Wskazówki i odpowiedzi do zadań domowych z SMWD
Laboratorium 4: Zadanie 2, Zadanie 6, Zadanie 8, Zadanie 10

Zadanie 2

1) Model regresji: $\mathbf{Y = \beta_0 + \beta_1 X + \varepsilon}$, gdzie Y – cholesterol, X - wiek.

2) Wprowadzić dane: lab4.zad2 (*cholesterol*, *wiek*).

3) Wykres rozproszenia i współczynnik korelacji:

Analizuj >> Korelacje

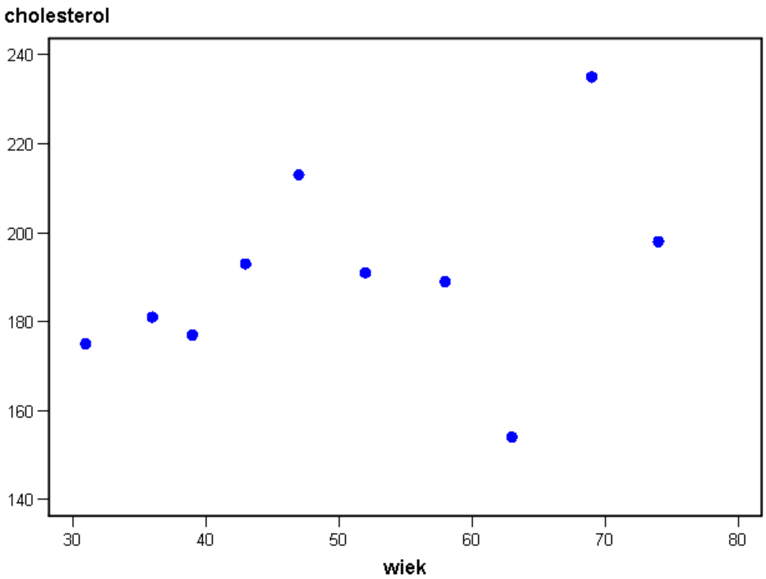
Role zadania: *cholesterol* >> zmienne analizowane, *wiek* >> koreluj z

Opcje: korelacja Pearsona

Rezultaty: utwórz wykresy punktowe

URUCHOM

Współczynniki korelacji Pearsona, N = 10 Prob > |r| przy H0: Ro=0	
	cholesterol
wiek	0.35672 0.3116



Zadanie 6

1) Model regresji: $\mathbf{sqrt(Y) = \beta_0 + \beta_1 X + \varepsilon}$, gdzie Y – wypadki, X - auta.

2) Wprowadzić dwa zbiory danych: lab4.zad6 (*auta*, *wypadki*), lab4.zad6p (*auta*)

3) Wykres rozproszenia i współczynnik korelacji (sqrt(wypadki) >> zmienne analizowane, auta >> koreluj z)
$\rho = 0.84888$, p-value = 0.0002.

4) Analizuj >> Regresja >> Liniowa

Role zadania: sqrt(wypadki) >> zmienna zależna, auta >> zmienne objaśniające

Dalej jak w zadaniach 3, 4

URUCHOM

5) Postawić odpowiednią hipotezę. Wynik dla testu F:
F = 28.37, p-value = 0.0002 $\Rightarrow$ istnieje związek między zmienną sqrt(Y) a X

6) sqrt(y) = 2.32509 + 0.00091517x

7) $R^2 = 0.7206$

8) Prognoza sqrt(Y)^ = 5.528 $\Rightarrow$ Y^ = $(5.528)^2$

9) Sprawdzenie założeń: dla testu Shapiro-Wilka pv-value = 0.6339

Zadanie 8

1) Model regresji: $\mathbf{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon}$, gdzie Y – wartość budynku, X_1 – powierzchnia, X_2 – odległość.

2) Tworzymy dwa zbiory: lab4.zad8 (wartość, powierzchnia, odległość), lab4.zad8p (powierzchnia, odległość)

3) Analizuj >> Regresja >> Liniowa

Role zadania: wartość >> zmienna zależna, powierzchnia, odległość >> zmienne objaśniające

Dalej jak w Materiałach pomocniczych dla regresji wielorakiej (lub zadaniu 7)

URUCHOM

4) Postawić odpowiednią hipotezę. Wynik dla testu F:

F = 26.67, p-value < 0.0010 $\Rightarrow$ istnieje związek między zmienną Y a którąkolwiek z X_1, X_2

5) Wyniki testów istotności t:

dla X_1 (powierzchnia) t = 6.15, p-value = 0.0008

dla X_2 (odległość) t = 2.71, p-value = 0.0352

6) $y = -19.87 + 1.93x_1 + 19.4x_2$

7) $R^2 = 0.8989$

8) prognoza Y^ = 346.988

Zadanie 10

1) Przyjmujemy następujący model regresji: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$,

gdzie: Y (mpg) – zużycie paliwa,

X_1 (accel) – przyśpieszenie,

X_2 (horsepower) – moc silnika,

X_3 (cylinders) – liczba cylindrów,

X_4 (weight) – waga samochodu,

X_5 (year) – rok produkcji.

2) Wynik dla testu F:

F = 88.88, p-value < 0.0001

3) Miara dopasowania $R^2 = 0.7553$

4) Wyniki testów istotności t:

dla X_1 (accel) p-value = 0.0998

dla X_2 (horsepower) p-value = 0.0468

dla X_3 (cylinders) p-value = 0.0180

dla X_4 (weight) p-value < 0.0001

dla X_5 (year) p-value = 0.0004

5) Usuwanie z modelu nieistotnej zmiennej X_1 (accel):

$Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$,

F = 109.1, p-value < 0.0001

$y = -10.09 - 0.11x_2 + 0.95x_3 - 0.01x_4 + 0.80x_5$, $R^2 = 0.7506$

Uwaga: Jeśli nie uwzględnimy w modelu regresji wyrazu wolnego, dostaniemy większą wartość $R^2 = 0.9847$