrm{SSX} = \sum \left(X_i - \bar{X}\right)^2 = 500$.

   **a)** Oblicz moc dla odrzucenia hipotezy zerowej $\mathrm{H}_0 : \beta_1 = 0$, na poziomie istotności $\alpha = 0,05$, jeżeli prawdziwa wartość $\beta_1 = 1$.

   **b)** Przedstaw na wykresie moc jako funkcję $\beta_1$ dla wartości $\beta_1$ pomiędzy -2 a 2.

   **c)** Powtórz zadanie b) dla wartości $\sigma^2 = 120$. Dodaj odpowiedni wykres do wykresu z zadania b). Porównaj wyniki.

9. Wygeneruj wektor $X = (X_1, \dots X_{200})^T$ z wielowymiarowego rozkładu normalnego $\mathcal{N}(0, \frac{1}{500}I)$. Następnie wygeneruj 1000 wektorów $Y$ z modelu $Y = 5 + \beta_1 X + \varepsilon$, gdzie

   **a)** $\beta_1 = 0$, $\varepsilon \sim \mathcal{N}(0, I)$.

   **b)** $\beta_1 = 0$, $\varepsilon_1, \dots, \varepsilon_{200}$ są iid z rozkładu wykładniczego Exp(1).

   **c)** $\beta_1 = 0$, $\varepsilon_1, \dots, \varepsilon_{200}$ są iid z rozkładu logistycznego $\mathrm{L}(0, 1)$.

   **d)** $\beta_1 = 2$, $\varepsilon \sim \mathcal{N}(0, I)$.

   **e)** $\beta_1 = 2$, $\varepsilon_1, \dots, \varepsilon_{200}$ są iid z rozkładu wykładniczego z parametrem $\lambda = 1$.

   **f)** $\beta_1 = 2$, $\varepsilon_1, \dots, \varepsilon_{200}$ są iid z rozkładu logistycznego $\mathrm{L}(0, 1)$.

   Dla każdego powtórzenia eksperymentu przetestuj hipotezę $H_0 : \beta_1 = 0$ i estymuj prawdopodobieństwo odrzucenia $H_0$ na podstawie częstości odrzuceń w próbie (osobno dla każdego z punktów (a)-(f)). Porównaj te estymatory prawdopodobieństwa z teoretycznym prawdopodobieństwem błędu I rodzaju (a, b, c) oraz teoretyczną mocą (d, e, f) obliczoną przy założeniu, że szum ma rozkład normalny. Podsumuj wyniki.

## ***Zadania teoretyczne*** $(+1pkt)$

Dla modelu liniowego $Y = \beta_0 + \beta_1 X + \varepsilon$ na podstawie $n = 20$ obserwacji uzyskano estymatory $b_0 = 1$, $b_1 = 3$ i $s = 4$.

1. $s(b_1) = 1$, gdzie $s(b_1)$ jest estymatorem odchylenia standardowego $b_1$. Skonstruuj 95% przedział ufności dla $\beta_1$.

2. Czy masz statystyczne uzasadnienie dla twierdzenia, że $Y$ zależy od $X$?

3. 95% przedział ufności dla $\mathrm{E}(Y)$, gdy $X = 5$ wynosi $[13, 19]$. Znajdź odpowiedni przedział predykcyjny.