**Практическое занятие №6**  **Корреляционный анализ. Регрессионный анализ. Анализ временных рядов.**

**Цель работы**: овладеть способами построения моделей линейной регрессии, и выработать умения и навыки оценки надежности коэффициента корреляции, уравнения регрессии и его коэффициентов.

**Задача**. При исследовании зависимости между выпуском готовой продукции *Y* (тыс. руб.) и энерговооруженностью труда *X* (кВт-час) получены следующие данные. Найти формулу, устанавливающую эту зависимость*.*

Таблица 2.1

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *X* | 1201 | 1300 | 1375 | 1412 | 1443 | 1500 | 1526 | 1516 | 1718 | 1783 | 1819 | 1877 | 1899 |
| *Y* | 14 | 21 | 27 | 29 | 30 | 31,5 | 35 | 34 | 41 | 38 | 39 | 46 | 51 |

Содержание работы: на основании данных необходимо:

1. Построить корреляционное поле и по характеру расположения точек в корреляционном поле определить общий вид регрессии.
2. Вычислить основные характеристики , , , , , , необходимые для построения модели регрессии.
3. Определить коэффициент корреляции , найти уровень его значимости и доверительный интервал (степень надежности определяется самостоятельно).
4. Написать эмпирические уравнения линий регрессий  на  и  на .
5. Вычислить коэффициент детерминации  и объяснить полученное значение.
6. Проверить адекватность уравнения регрессии.
7. Провести оценку величины погрешности уравнения регрессии  на  и его коэффициентов.
8. Построить уравнение регрессии  на  в первоначальной системе координат.

##### Выполнение работы

На основании, анализа взаимосвязи между выпуском готовой продукции и энерговооруженностью труда следует, что за факторный признак Х следует принять энерговооруженность труда, а выпуск готовой продукции за результативный признак Y.

**1.** Для определения формы связи между признаками X и Y строим на координатной плоскости точки (xi.yi), пользуясь табл. 2.1. Около построенных точек проводим так называемую линию тренда (рис. 2.1).

Рис. 2.1. Линейная регрессия *y* на *x* – прямая с уравнением

*y* = 0,0414 *x* – 31,322. Пунктированная прямая проведена «от руки».

Рис. 2.2. Линейная регрессия *x* на *y* – прямая с уравнением

*x* = 21,97 *y + 829,16*.

**2.** Произведем расчет статистик , , Sx, Sy, r, которые войдут в уравнения линий регрессий. Составим расчетную табл. 2.2.

Таблица 2.2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *xi* | *xi-ẍ* | *(xi-ẍ)2* | *yi* | *yi-ỹ* | *(yi-ỹ)2* | *x2* | *xy* |
| 1201 | -365,846 | 133843,408 | 14 | -19,577 | 383,256 | 1442401 | 16814 |
| 1300 | -266,846 | 71206,870 | 21 | -12,577 | 158,179 | 1690000 | 27300 |
| 1375 | -191,846 | 36804,947 | 27 | -6,577 | 43,256 | 1890625 | 37125 |
| 1412 | -154,846 | 23977,331 | 29 | -4,577 | 20,948 | 1993744 | 40948 |
| 1443 | -123,846 | 15337,870 | 30 | -3,577 | 12,794 | 2082249 | 43290 |
| 1500 | -66,846 | 4468,408 | 31,5 | -2,077 | 4,314 | 2250000 | 47250 |
| 1526 | -40,846 | 1668,408 | 35 | 1,423 | 2,025 | 2328676 | 53410 |
| 1516 | -50,846 | 2585,331 | 34 | 0,423 | 0,179 | 2298256 | 51544 |
| 1718 | 151,154 | 22847,485 | 41 | 7,423 | 55,102 | 2951524 | 70438 |
| 1783 | 216,154 | 46722,485 | 38 | 4,423 | 19,564 | 3179089 | 67754 |
| 1819 | 252,154 | 63581,562 | 39 | 5,423 | 29,410 | 3308761 | 70941 |
| 1877 | 310,154 | 96195,408 | 46 | 12,423 | 154,333 | 3523129 | 86342 |
| 1899 | 332,154 | 110326,178 | 51 | 17,423 | 303,564 | 3606201 | 96849 |
| 20369 | - | 629565,692 | 436,5 | - | 1186,923077 | 32544655 | 710005 |

Теперь найдем средние величины:

— средняя энерговооруженность труда.

— средний выпуск готовой продукции.

Рассчитаем следующие числовые характеристики:

Подставим найденные значения в формулу для нахождения коэффициента корреляции:

Коэффициент корреляции получился достаточно большим, его значение близко к единице. Следовательно, между признаками существует функциональная связь.

**3.** Теперь проверим значимость коэффициента корреляции. Вычислим статистику *t*p по формуле:

По таблице критических точек распределения Стьюдента по уровню значимости  и числу степеней свободы k = n – 2 = 13 – 2 = 11 находим tT = 2,201. Так как , то выборочный коэффициент корреляции значимо отличается от нуля. Следовательно, средний выпуск готовой продукции *Y* и средняя энерговооруженность труда *X* коррелированны.

Находим доверительный интервал для выборочного коэффициента корреляции  с надежностью . Так как объем выборки n = 13 < 50, то доверительный интервал находим по формуле: .

Так как по условию надежность (доверительная вероятность) равна , то по таблице функции Лапласа находим . Вычисляем среднюю квадратичную ошибку  по формуле:

Записываем доверительный интервал:

или ϵ [0,901;1].

Следовательно, с вероятностью 0,95 линейный коэффициент корреляции генеральной совокупности находится в пределах от 0,901 до 1. По имеющейся выборке следует ожидать влияние средней энерговооруженности труда на выпуск готовой продукции не менее чем на 90,1%.

**4.** Найдем эмпирические линейные уравнения регрессии *y* на *x* и *x* на *y*, которые являются приближенными уравнениями для истинных уравнений регрессий.

*ẏx* = 0,0414\**x* – 31,322

*ẋy*=21,97\**y*+829,16

Контроль вычислений:

*a1b1*=0,0414\*21,97=0,91

*r2*=0,954\*0,954=0,91

Получили, что *a1b1= r2*, а это значит, что вычисления выполнены верно.

Из уравнения *ẏx*=0,0414\**x*–31,322 следует, что при увеличении энерговооруженности труда на 1 кВт-час средний выпуск продукции растет на 0,0414 тыс. рублей.

**5.** Найдем коэффициент детерминации. Для линейной регрессии при вычисленном коэффициенте *r* он равен *r2*= 0,91. Это означает, что выпуск готовой продукции зависит от энерговооруженности труда на 91%, и только 9% рассеивания среднего показателя выпуска готовой продукции остались необъясненными.

**6.** Проверим адекватность уравнений линейной регрессии *y* на *x* по критерию Фишера-Снедекора. Для этого вычислим статистику *Fн* по формуле:

где *R2* определяем по формуле, используя расчетную табл. 2.3:

.

Таблица 2.3

|  |  |  |  |
| --- | --- | --- | --- |
| *yi* | *ẏx* | *yi-ẏx* | *(yi-ẏx)2* |
| 14 | 18,399 | -4,399 | 19,355 |
| 21 | 22,498 | -1,498 | 2,244 |
| 27 | 25,603 | 1,397 | 1,952 |
| 29 | 27,135 | 1,865 | 3,479 |
| 30 | 28,418 | 1,582 | 2,502 |
| 31,5 | 30,778 | 0,722 | 0,521 |
| 35 | 31,854 | 3,146 | 9,895 |
| 34 | 31,440 | 2,560 | 6,552 |
| 41 | 39,803 | 1,197 | 1,432 |
| 38 | 42,494 | -4,494 | 20,198 |
| 39 | 43,985 | -4,985 | 24,846 |
| 46 | 46,386 | -0,386 | 0,149 |
| 51 | 47,297 | 3,703 | 13,715 |
| Сумма | - | - | 106,839 |

При уровне значимости α=0,05 и числах степеней свободы *k1*=1, *k2*=13-2=11 по таблице критических точек распределения Фишера-Снедекора находим *FТ* = 4,844.

Так как *Fн*=111,204<4,844, то заключаем, что уравнение линейной регрессии *ẏх*=0,0414\*x–31,322 статистически значимо описывает результаты эксперимента.

**7.** Теперь найдем оценку величины погрешности уравнения регрессии  на  и его коэффициентов. Для этого составим табл. 2.4.

Таблица 2.4

|  |  |  |
| --- | --- | --- |
| *ui* | *ui-u* | *(ui-u)2* |
| -4,399 | -12,618 | 159,209 |
| -1,498 | -9,716 | 94,409 |
| 1,397 | -6,821 | 46,532 |
| 1,865 | -6,353 | 40,363 |
| 1,582 | -6,637 | 44,045 |
| 0,722 | -7,496 | 56,196 |
| 3,146 | -5,073 | 25,733 |
| 2,560 | -5,659 | 32,022 |
| 1,197 | -7,022 | 49,303 |
| -4,494 | -12,713 | 161,611 |
| -4,985 | -13,203 | 174,320 |
| -0,386 | -8,604 | 74,033 |
| 3,703 | -4,515 | 20,385 |
|  |  | 978,162 |

Проведем оценку величины погрешности уравнения регрессии *ẏх*=0,0414\*x–31,322.

Найдем относительную погрешность  уравнения по формуле:



где , , 

Для нахождения суммы  используем табл. 2.4.

Тогда

Так как величина  достаточно мала, то уравнение линейной регрессии *ẏх*=0,0414\*x–31,322 достаточно точно описывает опытные данные.

Оценим коэффициенты уравнения регрессии. У нас , . Для нахождения отношений  и  вычислим средние квадратические ошибки коэффициентов по формулам:

, , .

По табл. 2.2 находим: , . Учитывая, что , *r2*=0,91 и , находим:



Коэффициенты считаются значимыми, если выполняется условие: . Проверяя это условие для коэффициентов и получим, что коэффициент и коэффициент значимы.

Таким образом, полученное уравнение может быть принято для практического руководства.