ANALIZA REGRESJI DWÓCH ZMIENNYCH

PRZYKŁAD

W wyniku badania relacji między efektywnością pracy, a wysokością wynagrodzenia, w grupie 35 losowo wybranych pracowników produkcyjnych, otrzymywano następujące informacje:

A. Łączny fundusz płac w badanej grupie pracowników wynosił 53 340 PLN
B. Najliczniejsza grupa pracowników wykonywała 32,4 sztuki produktu w ciągu godziny
C. Zaobserwowano, iż rozkład wydajności pracy jest normalny, przy czym 95,45% pracowników charakteryzowało się wydajnością (16,6 ; 48,2)
D. Suma kwadratów wynagrodzenia pracowników wynosiła 85 517 494,5 PLN^2
E. Szacując wynagrodzenie na podstawie skonstruowanej funkcji regresji popełniano błąd względny wynoszący 7,1% średniej wysokości wynagrodzenia

1. Określ siłę i kierunek związku między wydajnością pracy a wysokością wynagrodzenia.
2. Wyznacz rachunkowo obydwa równania regresji i zinterpretuj ich parametry
3. Oceń jakość dopasowania funkcji regresji wysokości wynagrodzenia względem wydajności pracy
4. Oszacuj wysokość wynagrodzenia przy wydajności pracy 35 szt na godzinę
5. Zapisz model regresji dla zależności wynagrodzenia od wydajności pracy w pełnej postaci

Informacje :

X – wydajność pracy

Y – wysokość wynagrodzenia

A. $\sum y_i = 53\,340$ PLN $\Rightarrow$ $\bar{y} = \frac{\sum y_i}{n}$ $\Rightarrow$ $\bar{y} = \frac{53\,340}{35} = 1\,523{,}996$

B. $D_x = 32{,}4$

C. rozkład X jest normalny N(32,4 ; s) $\Rightarrow \bar{x} = 32{,}4$

$P\{16{,}6 ; 48{,}2\} = 0{,}9545$ $\Rightarrow$ $\{\bar{x} - 2s(x) ; \bar{x} + 2s(x)\}$ $\Rightarrow$ $\{16{,}6 ; 48{,}2\}$

$$\begin{matrix} 32{,}4 - 2s(x) = 16{,}6 \\ 32{,}4 + 2s(x) = 48{,}2 \end{matrix} \Rightarrow \begin{matrix} 32{,}4 - 16{,}6 = 2s(x) \\ 2s(x) = 48{,}2 - 32{,}4 \end{matrix} \Rightarrow \begin{matrix} 15{,}8 = 2s(x) \\ 2s(x) = 15{,}8 \end{matrix} \Rightarrow \begin{matrix} 7{,}9 = s(x) \\ s(x) = 7{,}9 \end{matrix}$$

D. $\sum y_i^2 = 85\,517\,494{,}5$ $\Rightarrow$ $s^2(y) = \frac{\sum y_i^2}{n} - \bar{y}^2$ $\Rightarrow$

$$s^2(y) = \frac{85517494{,}5}{35} - 1523{,}996^2 = 120\,793{,}2 \quad \Rightarrow \quad s(y) = \sqrt{s^2(y)} = 347{,}55$$

E. $V_{s_{y(x)}} = 0{,}071$ $\Rightarrow$ $V_{s_{y(x)}} = \frac{s_{y(x)}}{\bar{y}} = \frac{s_e}{\bar{y}} = 0{,}071$

Rozwiązanie :

Ad. 1

siła i kierunek związku miedzy zmiennymi $\Rightarrow$ współczynnik korelacji liniowej Pearsona

$$s^2_{y(x)} = s^2(y)(1-r^2)$$

$$V_{s_{y(x)}} = \frac{s_{y(x)}}{\bar{y}} = 0{,}071 \quad \Rightarrow \quad s_{y(x)} = V_{s_{y(x)}} \cdot \bar{y} = 0{,}071 \cdot 1523{,}996 = 108{,}8158$$

$$s^2_{y(x)} = s^2(y)(1-r^2) \Rightarrow \quad \frac{s^2_{y(x)}}{s^2(y)} = 1-r^2 \Rightarrow \quad 1-\frac{s^2_{y(x)}}{s^2(y)} = r^2 \Rightarrow \quad \sqrt{1-\frac{s^2_{y(x)}}{s^2(y)}} = r$$

$$\boxed{r = \sqrt{1-\frac{108{,}8158^2}{120\,793{,}2}} = 0{,}9497}$$

Pomiędzy wysokością otrzymywanego wynagrodzenia a wydajnością pracy istnieje silna zależność

O dodatnim kierunku zależności wnioskujemy na podstawie analizy merytorycznej

Ad. 2

Parametry równań regresji:

$$\hat{y}_i = b_{1y(x)} \cdot x_i + b_{oy(x)} \qquad \hat{x}_i = b_{1x(y)} \cdot y_i + b_{ox(y)}$$

$$b_{1y(x)} = r\frac{s(y)}{s(x)} \qquad b_{1x(y)} = r\frac{s(x)}{s(y)}$$

$$b_{1y(x)} = 0{,}9497 \cdot \frac{347{,}55}{7{,}9} = 41{,}78 \qquad b_{1x(y)} = 0{,}9497 \cdot \frac{7{,}9}{347{,}55} = 0{,}0216$$

$$b_{0y(x)} = \bar{y} - b_{1y(x)} \cdot \bar{x} \qquad b_{0x(y)} = \bar{x} - b_{1x(y)} \cdot \bar{y}$$

$$b_{0y(x)} = 1523{,}996 - 41{,}78 \cdot 32{,}4 \qquad b_{0x(y)} = 32{,}4 - 0{,}0216 \cdot 1523{,}996$$

$$b_{0y(x)} = 170{,}25 \qquad b_{0x(y)} = -0{,}4993$$

$$\boxed{\hat{y}_i = 41{,}78 \cdot x_i + 170{,}25} \qquad \boxed{\hat{x}_i = 0{,}0216 \cdot y_i - 0{,}4993}$$

Ad. 3

Ocena jakości dopasowania funkcji regresji do danych empirycznych:

Współczynnik determinacji $r^2 = 0{,}902$

Współczynnik indeterminacji $\varphi^2 = 1 - r^2 = 1 - 0{,}902 = 0{,}098$

Błąd standardowy szacunku funkcji regresji $s_{y(x)} = V_{s_{y(x)}} \cdot \bar{y} = 108{,}8158$

Względny błąd szacunku funkcji regresji $V_{s_{y(x)}} = 0{,}071$

Błąd standardowy szacunku współczynnika regresji

$$S_{b_{1y}} = \frac{s_{y(x)}}{\sqrt{\sum x_i^2 - n\bar{x}^2}} = \frac{s_{y(x)}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

$$s^2(x) = \frac{\sum x_i^2}{n} - \bar{x}^2 \quad \Rightarrow \quad n \cdot \left(s^2(x) + \bar{x}^2\right) = \sum x_i^2 \Rightarrow \quad 35 \cdot \left(7{,}9^2 + 32{,}4^2\right) = 38\,925{,}95$$

$$S_{b_{1y}} = \frac{108{,}82}{\sqrt{38\,925{,}95 - 35 \cdot 32{,}4^2}} = \frac{108{,}82}{\sqrt{2184{,}35}} = 2{,}2383$$

Względny błąd szacunku współczynnika regresji $V_{s_{b_{1y(x)}}} = \frac{s_{b_{1y(x)}}}{b_{1y(x)}} = \frac{2{,}2383}{41{,}78} = 0{,}0557$

Błąd standardowy szacunku wyrazu wolnego

$$S_{b_{0y}} = \sqrt{\frac{s^2_{y(x)} \sum x_i^2}{n(\sum x_i^2 - n\bar{x}^2)}} = \sqrt{\frac{108{,}82^2 \cdot 38\,925{,}95}{35 \cdot \left(38\,925{,}95 - 35 \cdot 32{,}4^2\right)}} = 77{,}6455$$

Względny błąd szacunku wyrazu wolnego $V_{s_{b_{0y(x)}}} = \frac{s_{b_{0y(x)}}}{b_{0y(x)}} = \frac{77{,}6455}{170{,}25} = 0{,}4561$

Ad. 4

Szacunek wysokości wynagrodzenia dla pracownika o wydajności pracy równej 35 sztuk na godzinę

$$\hat{y}_i = b_{1y(x)} \cdot x_i + b_{oy(x)} \quad \pm \quad s_{y(x)}$$

$$\hat{y}_i(35) = 41{,}78 \cdot 35 + 170{,}25 = 1632{,}63 \ \boxed{\pm 108{,}82}$$

Ad. 5

Pełen zapis modelu regresji:

$$\boxed{\begin{array}{ll} y_i = 41{,}78 \cdot x_i + 170{,}25 + e_i & R^2 = 0{,}902 \\ \quad (2{,}2383) \quad (77{,}6455) \quad (108{,}82) & \end{array}}$$