

Задача

За изследване на връзката между възрастта на продаваните коли втора употреба и тяхната продажна цена е проведено репрезентативно извадково наблюдение на 10 случайно избрани коли в една фирма за продажба на коли втора употреба.

Наблюдаваните резултати са поместени в таблицата по-долу:

Таблица 1:

№ на колата	Възраст на колата в години	Продажна цена в хил.лв.
1	4	12
2	5	8
3	6	8
4	11	4
5	12	5
6	6	7
7	9	10
8	12	5
9	13	2
10	4	11

1. Да се построи и оцени праволинеен регресионен модел на зависимостта, да се провери статистическата значимост на параметрите и да се интерпретират получените резултати.

2. Да се измери силата на зависимост между възрастта на продаваните коли втора употреба и тяхната продажна цена. Обосновете избор на метод.

3. Да се изчисли коефициент на детерминация и 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 продажната цена в хил.лв.;

- 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 възрастта на колата в години.

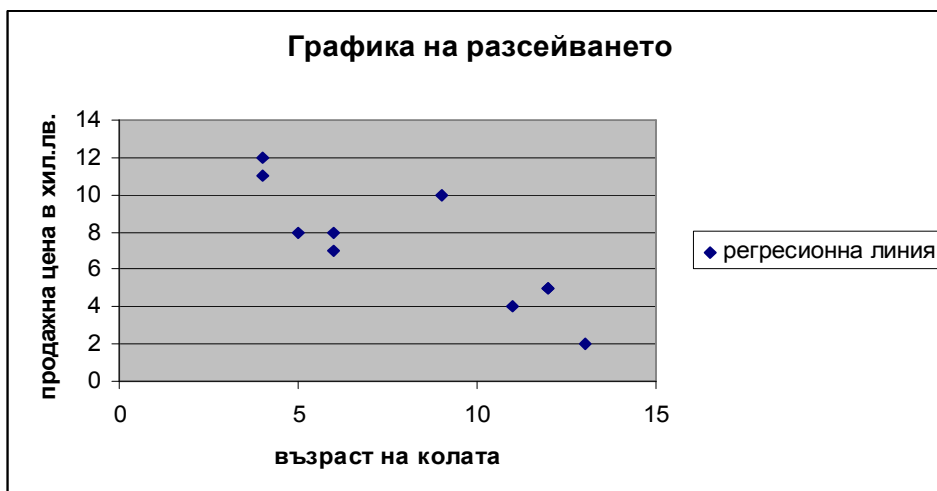
В зависимост от броя на факторите, които се включват в съответния модел, регресионните модели биват еднофакторни и многофакторни.

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

В дадения модел ще се изрази зависимостта между възрастта на колата в години и нейната продажна цена.

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

От построената графика се вижда, че съществува линейна зависимост.



Графика 1

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:

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

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

където:

N изразява поредният номер на наблюдението

Таблица 2: Работна таблица за построяване на праволинеен регресионен модел и определяне на неговите параметри

№	Възраст на колата в години X_i	Продажна цена в хил.лв. Y_i	$x \cdot y$	x^2	\hat{y}	$(\hat{y} - \bar{y})^2$	$(y - \hat{y})^2$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
1	2	3	4	5	6	7	8	9	10
1	4	12	48	16	10,4112	10,31180544	2,5242854	17,64	23,04
2	5	8	40	25	9,6465	5,98536225	2,7109623	10,24	0,64
3	6	8	48	36	8,8818	2,82845124	0,7775712	4,84	0,64
4	11	4	44	121	5,0583	4,58687889	1,1199989	7,84	10,24