**Описание КП-5 (регрессионный анализ)**

Митус Полина, группа ОБ19-8

**Вариант 6**

Средствами Excel с использованием мастера функций, мастера диаграмм и надстройки «Анализ данных» построить уравнение линейной парной регрессии yx=ax+b.

На уровне значимости α = 0,05 оценить модель и параметры уравнения регрессии.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***x*** | 5,6 | 4,1 | 12,1 | 3,1 | 5,3 | 2,4 | 18,7 | 2,8 | 9,8 | 7 |
| ***y*** | 0,9 | 10,3 | 12 | 4,4 | 6 | 13,4 | -5,5 | 21,3 | 7,1 | -3,8 |

**Задача регрессионного анализа состоит в определении аналитического выражения связи двух или нескольких величин (признаков) (Х и У, где Х – независимая переменная, а У - зависимая) в случае, когда изменение зависимой величины (результативного признака) обусловлено влиянием одной или нескольких независимых величин (факторных признаков).**

**Задача решается о линейной зависимости признака У от признака Х.**

**Предположим, что возможна линейная зависимость и У можно получить в виде: у=ах+b (уравнение парной линейной регрессии).**

**Определим параметры этого уравнения (а и b). Строим оценки данных параметров по исходной выборке, используя метод наименьших квадратов. Сделаем это тремя способами: 1) по расчетной таблице 2) по графическому представлению исходных данных и построения линии тренда 3) при помощи анализа данных**

**Вычисления по расчетной таблице:**

1. Для начала перенесем таблицу с исходными данными в файл Excel
2. Найдем произведение х\*у, х^2, y^2

Полученные результаты:

Изображение выглядит как стол

Автоматически созданное описание

1. Рассчитаем сумму в каждом столбце B, C, D, E, F, воспользовавшись функцией Excel СУММ

Полученные результаты:



1. Рассчитаем среднее значение в каждом столбце B, C, D, E, F, воспользовавшись функцией Excel СРЗНАЧ

Полученные результаты:



1. Посчитаем выборочные дисперсии для Х и У по формулам:



Получили значение, равное 23,7929 (ячейка В14)



Получили значение, равное 59,7689 (ячейка С14)

1. Посчитаем коэффициент регрессии а по формуле:



Получили значение, равное -0,836 (ячейка В16)

1. Посчитаем формальное значение у при х=0 (b) по формуле:

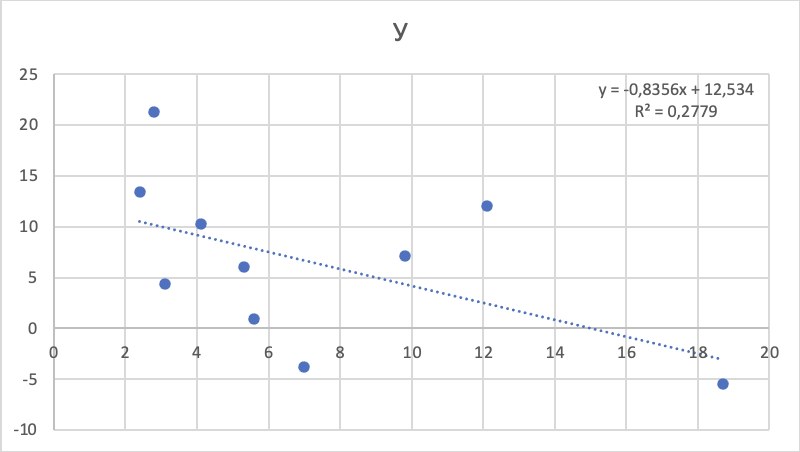


Получили значение, равное 12,53 (ячейка В17)

Полученное уравнение регрессии: y = -0,8356x + 12,534

**Графическое представление исходных данных и построение линии тренда:**

1. Строим точечную диаграмму для величин Х и У, а также линию тренда



1. Найдем линейный коэффициент корреляции по формуле:



Получили значение, равное -0,52719917 (ячейка Е16)

Следовательно, связь между Х и У является средней (1/3<R<2/3) и обратной (R<0)

1. Найдем коэффициент детерминации, возведя коэффициент корреляции в квадрат

Получили значение, равное 0,27793897 (изменение Х объясняет 28% изменения У)

Модель является не совсем удачной

1. Рассчитаем Утеор, Утеор – Усред, У-Утеор, Х-Хср для удобства дальнейших вычислений

Полученные значения:

Изображение выглядит как стол

Автоматически созданное описание

1. F- тест Фишера

Нулевая гипотеза – уравнение регрессии статистически незначимое, воздействие признака Х на У несущественно

1. Посчитаем F фактическое значение критерия Фишера по формуле:



Получили значение, равное 3,079 (ячейка Н16)

Также можно посчитать по формуле:



Для этого рассчитаем D фактическое по формуле:



Получили значение, равное 166,1211 (ячейка J16)

Рассчитаем D остаточное по формуле:



Получили значение, равное 53,94599 (ячейка J17)

Получаем F фактическое, равное 3,079 (ячейка Н17)

1. Вычислим критическое значение критерия Фишера (Fкрит.) при помощи функции Excel F.ОБР.ПХ, взяв за вероятность 0,05 (уровень значимости), а за степени свободы значения 1 и 8.

Получили значение, равное 5,317655072 (ячейка Н18)

1. Fфакт. < Fкрит. (3,079<5,317655072), значит, нулевая гипотеза принимается (уравнение регрессии статистически незначимое, воздействие признака Х на У несущественно).
2. Оценка статистической значимости параметров (t-статистика Стьюдента)

Нулевая гипотеза - показатели ***a***, ***b*** и ***rxy*** незначительно отличаются от нуля

1. Посчитаем стандартную ошибку коэффициента регрессии ma по формуле:



Получили значение, равное 0,47616328 (ячейка Е19)

1. Посчитаем стандартную ошибку сдвига mb по формуле:



Получили значение, равное 4,09779934 (ячейка Е20)

1. Посчитаем фактические значения t-статистики Стьюдента ta и tb по формулам:



Получили значение, равное -1,75482074 (ячейка G19)



Получили значение, равное 3,05878099 (ячейка G20)

1. Вычислим критическое значение t-статистики Стьюдента (tкрит.) при помощи функции Excel СТЬЮДЕНТ.ОБР.2Х, взяв за вероятность 0,05 (уровень значимости), а за степень свободы значение=8.

Получили значение, равное 2,30600414 (ячейка G21)

1. tb > tкрит (3,05878099 > 2,30600414), а ta < tкрит (-1,75482074 < 2,30600414), значит нулевая гипотеза (показатели ***a***, ***b*** и ***rxy*** незначительно отличаются от нуля) для показателя а принимается, для показателя b отвергается
2. Вычислим P-уровень для показателей a и b (ячейки I19 и I20) через функцию СТЬЮДЕНТ.РАСП.2Х, взяв наблюдаемые значения критерия (1,75482074\*(-1) и 3,05878099) за аргумент и значение степеней свободы (8).

Получили значения, равные 0,117368629 и 0,015611462.

1. Построим доверительные интервалы для a и b по формулам:





Получили интервал: (-1,93;0,26)





Получили интервал: (3,085;21,98)

Следовательно, параметр а является статистически незначимыми и несущественно отличен от нуля, параметр b не является статистически незначимыми и существенно отличен от нуля

**Вычисления при помощи анализа данных:**

1. Вышеперечисленные вычисления можно получить и через анализ данных (регрессия) в Excel:

Изображение выглядит как стол

Автоматически созданное описание

**Построение доверительного интервала прогноза:**

Доверительный интервал прогноза может быть построен только в том случае, если уравнение является статистически значимым и его коэффициенты являются статистически значимыми. В данном случае его строить нельзя, но покажем, как осуществляется построение.

1. Построим доверительный интервал прогноза. Хпр=10000. Посчитаем Yпр по формуле: *y = ax+b.*

Получили значение, равное -8343 (ячейка Е28)

1. Найдем ошибку прогноза по формуле:



Получили значение, равное 4758,26305 (ячейка Е29)

1. Строим доверительный интервал прогноза по формуле:



Полученный интервал: (-19316;2629)