Задача: Некоторая фирма занимается поставками различных грузов на короткие расстояния внутри города. Определите характер зависимости между расстоянием и затраченным временем. В качестве наиболее важного фактора, влияющего на время поставки, выбрано пройденное расстояние. Были собраны исходные данные о десяти поставках (таблица 1)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Расстояние, миль | 3,5 | 2,4 | 4,9 | 4,2 | 3 | 1,3 | 1 | 3 | 1,5 | 4,1 |
| Время, мин | 16 | 13 | 19 | 18 | 12 | 11 | 8 | 14 | 9 | 16 |

Требуется:

1. Определите характер зависимости между расстоянием и затраченным временем, используя мастер диаграмм MS Еxcel, проанализируйте применимость метода наименьших квадратов, постройте уравнение регрессии, используя МНК, проанализируйте силу регрессионной связи.
2. Проверить при α = 0,05 значимость коэффициента регрессии β1;
3. Определить при доверительной вероятности 0,95 интервальные опенки для β1, условного математического ожидания 𝑦̃ в точке x0=2 миль, а также интервальную опенку для значения y в точке предсказания xn+1=5.

# Решение

1. В первую очередь выделяем значение таблицы и с помощью мастера диаграмм строим диаграмму рассеяния

На основе диаграммы рассеяния определяем, что характер зависимости между расстоянием и временем является линейным. По линии тренда, можем сразу выяснить коэффициенты уравнения, где a = 0.3453, b = -1.8056.

Построим уравнение регрессии с помощью метода наименьших квадратов (МНК).

Найдем коэффициент детерминации по формуле:

где:

​ - сумма квадратов остатков (сумма квадратов разниц между фактическими значениями зависимой переменной и предсказанными значениями модели)

​ - общая сумма квадратов (сумма квадратов разниц между фактическими значениями зависимой переменной и их средним значением)

Таким образом .

Так как зависимость является линейной, МНК идеально подходит для нахождения силы регрессионной связи. отражает долю изменчивости зависимой переменной в данных, объясненную моделью. Чем ближе R² к 1, тем лучше модель объясняет данные.

1. Для проверки значимости коэффициента регрессии при α = 0,05 необходимо провести тест на значимость коэффициента наклона. По формуле находим значение

По таблице распределения Стьюдента также находим t при α = 0,05 и n = 10

t = 2,262

во много раз превышает t, следовательно можно отклонить нулевую гипотезу о равенстве коэффициента наклона нулю.

1. Для определения интервальных оценок для ​, условного математического ожидания в точке и интервальной оценки для значения в точке предсказания при доверительной вероятности 0.95, мы можем воспользоваться формулами для доверительных интервалов в линейной регрессии.
2. **Интервальная оценка для ​:**

Для этого мы используем стандартную ошибку коэффициента наклона ), которую мы ранее рассчитали. Доверительный интервал для при уровне доверия 0.95 можно найти с помощью формулы:

1. **Условное математическое ожидание ​ в точке** :

Для этого используем формулу:

1. **Интервальная оценка для значения *y* в точке предсказания**:

Для этого используем формулу: ​ Затем вычисляем стандартную ошибку прогноза:

Доверительный интервал для *y* при уровне доверия 0.95 можно найти с помощью формулы: