**Практическое занятие 1.**

**Парная регрессия и корреляция в эконометрических**

**исследованиях**

***Методические указания***

*Парная (простая) регрессия* – регрессия между двумя переменными, т.е. модель вида

*y= f (x)*,

где *y* – зависимая переменная (результативный признак);

*х* – независимая переменная (факторный признак).

В зависимости от вида функции, объясняющей связь переменных, простая регрессия подразделяется на линейную и нелинейную. Линейная регрессия, которая выражается уравнением прямой (линейной функцией) вида:

*yх = a +bx,*

 где *yх* – теоретическое значение результативного признака, найденное из соответствующей математической функции связи *y* и *х*, т.е. уравнения регрессии

*или y = a + bx + ė,*

где *y* – фактическое значение результативного признака *y*;

*x* – значение факторного признака *х*;

*a, b* – параметры парной линейной регрессии;

*ė* – случайная величина [6].

Построение линейной регрессии сводится к оценке параметров – *a, b.* Для оценки этих параметров используется метод наименьших квадратов (МНК), которые вычисляются по формулам

*a=*$\overbar{y}-b\overbar{x,}$ (1)

$b=\frac{\overbar{x∙y}-\overbar{y}∙\overbar{x}}{\overbar{x^{2}}-\overbar{x}^{2}}$. (2)

Тесноту связи изучаемых явлений оценивает линейный коэффициент корреляции

$r\_{xy}=\frac{\overbar{x∙y}-\overbar{y}∙\overbar{x}}{σ\_{x}∙σ\_{y}},$ (3)

где ‾*x* – среднее значение переменной *x* (выборочное среднее);

 ‾*y* – среднее значение переменной *y* (выборочное среднее);

  - среднее значение произведений;

 *σ*х и *σ*y – выборочные средние квадратичные отклонения переменных  *х* и *y*:

**; (4)

Коэффициент корреляции изменяется в диапазоне от -1 до 1.

Для оценки качества подбора линейного уравнения регрессии используют коэффициент детерминации

$R^{2}=(r\_{xy})^{2}$. (5)

Оценка значимости уравнения регрессии в целом осуществляется с помощью *F*–критерия Фишера.

*F*–критерий Фишера рассчитывается по формуле

$F=\frac{D факт}{Dост},$ (6)

где *F*–критерий для проверки нулевой гипотезы $D\_{факт}= D\_{ост};$

$D\_{факт}$ – факторная дисперсия.

$D\_{факт}= \frac{\sum\_{}^{}(y\_{x}-\overbar{y})^{2}}{1}$,(7)

$D\_{ост}$ – остаточная дисперсия.

$D\_{ост }= \frac{\sum\_{}^{}(y- y\_{x})^{2}}{n-2}$. (8)

Рассчитанное значение *F*–отношений признается достоверным, если оно больше табличного, т. е связь между признаками существенна: $F\_{факт}>F\_{табл}$. Если $F\_{факт}<F\_{табл}$, то уравнение регрессии считается статистически незначимым [6].

Характеристикой общего качества построенной регрессии является средняя ошибка аппроксимации, которая находится по формуле

$\overbar{A}= \frac{1}{n}∙\left|\frac{y-y\_{x}}{y}\right|∙100.$ (9)

Для оценки значимости параметров уравнения регрессии, а также коэффициента корреляции рассчитывается стандартная ошибка по каждому параметру:

$m\_{b}=\sqrt{\frac{\sum\_{}^{}(y- y\_{x})^{2}/(n-2)}{\sum\_{}^{}(x-\overbar{x})^{2}}}=\sqrt{\frac{s^{2}}{\sum\_{}^{}(x-\overbar{x})^{2}}},$ (10)

$m\_{a}=\sqrt{s^{2}∙\frac{\sum\_{}^{}x^{2}}{n\*\sum\_{}^{}(x-\overbar{x})^{2}}}$, (11)

$m\_{r}=\sqrt{\frac{1-r^{2}}{n-2}},$ (12)

где *S*2 – остаточная дисперсия на одну степень свободы.

Для оценки существенности коэффициента регрессии рассчитывается *t* **–** критерий Стьюдента

$t\_{b}=\frac{b}{m\_{b}},$ (13)

$t\_{a}=\frac{a}{m\_{a}},$ (14)

$t\_{r}=\frac{r\_{xy}}{m\_{r}}.$ (15)

Фактическое значение критерия Стьюдента сравнивается с табличным при определенном уровне надежности *α* и числе степеней свободы *df* = *n –* 2. Если фактическое значение критерия Стьюдента больше табличного, гипотеза о несущественности связи отвергается, что свидетельствует о существенности уравнения регрессии в целом [3].

На основе стандартной ошибки параметров и табличного значения критерия Стьюдента рассчитываются доверительные интервалы

$γ\_{а}=a\pm t\_{таб}∙m\_{a},$ (16)

$γ\_{b}=b\pm t\_{таб}∙m\_{b}.$ (17)

Используя уравнение регрессии, можно получить предсказываемое значение результата ($y\_{р})$с помощью точечного прогноза при заданном значение фактора $х\_{р}$, путем подстановки в уравнение *yх = a +bx* соответствующего значения *х*.

Стандартная ошибка прогнозирования рассчитывается по формуле

$m\_{y\_{x}}=S∙\sqrt{\frac{1}{n}+\frac{(x\_{p}-\overbar{x})^{2}}{\sum\_{}^{}(x-\overbar{x})^{2}}}.$ (18)

Ошибка прогнозирования индивидуального значения результата включает не только стандартную ошибку $m\_{y\_{x}}$, но и случайную ошибку S

$m\_{y\_{p}}=S∙\sqrt{1+\frac{1}{n}+\frac{(x\_{p}-\overbar{x})^{2}}{\sum\_{}^{}(x-\overbar{x})^{2}}}.$ (19)

Доверительный интервал для прогнозируемого значения [1]:

$γ\_{y\_{p}}=a\pm t\_{таб}∙m\_{y\_{p}}.$ (20)

**Решение типовой задачи**

**Пример.** По территориям регионов приводятся данные зависимости затрат на производство от объемов выпуска продукции, представленные в табл. 1.

Таблица 1

**Исходные данные для расчета**

|  |  |  |
| --- | --- | --- |
| Номер наблюдения | Затраты на производство *у*, тыс. руб. | Объем выпуска *х*, тыс. шт. |
| 1 | 68,80 | 45,10 |
| 2 | 61,20 | 41,30 |
| 3 | 59,90 | 37,70 |
| 4 | 56,70 | 36,50 |
| 5 | 55,00 | 36,20 |
| 6 | 54,30 | 32,40 |
| 7 | 49,30 | 28,10 |

**Требуется:**

1. Построить линейное уравнение парной регрессии *y* от *x*.
2. Рассчитать линейный коэффициент корреляции и среднюю ошибку аппроксимации.

**Решение**

Для определения параметров (коэффициентов) уравнения регрессии используем формулы метода наименьших квадратов (1), (2) и рассчитанные в табл. 2 значения средних величин.

Таблица 2

**Промежуточные расчеты для расчета параметров**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Номер наблюдения | Затраты на производство *у*, тыс. руб. | Объем выпуска *х*, тыс. шт. | *ух* | *х*2 | *у*2 |
| 1 | 68,80 | 45,10 | 3102,88 | 2034,01 | 4733,44 |
| 2 | 61,20 | 41,30 | 2527,56 | 1705,69 | 3745,44 |
| 3 | 59,90 | 37,70 | 2318,13 | 1497,69 | 3588,01 |
| 4 | 56,70 | 36,50 | 2069,55 | 1332,25 | 3214,89 |
| 5 | 55,00 | 36,20 | 1991,00 | 1310,44 | 3025,00 |
| 6 | 54,30 | 32,40 | 1759,32 | 1049,76 | 2948,49 |
| 7 | 49,30 | 28,10 | 1385,33 | 789,61 | 2430,49 |
| Итого | 405,20 | 258,30 | 15153,77 | 9719,45 | 23685,76 |
| Среднее значение | 57,89 | 36,90 | 2164,82 | 1388,49 | 3383,68 |

$$b=\frac{2164,82-57,89×36,90}{1388,49-36,90^{2}}=\frac{2164,82-2136,14}{1388,49-1361,61}=\frac{28,68}{26,88}=1,07;$$

$$a=57,89-1,07×36,9=18,41;$$

$y\_{x}=18,41+1,07x$.

При увеличении объема выпуска на 1 тыс. ед. затраты на производство возрастут на 1 070 руб. [7].

Используя формулы (3), (4) и рассчитанные в табл.2 значения средних величин, определим коэффициент корреляции.

$$σ\_{x}^{2}=1388,49-(36,90)^{2}=27;$$

$$σ\_{y}^{2}=3383,64-(57,89)^{2}=31;$$

$$r\_{xy}=\frac{2164,82-36,90×57,89}{\sqrt{27×31}}=0,98.$$

Величина коэффициента корреляции означает, что наблюдается достаточно тесная связь рассматриваемых параметров, т.е. связь между факторным и результативным признаками сильная.

Расчет средней ошибки аппроксимации представлен в табл.3.

Таблица 3

**Промежуточные расчеты для расчета параметра**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Номер наблюдения | у | $$y\_{x}$$ | $$y-y\_{x}$$ | $$\left|\frac{y-y\_{x}}{y}\right|$$ |
| 1 | 68,8 | 66,68 | 2,12 | 0,03 |
| 2 | 61,2 | 62,61 | -1,41 | 0,02 |
| 3 | 59,9 | 59,82 | 0,08 | 0,01 |
| 4 | 56,7 | 57,46 | -0,76 | 0,01 |
| 5 | 55,0 | 57,13 | -2,13 | 0,04 |
| 6 | 54,3 | 53,06 | 1,24 | 0,02 |
| 7 | 49,3 | 48,44 | 0,86 | 0,02 |
| Итого | 405,2 | 405,20 | 0 | 0,15 |

$$\overbar{А}= \frac{1}{7 }∙0,15∙100=2,1\%.$$

Ошибка аппроксимации показывает хорошее соответствие расчетных и фактических данных: среднее отклонение 2,1% [7].

**Задания для самостоятельного решения**

***Задание 1.*** Торговое предприятие имеет сеть, состоящую из 12 магазинов, информация о деятельности которых представлена в табл.4.

Таблица 4

**Исходные данные для расчетов**

|  |  |  |
| --- | --- | --- |
| Номермагазина | Годовой товарооборот, млн руб. (переменная *y*) | Торговая площадь, тыс. м2 (переменная *x*) |
| 1 | 19,76 | 0,24 |
| 2 | 38,09 | 0,31 |
| 3 | 40,95 | 0,55 |
| 4 | 41,08 | 0,48 |
| 5 | 56,29 | 0,78 |
| 6 | 68,51 | 0,98 |
| 7 | 75,01 | 0,94 |
| 8 | 89,05 | 1,21 |
| 9 | 91,13 | 1,29 |
| 10 | 91,26 | 1,12 |
| 11 | 99,84 | 1,29 |
| 12 | 108,55 | 1,49 |

**Требуется:**

1. Построить линейное уравнение парной регрессии *y* от *x*.
2. Рассчитать линейный коэффициент корреляции и среднюю ошибку аппроксимации.