Zadania do lab 12

Krystian Baran 145000

25 maja 2021

Spis treści

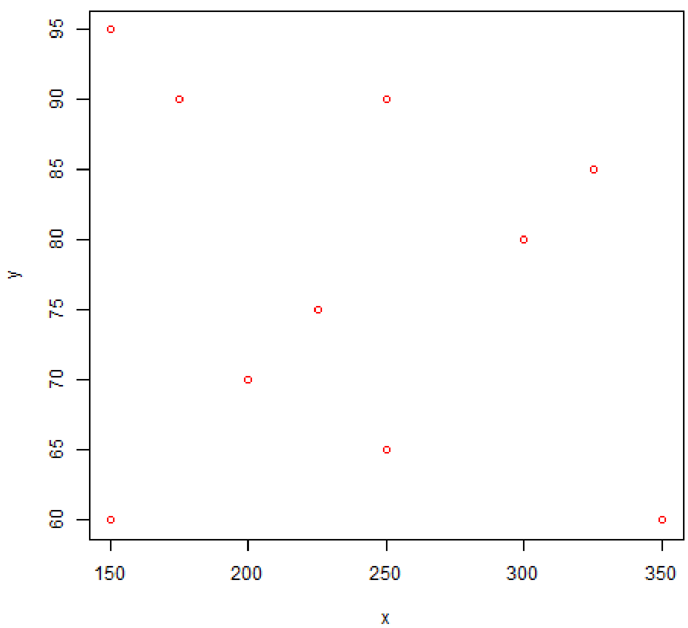
1 Zadanie 2

Odnotowano miesięczne dochody przypadające na jednego członka rodziny (w zł) - cecha X oraz wyrażoną w procentach część budżetu rodzinnego przeznaczoną na zakup artykułów żywnościowych i utrzymanie mieszkania - cecha Y.

X	200	300	150	225	175	350	150	250	325	250
Y	70	80	95	75	90	60	60	65	85	90

Sporządzić diagram rozrzutu, wyznaczyć oceny współczynników korelacji i determinacji między dochodem przypadającym na jednego członka rodziny a wydatkami na artykuły żywnościowe i utrzymanie mieszkania.

Wykres sporządzono w R wygląda następująco:



Obliczymy teraz współczynnik korelacji zgodnie ze wzorem:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} \cdot SS_{yy}}}$$

Gdzie:

$$SS_{xx} = \sum(x_i^2) - \frac{(\sum x_i)^2}{n} \overset{R}{=} sum(x^2) - sum(x)^2/length(x)$$

$$SS_{xy} = \sum x_i y_i - \frac{\sum x_i \cdot \sum y_i}{n} \stackrel{R}{=} sum(x * y) - sum(x) * sum(y)/length(x)$$

Otrzymano następujące wartości:

SS_{xx}	SS_{yy}	SS_{xy}	r
45312.5	1510	-1625	-0.1964517

Aby sprawdzić ten współczynnik to wyznaczymy hipotezę zerową orzekającą, że istnieje dodatnia korelacja:

H_0	$\rho \geqslant 0$
H_1	$\rho < 0$

Skorzystamy ze statystki testowej:

$$Z = (U - u_0) \cdot \sqrt{n-3}$$

Która ma w przybliżeniu standardowy rozkład normalny.

$$U = \frac{1}{2} \ln \frac{1+r}{1-r} = \frac{1}{2} \ln(\frac{0.8035483}{1.1964517} \approx -0.199039$$

$$u_0 = \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0} + \frac{\rho_0}{2n-2} = 0$$

$$Z_0 = -0.199039 \cdot \sqrt{10-3} \approx -0.526608$$

Obliczymy teraz *p-value* która wynosi:

$$\text{p-value} = \Phi(-0.526608) \stackrel{R}{=} pnorm(-0.526608, 0, 1) \approx 0.2992329$$

Wartość ta jest większa niż większość standardowo przyjętych α zatem nie możemy odrzucić hipotezę zerową więc nie istnieje ujemna korelacja między cechą X i Y.

Aby wyznaczyć współczynnik determinacji potrzebne jest wyliczenie SSE, które wyraża się następująco:

$$\text{SSE} = \sum (y_i - \hat{y}_i)^2$$

Zatem potrzebujemy najpierw wyznaczyć równanie regresji. Do tego równania potrzebujemy dwa współczynniki:

$$\beta_0 = \overline{y} - \beta_1 \overline{x} \stackrel{R}{=} mean(y) - b1 * mean(x) \approx 85.51724$$

$$\beta_1 = \frac{SS_{xy}}{SS_{xx}} = b1 \approx -0.03586207$$

Wtedy podstawiając kolejne wartosci x_i do równania regresji możemy obliczyć SSE:

$$\hat{y}_i = \beta_0 + \beta_1 \cdot x_i = 85.51724 - 0.03586207 \cdot x_i$$

Wtedy $SSE = 1451.724$ i można obliczyć współczynnik determinacji:

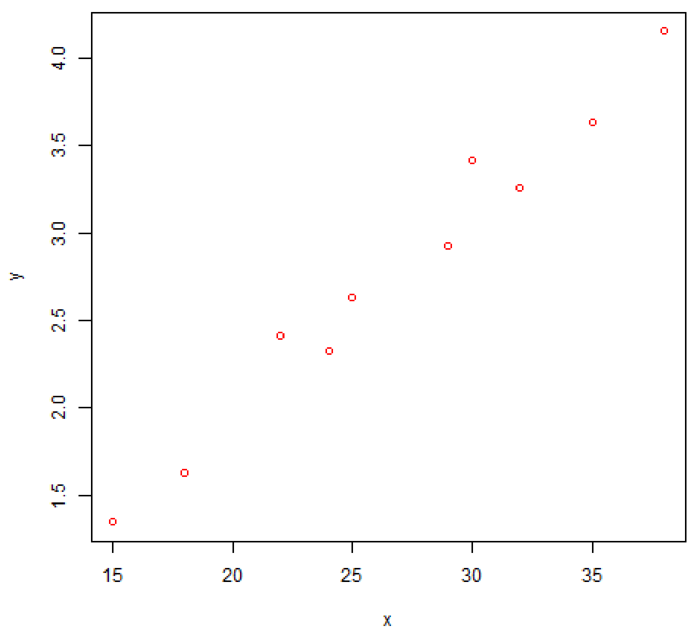
$$r^2 = 1 - \frac{SSE}{SS_{yy}} = 1 - \frac{1451.724}{1510} \approx 0.03859329$$

Zatem równanie regresji zmniejsza całkowitą sumę kwadratów próby o 3% od średniej arytmetycznej.

2 Zadanie 3

Naturalne jest przekonanie, że powinna być silna korelacja pomiędzy miesięcznymi obrotami firmy a jej liczebnością personelu handlowego. Dla pewnej firmy zostały zebrane dane dotyczące liczby sprzedawców w ostatnich 10 kwartałach oraz osiągane średniomiesięczne obroty (w mln zł) w tym czasie.
Wynoszą one: [15, 1.35], [18, 1.63], [24, 2.33], [22, 2.41], [25, 2.63], [29, 2.93], [30, 3.41], [32, 3.26], [35, 3.63], [38, 4.15].
Sprawdzić, czy to przekonanie potwierdziło się dla badanej firmy.

Sporządzono wykres rozrzutu w celu wstępnego sprawdzenia danych; wykres przygotowany w R.



Z wykresu widać że może istnieć dodatnia korelacja pomiędzy danymi. x oznacza liczbę pracowników, natomiast y oznacza średnio-miesięczne obroty. Następnie został obliczony współczynnik korelacji zgodnie ze wzorami podanymi w poprzednim zadaniu. Uzyskano następujące wartości:

SS_{xx}	SS_{yy}	SS_{xy}	r
485.6	6.97801	57.456	0.9870298

Wartość ta jest dodatnia, zatem jest możliwe że korelacja jest typu dodatniego. W celu sprawdzenia tego sporządzona została teza alternatywna i zerowa wyznaczona poniżej. ρ jest rzeczywistym współczynnikiem korelacji.

H_0	$\rho \leqslant 0$
H_1	$\rho > 0$

Ponieważ liczba obserwacji $n = 10 > 7$ możemy zastosować statystykę jak w poprzednim zadaniu.

$$Z = (U - u_0) \cdot \sqrt{n-3}$$

$$U = \frac{1}{2} \ln \frac{1+r}{1-r} = \frac{1}{2} \ln(\frac{1.9870298}{0.0129702} \approx 2.51587$$

$$u_0 = \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0} + \frac{\rho_0}{2n-2} = 0$$

$$Z_0 = 2.51587 \cdot \sqrt{10-3} \approx 6.656366$$

Obliczymy teraz *p-value* zgodnie ze wzorem:

$$\text{p-value} = 1 - \Phi(6.656366) \stackrel{R}{=} 1 - pnorm(6.656366, 0, 1) \approx 1.4034e - 11$$

Przyjmując $\alpha = 0.05$ widzimy że *p-value* jest od tej wartości mniejsze; zatem możemy odrzucić hipotezę zerową i wzioskować że istnieje dodatnia korelacja między liczbą personelu i miesięcznymi obrotami.