

Задача 8.3. стр. 269

За изследване на връзката между броя на заетите и средния размер на работната заплата в сферата на обслужването е проведено наблюдение на 10 случайно избрани фирми. Получените резултати са поместени в следната таблица:

<i>№ на фирма</i>	<i>Средномесечен доход на човек във фирмата (в лв.)</i>	<i>Брой заети</i>
	<i>x_i</i>	<i>y_i</i>
1	320	24
2	580	27
3	440	26
4	300	28
5	450	30
6	320	32
7	360	35
8	240	38
9	280	40
10	200	50

1. Да се построи и анализират качествата на праволинеен регресионен модел на зависимостта между средния месечен доход на човек и броя на заетите във фирмите. Да се интерпретират получените резултати.

2. Да се измери силата на зависимост между средния месечен доход на човек и броя на заетите във фирмите. Да се интерпретират получените резултати.

Решение:

Целта на регресионния модел е едно явление следствие (Y) да се представи, като функция от едно или повече явления фактори (X) т.е. $Y_i = f(X_{1j}, X_{2j}, \dots, X_{kj}) + \varepsilon_i$

където:

ε_i е случаен компонент дължащ се на фактори, невключени в модела, на неправилно определяне вида на функцията f.

Величините които се включват в такъв модел са:

- ендогенни (зависими) - променлива, която се детерминира чрез съответния модел. Отбелязва се чрез символа Y_i . За конкретната задача това ще бъде броя на заетите;

- екзогенни (фактори) – те не се детерминират чрез модела, а се включват в него като дадени стойности. Отбелязват се чрез символа X_i . Това в случая ще бъде средномесечен доход на човек във фирмата в лева.

В зависимост от броя на факторите, които се включват в съответния модел, регресионните модели биват еднофакторни и многофакторни.

Чрез едно факторните регресионни модели се описва зависимостта между едно явление фактор и едно явление следствие, т.е. $Y_i = f(X_1) + \varepsilon_i$

В дадения модел ще се изрази зависимостта на броя на заетите от средномесечния доход на човек от фирмата.

В зависимост от вида на функцията f еднофакторните регресионни модели биват линейни и нелинейни.

Обективните зависимости имащи линеен характер се описват чрез модела:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Където:

$$\hat{Y}_i = \beta_0 + \beta_1 X_i$$

Величините \hat{Y}_i са теоретичните стойности за Y_i установени по дадения модел, съответстващи на стойностите на фактора X_i . Те отговарят на допускането, че между X и Y съществува функционална зависимост.

Избирайки вида на функцията трябва да се определят числените стойности на параметрите за конкретните статистически данни.

Параметрите на модела β_0 и β_1 се установяват по метода на най-малките квадрати чрез решаването на система от нормални уравнения от вида:

$$\begin{cases} \sum Y_i = N \cdot \beta_0 + \beta_1 \sum X_i \\ \sum X_i \cdot Y_i = \beta_0 \sum X_i + \beta_1 \sum X_i^2 \end{cases}$$

Чрез решаването на тази система по отношение на β_1 се получава:

$$\beta_1 = \frac{N \cdot \sum X_i \cdot Y_i - \sum X_i \cdot \sum Y_i}{N \cdot \sum (X_i)^2 - (\sum X_i)^2}$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}_i = \frac{\sum Y_i}{N} - \beta_1 \frac{\sum X_i}{N}$$