Подробный отчёт

корреляционно-регрессионного анализа

С О Д Е Р Ж А Н И Е

Раздел 1. Двухфакторный регрессионный анализ исследуемых массивов данных. 3

1.1. Построение уравнения и нахождение параметров регрессии. 3

1.2. Точечные и интервальные оценки уравнения и параметров регрессии. 6

Раздел 2. Двухфакторный корреляционный анализ массивов данных. 13

2.1. Установление наличия и характеристика тесноты линейной связи. 13

2.2. Точечная и интервальные оценки множественной корреляции 15

Раздел 1. Двухфакторный регрессионный анализ исследуемых массивов данных.

1.1. При выполнении множественного корреляционно-регрессионного анализа, в отличие от парного, имеет смысл обратить порядок действий и сначала выразить предполагаемую зависимость в форме математической модели (регрессионного уравнения), а затем подтверждать или опровергать наличие линейной связи между анализируемыми массивами данных.

1.1. Построение уравнения и нахождение параметров регрессии.

1.2. Формулы построения выборочного уравнения **двухфакторной линейной регрессии** ($\hat{y}_{x_1x_2}$), а также нахождения его **параметров** ($\boldsymbol{b_0}, \boldsymbol{b_1}$ и $\boldsymbol{b_2}$), приведены ниже (1.1 – 1.4). Также, как и в парном регрессионном анализе, формулы были выведены **методом наименьших квадратов**:

$$\hat{y}_{x_1x_2} = b_0 + b_1x_1 + b_2x_2 \tag{1.1}$$

$$b_1 = \frac{\sum dx_1dy \cdot \sum dx_2^2 - \sum dx_2dy \cdot \sum dx_1dx_2}{\sum dx_1^2 \cdot \sum dx_2^2 - (\sum dx_1dx_2)^2} \tag{1.2}$$

$$b_2 = \frac{\sum dx_2dy \cdot \sum dx_1^2 - \sum dx_1dy \cdot \sum dx_1dx_2}{\sum dx_1^2 \cdot \sum dx_2^2 - (\sum dx_1dx_2)^2} \tag{1.3}$$

$$b_0 = \bar{y} - b_1 \cdot \overline{x_1} - b_2 \cdot \overline{x_2} \tag{1.4}$$

1.3. Под обозначениями dx_1, dx_2, dy здесь следует понимать отклонения индивидуальных значений двух предикторов и одного предиката от их средних арифметических (по отдельности: $x_{i1} - \bar{x}_1$; $x_{i2} - \bar{x}_2$; $y_i - \bar{y}$). Их индивидуальные значения, суммы и произведения приведены во вспомогательных таблицах. Вычисления представлены ниже (1.5 – 1.8):

$$b_1 = \frac{24{,}86 \cdot 4743{,}39 - 1055{,}54 \cdot 118{,}73}{4{,}78 \cdot 4743{,}39 - 118{,}73^2} = -0{,}87 \tag{1.5}$$

$$b_2 = \frac{1055{,}54 \cdot 4{,}78 - 24{,}86 \cdot 118{,}73}{4{,}78 \cdot 4743{,}39 - 118{,}73^2} = 0{,}24 \tag{1.6}$$

$$b_0 = 12{,}61 + 0{,}87 \cdot 2{,}32 - 0{,}24 \cdot 49{,}10 = 2{,}63 \tag{1.7}$$

$$\hat{y}_{x_1 x_2} = 2{,}63 - 0{,}87x_1 + 0{,}24x_2 \tag{1.8}$$

1.4. *Построенное выборочное уравнение регрессии позволяет предполагать, что с увеличением первого предиктора (x_1) на единицу, предикат (y) уменьшится на 0,87 при неизменном втором предикторе (x_2), а если возрастёт на единицу второй предиктор, то предикат увеличится на 0,24 при постоянном значении первого, а при нулевых значениях обоих предикторов ($x_1 = 0$; $x_2 = 0$), предикат примет значение 2,63.*

1.5. Для обеспечения сравнимости вычисленных параметров уравнения двухфакторной регрессии следует воспользоваться **бета-коэффициентом**, позволяющим определить на какую часть величины среднего квадратического изменится в среднем значение предиката при изменении предиктора на величину своего СКО, и **средним коэффициентом эластичности**, дающем понимание о том, на сколько процентов в среднем изменится предикат при изменении предиктора на один процент. Формулы и вычисления для анализируемых массивов данных приведены ниже (1.9 – 1.12):

$$\beta_{x_1 y} = b_1 \cdot \frac{s_{x_1}}{s_y} = -0{,}87 \cdot \frac{0{,}69}{6{,}94} = -0{,}09 \tag{1.9}$$

$$\beta_{x_2 y} = b_2 \cdot \frac{s_{x_2}}{s_y} = 0{,}24 \cdot \frac{21{,}78}{6{,}94} = 0{,}77 \tag{1.10}$$

$$\overline{Э}_{x_1 y} = b_1 \cdot \frac{\overline{x_1}}{\bar{y}} = -0{,}87 \cdot \frac{2{,}32}{12{,}61} = -0{,}16\% \tag{1.11}$$

$$\overline{Э}_{x_2 y} = b_2 \cdot \frac{\overline{x_2}}{\bar{y}} = 0{,}24 \cdot \frac{49{,}10}{12{,}61} = 0{,}95\% \tag{1.12}$$

1.6. *Исходя из результатов, можно предпожить,* что *с увеличением первого предиктора на одно его среднеквадратическое отклонение и при постоянстве значения второго предиктора, предикат уменьшится в среднем на 0,09 своего среднеквадратического отклонения, а при увеличении второго* *предиктора и неизменном состоянии первого – предикат возрастёт на 0,77 своей сигмы. С увеличением первого предиктора на один процент и при неизменности второго, уменьшение предиката составит 0,16%, а при увеличении второго и постоянстве первого – предикат увеличится на 0,95%.*

1.7. Множественную линейную регрессию графически следует отобразить в виде матричной диаграммы рассеяния (см. рис. 1.1). Результаты вычислений в подразделе также сведены в таблицу ниже (см. табл. 1.1):

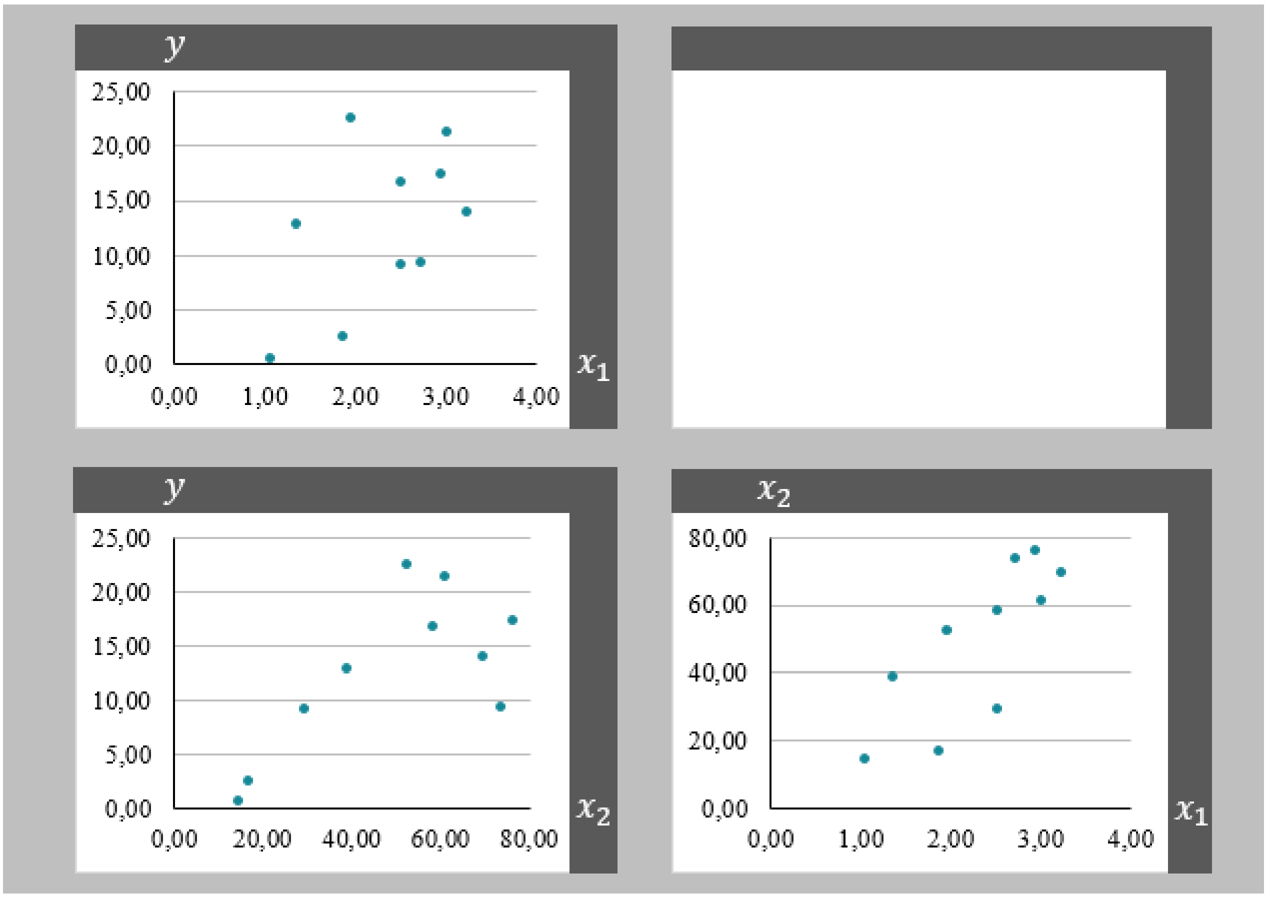

Рис. 1.1 Матричная диаграмма рассеяния анализируемых массивов данных.

Таблица 1.1

Уравнение и параметры регрессионного анализа.

Выборочное уравнение двухфакторной линейной регрессии	$\hat{y}_{x_1x_2} = 2{,}63 - 0{,}87x_1 + 0{,}24x_2$

Продолжение таблицы 1.1

Параметр/мера/оценка/показатель	Обозначение	Значение
Параметр b_1 уравнения двухфакторной регрессии	b_1	–0,87
Параметр b_2 уравнения двухфакторной регрессии	b_2	0,24
Параметр b_0 уравнения двухфакторной регрессии	b_0	2,63
Бета-коэффициент (для b_1)	$\beta_{x_1 y}$	–0,09
Бета-коэффициент (для b_2)	$\beta_{x_2 y}$	0,77
Средний коэффициент эластичности (для b_1)	$\bar{Э}_{x_1 y}$	–0,16%
Средний коэффициент эластичности (для b_2)	$\bar{Э}_{x_2 y}$	0,95%

1.2. Точечные и интервальные оценки уравнения и параметров регрессии.

1.8. При характеристике точности построенного уравнения множественной линейной регрессии, также, как и в случае с парной регрессией, невозможно обойтись без нахождения **остатков**, т.е. средних отклонений фактических значений предиката от вычисленных теоретических $(y_i - \hat{y}_i)$. Их значения и квадраты значений приведены во вспомогательных таблицах, а графическое представление в виде гистограммы дано ниже (см. рис. 1.2):

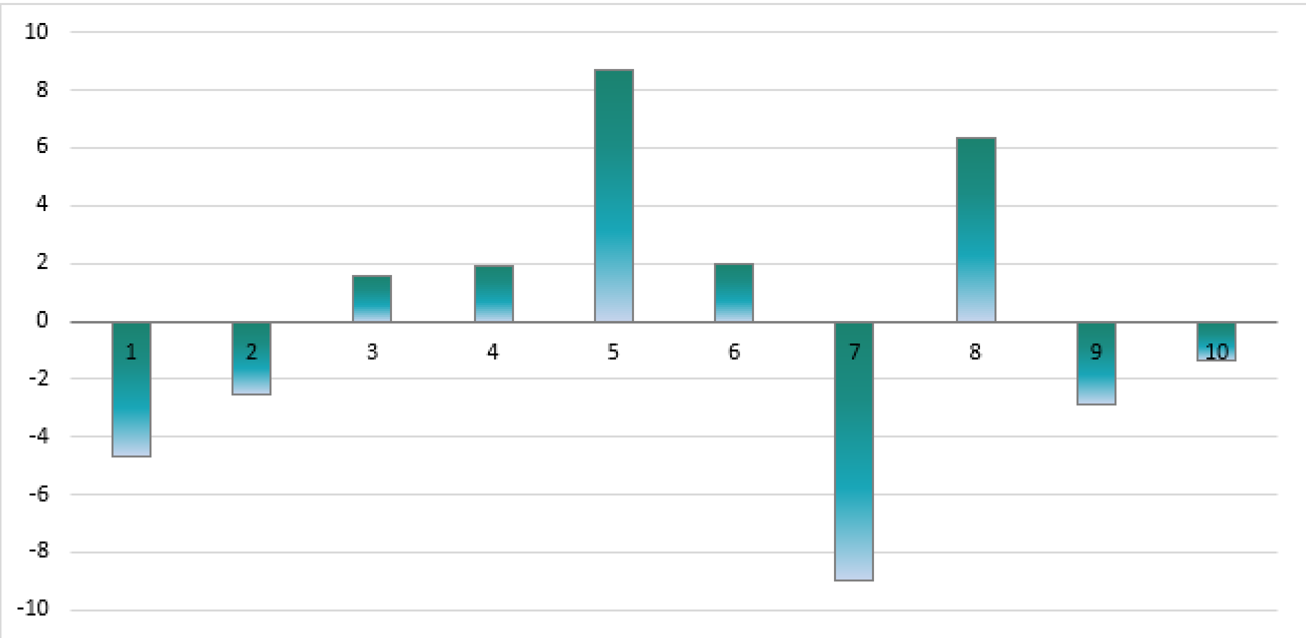

Рис. 1.2 Гистограмма остатков множественной регрессии.

1.9. Для характеристики качества подбора уравнения множественной линейной регрессии будут рассмотрены, во-первых, **средняя ошибка аппроксимации** ($\bar{A}$), представляющая собой среднее арифметическое модулей отношений остатков к фактическим значениям предикатного массива данных (1.13). Значение, лежащее в промежутке $[0{,}05;\ 0{,}07]$, свидетельствует о высокой точности подбора регрессионного уравнения, а в качестве допустимого значения коэффициента следует принять 0,15. Во-вторых, **остаточная дисперсия**, которая является несмещенной оценкой дисперсии остатков по степеням свободы k (1.15). Следует учитывать, что число степеней свободы здесь составит на одну константу меньше, чем при парной линейной регрессии (1.14). Наконец, нахождением квадратного корня из остаточной дисперсии получена **стандартная ошибка регрессии** (1.16):

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y_i - \hat{y}_i}{y_i}\right| = \frac{11{,}67}{10} = 1{,}17 \qquad (1.13)$$

$$k = n - m - 1 = 10 - 2 - 1 = 7 \qquad (1.14)$$

$$S_{\hat{y}}^2 = \frac{1}{k}\sum(y_i - \hat{y}_i)^2 = \frac{245{,}31}{7} = 35{,}04 \qquad (1.15)$$

$$S_{\hat{y}} = \sqrt{S_{\hat{y}}^2} = \sqrt{35{,}04} = 5{,}92 \qquad (1.16)$$

1.10. *Качество построенного уравнения регрессии оценивая средней ошибкой аппроксимации следует охарактеризовать как крайне неудовлетворительное, т.к. значение ошибки превышает допустимое почти в 8 раз* $(1{,}17 > 0{,}15)$. *Разброс фактических значений предикатного массива данных относительно теоретических в данной регрессионной модели составляет в среднем 5,92.*

1.11. Для дальнейшей характеристики точности построенного регрессионного уравнения будет использован **коэффициент множественной детерминации ($R^2_{x_1x_2y}$)**, рассчитываемый как отношение остаточной вариации к общей. Формула и вычисление ниже (1.17):

$$R^2_{yx_1x_2} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{236{,}24}{481{,}55} = 0{,}49 \quad (1.17)$$

1.12. Следует учитывать, что данный коэффициент для множественной линейной регрессии является неубывающей величиной, завышающей своё значение за счёт увеличения числа предикторов. В связи с этим, следует найти значение **скорректированного коэффициента множественной детерминации ($\boldsymbol{R^2_{x_1x_2y}}$)**, вычисленного с учётом степеней свободы дисперсий (1.18), который позволяет наиболее адекватно оценить качество подбора уравнения регрессии. А выражение данного показателя в процентах даст представление о том, какая доля вариации предиката может быть объяснена вариацией включённых в модель предикторов.

$$\overline{R^2_{yx_1x_2}} = 1 - \frac{n-1}{k}\left(1 - R^2_{yx_1x_2}\right) = 1 - \frac{9}{7}(1 - 0{,}49) = 0{,}34 \quad (1.18)$$

1.13. *Следовательно, точность подбора уравнения регрессии довольно низкая, потому как вариация предиката лишь на 34,50% может быть объяснена вариацией предикторов (доля независимой от предикторов вариации предиката составляет 65,50%).*

1.14. Теперь, также, как и в случае с парной регрессией, следует оценить статистическую значимость регрессионного уравнения по **F-критерию**, а для этого выдвинуть нулевую гипотезу (H_0), предполагающую для коэффициента множественной детерминации незначимое отличие от нуля (1.19). Отклоняется она при соблюдении неравенства $F_{\text{эмп}} \geq F_{\text{кр}}$. Формулы и вычисления ниже (1.20, 1.21):

$$H_0: R^2_{yx_1x_2} = 0 \quad (1.19)$$

$$F_{\text{эмп}} = \frac{R^2_{yx_1x_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{k}{m} = \frac{0{,}49}{1 - 0{,}49} \cdot \frac{7}{2} = 3{,}37 \quad (1.20)$$

$$F_{\text{кр}} = F_{\text{кр}}(m; k; \alpha) = F_{\text{кр}}(2; 7; 0{,}05) = 4{,}74 \quad (1.21)$$

1.15. *В связи с тем, что* $3{,}37 < 4{,}74$ *(то есть* $F_{\text{эмп}} < F_{\text{кр}}$*), нулевая гипотеза не отклоняется, и, следовательно, следует предполагать, что коэффициент детерминации и выборочное уравнение регрессии исследуемых* *массивов данных незначимо отличаются от нуля (статистически незначимы) с доверительной вероятностью* $P = 0{,}95$.

1.16 Для точечной оценки параметров b_1 и b_2 множественного регрессионного уравнения вычисляются их **стандартные ошибки** ($\boldsymbol{S_{b_1}}$ **и** $\boldsymbol{S_{b_2}}$), задействуя значение стандартной ошибки регрессии. Формулы и вычисления приведены ниже (1.22, 1.23):

$$S_{b_1} = \sqrt{\frac{\sum dx_2^2 \cdot S_{\hat{y}}^2}{\sum dx_1^2 \cdot \sum dx_2^2 - (\sum dx_1 dx_2)^2}} = \sqrt{\frac{4743{,}39 \cdot 35{,}04}{4{,}78 \cdot 4743{,}39 - 118{,}73^2}} = 4{,}41 \qquad (1.22)$$

$$S_{b_2} = \sqrt{\frac{\sum dx_1^2 \cdot S_{\hat{y}}^2}{\sum dx_1^2 \cdot \sum dx_2^2 - (\sum dx_1 dx_2)^2}} = \sqrt{\frac{4{,}78 \cdot 35{,}04}{4{,}78 \cdot 4743{,}39 - 118{,}73^2}} = 0{,}14 \qquad (1.23)$$

1.17. Для проверки статистической значимости параметра b_1 для него будет выдвинута нулевая гипотеза (H_0; 1.24), затем найден эмпирический **t-критерий Стьюдента** (t_{b_1}), вычисляемый отношением модуля параметра к его стандартной ошибке (1.25), и наконец, полученное значение будет сравнено с найденным по таблице критическим (1.26):

$$H_0\colon\ b_1 = 0 \qquad (1.24)$$

$$t_{b_1} = \frac{|b_1|}{S_{b_1}} = \frac{0{,}87}{4{,}41} = 0{,}20 \qquad (1.25)$$

$$t_{\text{кр}} = t_{\text{кр}}(k;\alpha) = t_{\text{кр}}(7; 0{,}05) = 2{,}36 \qquad (1.26)$$

1.18. *Так как* $0{,}20 < 2{,}36$ ($t_{b_1} < t_{\text{кр}}$), *нулевая гипотеза не отклоняется, и с* 95%*-ной вероятностью можно полагать, что параметр* b_1 *выборочного уравнения регрессии статистически незначим.*

1.19. Аналогичным образом следует выдвинуть и осуществить проверку нулевой гипотезы (H_0) для параметра b_2 (1.27, 1.28):

$$H_0\colon b_2 = 0 \qquad (1.27)$$

$$t_{b_2} = \frac{|b_2|}{S_{b_2}} = \frac{0{,}24}{0{,}14} = 1{,}75 \quad (1.28)$$

1.20. *Исходя из полученного неравенства* $1{,}75 < 2{,}36$ *(* $t_{b_0} < t_{кр}$ *), нулевая гипотеза не отклоняется. Параметр* b_0 *выборочного уравнения регрессии статистически незначим с вероятностью* $P = 0{,}95$.

1.21. Интервальная оценка способна подтвердить вышесказанное о параметрах b_1 и b_2 и определить границы числового множества для параметров β_1 и β_2 регрессионного уравнения генеральных совокупностей (1.29), в котором они будут располагаться с доверительной вероятностью $P = 0{,}95$.

$$\hat{Y} = \beta_0 + \beta_1 X_i + +\beta_2 X_2 \quad (1.29)$$

1.22. Формула построения доверительного интервала для параметра b_1 в двух видах (1.30, 1.31), графическое представление (см. рис. 2.3) и все вычисления приведены ниже (1.32, 1.33):

$$\beta_1 \in b_1 \pm t_{кр} \cdot S_{b_1} \quad (1.30)$$

$$b_1 - t_{кр} \cdot S_{b_1} < \beta_1 < b_1 + t_{кр} \cdot S_{b_1} \quad (1.31)$$

$$-0{,}87 \pm 2{,}36 \cdot 4{,}41 = -0{,}87 \pm 10{,}42 \quad (1.32)$$

$$-11{,}29 < \beta_1 < 9{,}56 \quad (1.33)$$

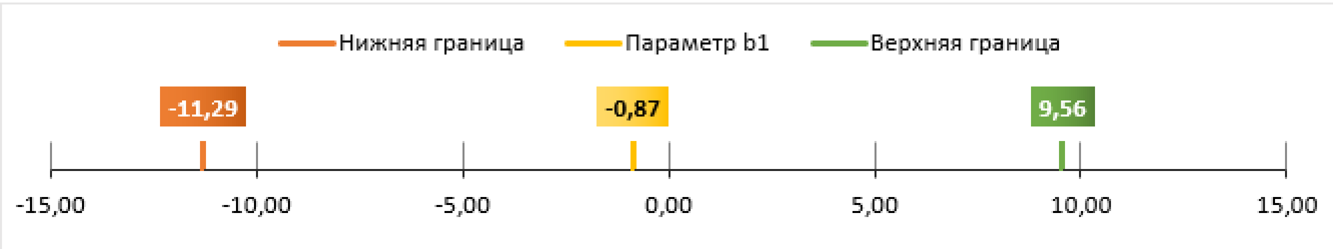

Рис. 1.3 Доверительный интервал для значения параметра β_1 (t-распределение Стьюдента; $P = 0{,}95$).

1.23. *Построенный интервал подтверждает незначимость параметра* b_1 *выборочного уравнения регрессии, так как точка 0 лежит внутри него, и с доверительной вероятностью* $P = 0{,}95$ *выявляет границы для параметра* β_1 *уравнения регрессии генеральных совокупностей в открытом числовом множестве* $(-11{,}29; 9{,}56)$.

1.24. Аналогичным образом ниже проведена интервальная оценка параметра b_2 выборочного уравнения регрессии (1.34 – 1.37; см. рис. 2.4):

$$\beta_2 \in b_2 \pm t_{кр} \cdot S_{b_2} \tag{1.34}$$

$$b_2 - t_{кр} \cdot S_{b_2} < \beta_2 < b_2 + t_{кр} \cdot S_{b_2} \tag{1.35}$$

$$0{,}24 \pm 2{,}31 \cdot 0{,}14 = 0{,}24 \pm 0{,}33 \tag{1.36}$$

$$-0{,}09 < \beta_2 < 0{,}57 \tag{1.37}$$

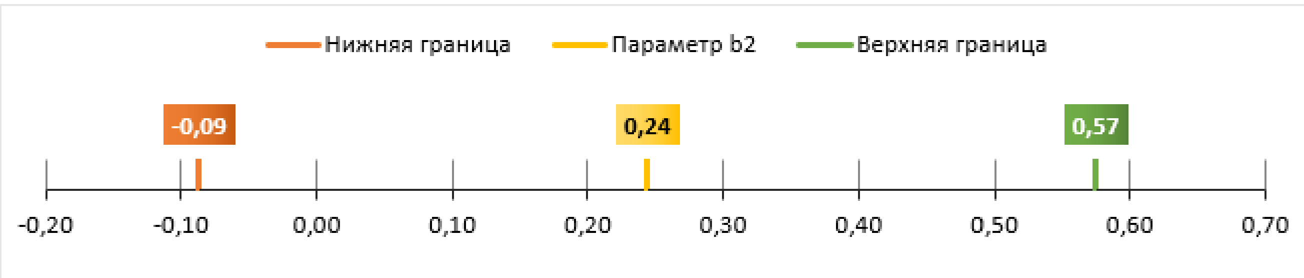

Рис. 1.4 Доверительный интервал для значения параметра β_2
(t-распределение Стьюдента; $P = 0{,}95$).

1.25. *Статистическая незначимость параметра* b_2 *выборочного уравнения регрессии подтверждена, так как точка* 0 *лежит внутри доверительного интервала, и параметр* β_2 *уравнения регрессии генеральной совокупности с* 95%*-ной вероятностью лежит в открытом числовом промежутке* $(-0{,}09; 0{,}57)$.

1.26. *Из всего этого следует заключить, что построенная выборочная модель регрессии для анализируемых массивов данных недостаточно репрезентативна и её использование имеет низкое практическое значение.* Все вычисления в подразделе сведены в таблицу ниже (см. табл. 1.2.):

Таблица 1.2.

Точечные и интервальные оценки выборочного уравнения регрессии.

Параметр/мера/оценка/показатель	Обозначение	Значение
Средняя ошибка аппроксимации	$\bar{A}$	1,17
Исправленная остаточная дисперсия	$S_{\hat{y}}^2$	35,04
Стандартная ошибки регрессии	$S_{\hat{y}}$	5,92
Коэффициент детерминации	$R_{yx_1x_2}^2$	0,49
Скорректированный коэффициент детерминации	$\overline{R_{yx_1x_2}^2}$	0,34
Эмпирическое значение F-критерия	$F_{эмп}$	3,37
Критическое значение F-критерия	$F_{кр}$	4,74
Стандартная ошибка параметра b_1	S_{b_1}	4,41
Стандартная ошибка параметра b_0	S_{b_2}	0,14
Эмпирическое значение t-критерия для b_1	t_{b_1}	0,20
Эмпирическое значение t-критерия для b_0	t_{b_2}	1,75
Критическое значение t-критерия	$t_{кр}$	2,36
Доверительный интервал для параметра β_1	β_1	(-11,29; 9,56)
Доверительный интервал для параметра β_0	β_2	(-0,09; 0,57)

Раздел 2. Двухфакторный корреляционный анализ массивов данных.

2.1. Прежде чем приступать к множественному корреляционному анализу исследуемых данных, следует графически сопоставить вычисленные ранее коэффициенты ковариации и корреляции анализируемых массивов данных в двух матрицах ниже (см. рис. 2.1). Для интерпретации обоих показателей ниже также приведены описательные шкалы (размещены на следующей странице ввиду громоздкости; см. рис. 2.2).

	x_1	x_2	y
x_1	0,48	11,87	2,06
x_2	11,87	474,34	105,55
y	2,06	105,55	48,15

	x_1	x_2	y
x_1	1,00	0,79	0,52
x_2	0,79	1,00	0,70
y	0,52	0,70	1,00

Рис. 2.1 Ковариационная и корреляционная матрицы для исследуемых массивов данных.

2.2. По корреляционной матрице можно предполагать о наличии **мультиколлинеарности**, т.е. наличии линейной связи между предикторами построенной регрессионной модели (отбор предикторов в исходных данных произведен неудовлетворительно), что также упоминает о её низком качестве в связи с неустойчивостью оценок параметров регрессии.

2.3. Линейную связь первого предиктора с предикатом следует охарактеризовать как прямую и заметную, также, как и связь второго предиктора с ним (число 0,70 округлено до двух знаков после запятой с 0,6984).

2.1. Установление наличия и характеристика тесноты линейной связи.

2.4. **Коэффициент множественной корреляции** ($\boldsymbol{r_{yx_1x_2}}$) позволяет оценить тесноту совместной связи сразу нескольких массивов данных. Вычислить

Шкала для интерпретации коэффициента ковариации:

	Прямая связь	$\mathrm{cov} > 0$
	Обратная связь	$\mathrm{cov} < 0$

Шкала Чеддока для описания коэффициента корреляции:

	Функциональная	$\lvert r \rvert = 1$
	Очень высокая	$0{,}90 < \lvert r \rvert < 0{,}99$
	Высокая	$0{,}70 < \lvert r \rvert < 0{,}90$
	Заметная	$0{,}50 < \lvert r \rvert < 0{,}70$
	Умеренная	$0{,}30 < \lvert r \rvert < 0{,}50$
	Слабая	$0 < \lvert r \rvert < 0{,}30$
	Отсутствует	$\lvert r \rvert = 0$

Рис. 2.2 Шкалы для характеристики коэффициента ковариации
и модуля коэффициента корреляции.

его можно путём извлечения квадратного корня из коэффициента множественной детерминации (2.1). Чем ближе модуль значения к единице, тем теснее связь предиката с набором предикторов. А чем больше он значения максимального парного коэффициента корреляции, тем качественнее построенная регрессионная модель ($r_{yx_1x_2} > r_{yx_j(\max)}$).

$$r_{yx_1x_2} = \sqrt{R^2_{yx_1x_2}} = \sqrt{0{,}49} = 0{,}70 \tag{2.1}$$

2.5. *Линейную связь предикторов с предикатом следует охарактеризовать как прямую и высокую. Значение коэффициента множественной корреляции не существенно отличается от значения максимальной по величине парной корреляции* ($r_{yx_2} = 0{,}6984 < r_{yx_1x_2} = 0{,}70$), *что в очередной раз подтверждает низкое качество регрессионной модели*.

2.2. Точечная и интервальные оценки множественной корреляции

2.6. Оценка значимости для коэффициента множественной корреляции проводится путем сравнения эмпирического значения t-критерия с критическим. Вначале следует выдвинуть нулевую гипотезу, предполагающую незначимость от нуля для коэффициента корреляции (2.2), затем рассчитать эмпирический t-критерий (2.3):

$$H_0\colon\ r_{yx_1x_2} = 0 \tag{2.2}$$

$$t_{\text{эмп}} = \frac{r_{yx_1x_2} \cdot \sqrt{k}}{\sqrt{1 - r^2_{yx_1x_2}}} = \frac{0{,}70 \cdot \sqrt{7}}{\sqrt{1 - 0.49}} = 3{,}64 \tag{2.3}$$

2.7. *Так как* $3{,}63 > 2{,}36$ *(соблюдено неравенство* $t_{\text{эмп}} \geq t_{\text{кр}}$ *), нулевая гипотеза* H_0 *отклоняется, и, следовательно, с вероятностью* $P = 0{,}95$ *можно полагать, что коэффициент множественной корреляции исследуемых массивов данных значимо отличается от нуля (статистически значим).*

2.8. Интервальный метод оценки, также, как и в случае с парной корреляцией, будет произведен двумя методами, первый из которых построение интервала задействованием t-критерия Стьюдента, формула (2.4, 2.5) вычисление (2.6, 2.7) и графическое представление даны ниже (см. рис. 2.3):

$$\rho_{YX_1X_2} \in r_{yx_1x_2} \pm t_{\text{кр}} \cdot \sqrt{\frac{1 - r^2_{yx_1x_2}}{k}} \tag{2.4}$$

$$r_{yx_1x_2} - t_{\text{кр}} \cdot \sqrt{\frac{1 - r^2_{yx_1x_2}}{k}} < \rho_{YX_1X_2} < r_{yx_1x_2} + t_{\text{кр}} \cdot \sqrt{\frac{1 - r^2_{yx_1x_2}}{k}} \tag{2.5}$$

$$0{,}70 \pm 2{,}36 \cdot \sqrt{\frac{1-0{,}49}{7}} = 0{,}79 \pm 0{,}64 \tag{2.6}$$

$$0{,}06 < \rho_{YX_1X_2} < 1 \tag{2.7}$$

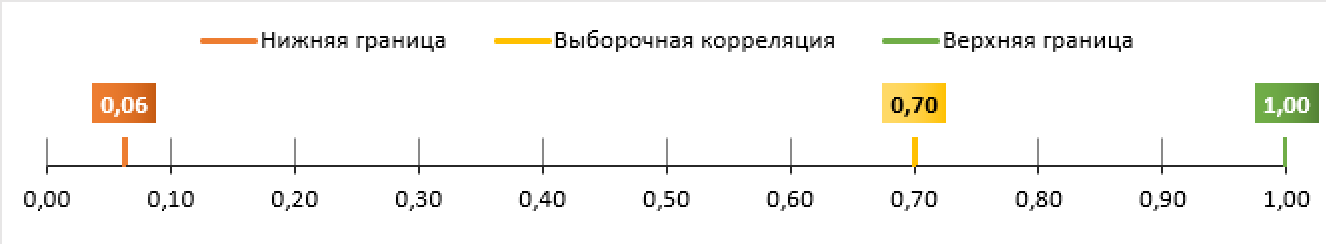

Рис. 2.3 Доверительный интервал для коэффициента множественной корреляции (t-распределение; P=0,95).

2.9. Так как значение верхней границы интервала составило больше единицы, оно было доведено до неё в соответствии с пределами коэффициента корреляции $[-1; 1]$. *Следует сделать предположение о том, что коэффициент множественной корреляции генеральных совокупностей будет располагаться в пределах* $(0{,}06; 1]$ *с* 95%*-ной вероятностью.*

2.10 Для получения интервала с более узкими и точными границами будет задействованы нижняя и верхняя границы **z-преобразования Фишера** для выборочного коэффициента множественной корреляции. Формулы и вычисления (2.8 – 2.12), а также графическое представление (см. рис. 2.4) приведены ниже:

$$z_L = \frac{1}{2} \cdot \ln\left(\frac{1+r_{yx_1x_2}}{1-r_{yx_1x_2}}\right) - \frac{1{,}96}{\sqrt{k-1}} = \frac{1}{2} \cdot \ln\left(\frac{1+0{,}70}{1-0{,}70}\right) - \frac{1{,}96}{\sqrt{6}} = 0{,}07 \tag{2.8}$$

$$z_U = \frac{1}{2} \cdot \ln\left(\frac{1+r_{yx_1x_2}}{1-r_{yx_1x_2}}\right) + \frac{1{,}96}{\sqrt{k-1}} = \frac{1}{2} \cdot \ln\left(\frac{1+0{,}70}{1-0{,}70}\right) + \frac{1{,}96}{\sqrt{6}} = 1{,}67 \tag{2.9}$$

$$\frac{e^2 \cdot z_L - 1}{e^2 \cdot z_L + 1} < \rho_{YX_1X_2} < \frac{e^2 \cdot z_U - 1}{e^2 \cdot z_U + 1} \tag{2.10}$$

$$\frac{7.39 \cdot 0{,}07 - 1}{7.39 \cdot 0{,}07 + 1} < \rho_{YX_1X_2} < \frac{7.39 \cdot 1{,}67 - 1}{7.39 \cdot 1{,}67 + 1} \tag{2.11}$$

$$0{,}07 < \rho_{YX_1X_2} < 0{,}93 \qquad (2.12)$$

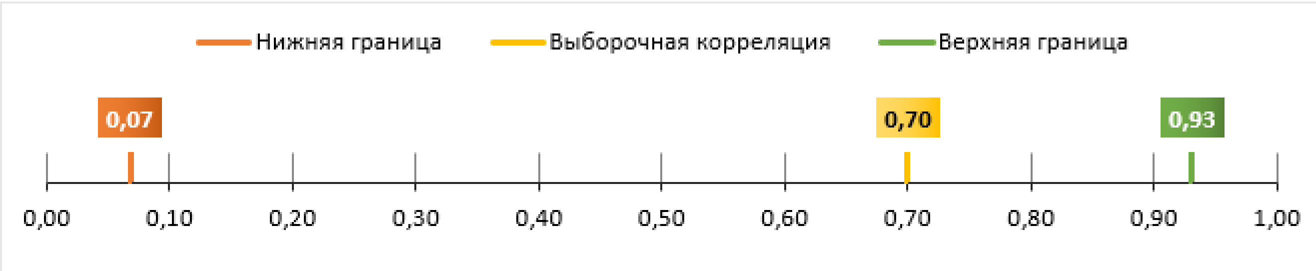

Рис. 2.4 Доверительный интервал для коэффициента множественной корреляции (z-преобразование Фишера; P=0,95).

2.11. *Построенный интервал позволяет с вероятностью* $P = 0{,}95$ *предположить, что множественный коэффициент корреляции генеральных совокупностей будет находится в пределах* $(0{,}07; 0{,}93)$. Все результаты вычислений в разделе сведены в таблицу ниже (см. табл. 2.1):

Таблица 2.1

Коэффициент множественной корреляции и его точечная и интервальные оценки.

Параметр/мера/оценка/показатель	Обозначение	Значение
Выборочный коэффициент множественной корреляции	$r_{yx_1x_2}$	0,70
Эмпирическое значение t-критерия	$t_{\text{эмп}}$	3,41
Критическое значение t-критерия	$t_{\text{кр}}$	2,36
Доверительный интервал для коэффициента корреляции (t-распределение Стьюдента; P=0,95).	$\rho_{YX_1X_2}$	(0,06; 1]
Доверительный интервал для коэффициента корреляции (z-преобразование Фишера; P=0,95).	$\rho_{YX_1X_2}$	(0,07; 0,93)