По семи территориям Уральского экономического района за 199Х г. Известны значения двух признаков (см. табл. 4) показателей «Среднедневная заработная плата одного работающего» (х, руб.) и «Доля расходов на покупку продовольственных товаров в общих расходах» (у, %).

Таблица 4

| **Регион** | **Доля расходов на покупку продовольственных товаров в общих расходах, % (у)** | **Среднедневная заработная плата одного работающего, руб. (х)** |
|---|---|---|
| Удмуртская респ. | 68,8 | 45,1 |
| Свердловская обл. | 61,2 | 59,0 |
| Башкортостан | 59,9 | 57,2 |
| Челябинская обл. | 56,7 | 61,8 |
| Пермская обл. | 55,0 | 58,8 |
| Курганская обл. | 54,3 | 47,2 |
| Оренбургская обл. | 49,3 | 55,2 |

**Требуется:**

1) оценить тесноту линейной корреляционно-регрессионной зависимости;

2) построить уравнение парной линейной регрессии;

3) записать модель парной линейной регрессии;

4) оценить качество уравнения регрессии.

**Решение.**

Объем выборки n = 7.

1) Теснота линейной корреляционно-регрессионной зависимости оценивается с помощью коэффициента парной корреляции.

Коэффициент парной корреляции найдем по формуле

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Составим расчетную таблицу (см. табл. 2) (расчеты выполнены в Excel).

Таблица 2

**Расчетная таблица**

| № п/п | $x_i$ | $y_i$ | $(x_i - \bar{x})$ | $(y_i - y_i)$ | $(x_i - \bar{x})\ (y_i - y_i)$ | $(x_i - \bar{x})^2$ | $(y_i - y_i)^2$ |
|---|---|---|---|---|---|---|---|
| 1 | 45,1 | 68,8 | -9,8 | 10,91 | -106,96 | 96,04 | 119,12 |
| 2 | 59 | 61,2 | 4,1 | 3,31 | 13,59 | 16,81 | 10,98 |
| 3 | 57,2 | 59,9 | 2,3 | 2,01 | 4,63 | 5,29 | 4,06 |
| 4 | 61,8 | 56,7 | 6,9 | -1,19 | -8,18 | 47,61 | 1,41 |
| 5 | 58,8 | 55 | 3,9 | -2,89 | -11,25 | 15,21 | 8,33 |
| 6 | 47,2 | 54,3 | -7,7 | -3,59 | 27,61 | 59,29 | 12,86 |
| 7 | 55,2 | 49,3 | 0,3 | -8,59 | -2,58 | 0,09 | 73,71 |
| **Сумма** | **384,3** | **405,2** | | | **-83,14** | **240,34** | **230,47** |

Средние значения признаков равны

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{384{,}3}{7} = 54{,}9$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{405{,}2}{7} = 57{,}9$$

Коэффициент парной корреляции равен

$$r_{xy} = \frac{-83{,}14}{\sqrt{240{,}34 \cdot 230{,}47}} = -0{,}353 \quad .$$

Статистическую значимость коэффициента корреляции проверим с помощью критерия Стьюдента. Найдем

$$t_{расч} = \frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \cdot \sqrt{n - m - 1} = \frac{-0{,}353}{\sqrt{1 - (-0{,}353)^2}} \sqrt{7 - 1 - 1} = -0{,}844 \quad (m = 1).$$

Табличное значение t-критерия Стьюдента при уровне значимости α = 0,05 и числе степеней свободы df = n - m - 1 = 7 - 1 - 1 = 14 составляет $t_{табл}$ = 2,57. (табличное значение найдено с помощью встроенной функции Excel «СТЬЮДРАСПОБР»).

Так как $|t_{расч}| > t_{табл}$ (|-0,844| < 2,57) , значение коэффициента корреляции статистически не значимо.

***Величина статистически не значимого коэффициента парной корреляции свидетельствует об умеренной связи (близкой к слабой) между***

***Долей расходов на покупку продовольственных товаров в общих расходах и Среднедневной заработной платой одного работающего.***

Коэффициент детерминации

$$R^2 = (r_{yx})^2 = (-0{,}353)^2 = 0{,}125$$

Показывает, что ***на 12,5% изменение Доли расходов на покупку продовольственных товаров в общих расходах объясняется изменениями Среднедневной заработной платой одного работающего***. Оставшиеся 87,5% приходятся на другие факторы, не включенные в модель.

***Построение уравнения парной линейной регрессии теоретически обосновано.***

2) Уравнение парной линейной регрессии $\hat{y} = b_0 + b_1 x$

найдем методом наименьших квадратов по формулам, предварительно проведя необходимые расчеты (здесь и далее см. табл. 3):

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{22162{,}34 - 405{,}2 \cdot 384{,}3 : 7}{21338{,}41 - (384{,}3)^2 : 7} = -0{,}346 \quad ,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 57{,}9 - (-0{,}346) \cdot 54{,}9 = 76{,}877$$

***Уравнение парной линейной регрессии имеет вид:***

$$\hat{y} = 76{,}877 - 0{,}346 \cdot x \ .$$

Коэффициент регрессии $b_1$ = -0,346 показывает, что с увеличением **Среднедневной заработной платы одного работающего** на 1 руб. Доля расходов на покупку продовольственных товаров в общих расходах уменьшается на 0,346%.

Подставляя в полученное уравнение регрессии значения $x_i$ можно определить ***теоретические значения*** $\hat{y}_i$ .

Таблица 3

**Расчетная таблица**

| % п/п | $x_i$ | $y_i$ | $x_i^2$ | $x_i \cdot y_i$ | $\hat{y}_i$ | $\left\lvert \frac{y_i - \hat{y}_i}{y_i} \right\rvert$ | $(\hat{y}_i - \bar{y})^2$ | $(\hat{y}_i - y_i)^2$ | $(x_i - \bar{x})^2$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 45,1 | 68,8 | 3102,88 | 2034,01 | 61,28 | 0,11 | 11,49 | 56,61 | 96,04 |
| 2 | 59 | 61,2 | 3610,80 | 3481,00 | 56,47 | 0,08 | 2,01 | 22,40 | 16,81 |
| 3 | 57,2 | 59,9 | 3426,28 | 3271,84 | 57,09 | 0,05 | 0,63 | 7,90 | 5,29 |
| 4 | 61,8 | 56,7 | 3504,06 | 3819,24 | 55,50 | 0,02 | 5,70 | 1,44 | 47,61 |
| 5 | 58,8 | 55 | 3234,00 | 3457,44 | 56,54 | 0,03 | 1,82 | 2,36 | 15,21 |
| 6 | 47,2 | 54,3 | 2562,96 | 2227,84 | 60,55 | 0,12 | 7,09 | 39,05 | 59,29 |
| 7 | 55,2 | 49,3 | 2721,36 | 3047,04 | 57,78 | 0,17 | 0,01 | 71,94 | 0,09 |
| **Сумма** | **384,3** | **405,2** | **22162,34** | **21338,41** | **405,20** | **0,57** | **28,76** | **201,71** | **240,34** |
| **Среднее** | 54,9 | 57,9 | | | 57,89 | 0,081 | | | |

3) ***Модель парной линейной регрессии имеет вид***

$$y = 76{,}877 - 0{,}346 \cdot x + \varepsilon$$

или

$$y_i = 76{,}877 - 0{,}346 \cdot x_i + \varepsilon_i$$

4) Оценим качество уравнения регрессии;

а) Найдем среднюю относительную ошибку аппроксимации

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left\lvert \frac{y_i - \hat{y}_i}{y_i} \right\rvert \cdot 100\% = \frac{0{,}57}{7} \cdot 100\% = 8{,}14\% \;.$$

Так как $\overline{A} < 10\%$, ***уравнение имеет хорошую точность***.

б) Проверим статистическую значимость уравнения регрессии в целом с помощью критерия Фишера. Необходимые расчеты приведены в табл. 3.

$$F_{расч} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \cdot \frac{n - m - 1}{m} = \frac{28{,}76}{201{,}71} \cdot \frac{7 - 1 - 1}{1} = 0{,}713 \;.$$

Табличное значение критерия Фишера с $df_1 = m = 1$ и $df_2 = n - m - 1 = 7 - 1 - 1 = 5$ степенями свободы при уровне значимости $\alpha = 0{,}05$ найдем с помощью встроенной функции Excel «FРАСПОБР». $F_{табл} = 6{,}61$ .

Поскольку $F_{расч} < F_{табл}$ , ***уравнение регрессии статистически не значимо в целом***.

в) Проверим статистическую значимость параметров уравнения регрессии с помощью критерия Стьюдента. Необходимые расчеты приведены в табл. 3.

Расчетные значения критерия равны

$$t_{b0} = \left|\frac{b_0}{m_{b0}}\right| = \frac{76{,}877}{22{,}62} = 3{,}40 \ ,$$

$$t_{b1} = \left|\frac{b_1}{m_{b1}}\right| = \frac{|-0{,}346|}{0{,}41} = 0{,}844 \ .$$

Средние квадратические ошибки параметров равны

$$m_{b0} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{(n-2)} \cdot \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}} = \sqrt{\frac{201{,}71}{7-2} \cdot \frac{21338{,}41}{7 \cdot 240{,}34}} = 22{,}62 \ ,$$

$$m_{b1} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{(n-2) \cdot \sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}} = \sqrt{\frac{201{,}71}{(7-2) \cdot 240{,}34}} = 0{,}41 \ .$$

Табличное значение критерия Стьюдента $t_{табл} = 2{,}57$ найдено в пункте 1).

Так как $t_{b0} > t_{табл}$ , ***параметр $b_0$ статистически значим***; $t_{b1} < t_{табл}$ , ***параметр $b_1$ статистически не значим***.

г) Найдем интервальные оценки параметров по формулам:

$$\left(b_0 - m_{b0} \cdot t_{табл} \ ; \ b_0 + m_{b0} \cdot t_{табл}\right);$$

$$\left(76{,}877 - 22{,}62 \cdot 2{,}57 \ ; \ 76{,}877 + 22{,}62 \cdot 2{,}57\right);$$

$$\left(b_1 - m_{b1} \cdot t_{табл} \quad ; \quad b_1 + m_{b1} \cdot t_{табл}\right);$$

$$\left(-0{,}346 - 0{,}41 \cdot 2{,}57 \quad ; \quad -0{,}346 + 0{,}41 \cdot 2{,}57\right).$$

***Интервальная оценка b₀ :*** $\left(18{,}73 \quad ; \quad 135{,}02\right)$ **.**

***Интервальная оценка b₁ :*** $\left(-1{,}40 \quad ; \quad 0{,}71\right)$ **.**

***Вывод:*** *полученное уравнение не значимо в целом, коэффициент* $b_1$ *и коэффициент корреляции не значимы*