ELEMETRIX

Подробный отчёт

корреляционно-регрессионного анализа

С О Д Е Р Ж А Н И Е

Раздел 1. Корреляционный анализ исследуемых массивов данных.

1.1. Поиск предполагаемой линейной зависимости между двумя числовыми совокупностями предполагает, во-первых, установление факта наличия линейной связи между ними, оценку её тесноты и значимости, а во-вторых, выражение предполагаемой линейной зависимости в форме математической модели путём построения регрессионного уравнения.

1.1. Установление наличия линейной связи между массивами данных.

1.2. Первый показатель, позволяющий определить существует ли взаимосвязь у имеющихся массивов данных – это **коэффициент ковариации (cov_{xy})**, который рассчитывается разностью среднего арифметического произведений индивидуальных значений двух массивов данных и произведения их средних арифметических. Формула и вычисление даны ниже (1.1):

$$cov(x,y) = \overline{xy} - \bar{x} \cdot \bar{y} = 125{,}85 - 113{,}98 = 11{,}87 \tag{1.1}$$

1.3. Интерпретация коэффициента проста: если его значение положительно, то с ростом значений в первом массиве, значения второго также будут возрастать (прямая связь), а если знак коэффициента отрицателен – во втором массиве существует тенденция на убывание (обратная связь).

1.4. Однако только по абсолютному значению ковариации нет возможности судить о том, насколько сильна взаимосвязь двух совокупностей, так как коэффициент зависит от их дисперсий. Именно поэтому необходимо также найти значение выборочного **коэффициента парной линейной корреляции (r_{xy})**, равное отношению найденного выше коэффициента ковариации к произведению средних квадратических отклонений двух массивов. Формула и расчет даны ниже (1.2):

$$r_{xy} = \frac{cov(x,y)}{s_x \cdot s_y} = \frac{11{,}87}{15{,}05} = 0{,}79 \tag{1.2}$$

1.6. Коэффициент корреляции изменяется в пределах от минус единицы до единицы, следовательно, по положительному или отрицательному знаку судят о наличии прямой или обратной линейной связи.

1.7. А для простой характеристики тесноты связи задействуется **шкала Чеддока**, по которой модуль коэффициента корреляции сравнивается с неравенствами, приведенными ниже графически (см. рис. 1.1). Следует учитывать, что найденная линейная связь не обязательно носит причинно-следственный характер.

1.8. *Таким образом, линейную связь между анализируемыми массивами данных можно охарактеризовать как прямую и высокую*. Результаты вычислений в подразделе представлены ниже в табличном виде (см. табл. 1.1).

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. 1.1 Характеристика тесноты линейной связи по модулю коэффициента корреляции.

Таблица 1.1

Коэффициенты линейной связи анализируемых массивов данных.

Параметр/мера/оценка/показатель	Обозначение	Значение
Коэффициент ковариации	$cov(x, y)$	11,87
Коэффициент корреляции	r_{xy}	0,79

1.2. Точечная и интервальные оценки выборочной корреляции.

1.9. Следует сказать, что приведённая выше оценка носит лишь обобщенный характер и не даёт гарантий на вероятностную достоверность и репрезентативность. Именно поэтому следует проверить коэффициент корреляции на **статистическую значимость**, т.е. узнать обусловлена ли линейная связь в исследуемых массивах совершенной случайностью или же имеет место быть на самом деле. Для этого необходимо выдвинуть нулевую гипотезу (H_0), предполагающую, что выборочный коэффициент корреляции незначимо отличается от нуля (1.3):

$$H_0\colon\ r_{xy} = 0 \tag{1.3}$$

1.10. Чтобы проверить данную гипотезу, необходимо найти и сравнить друг с другом значения двух t-критериев – эмпирического и критического, формулы и нахождение которых даны ниже (1.5, 1.6). Неравенство, при котором нулевая гипотеза может быть отклонена, и с заданной доверительной вероятностью $P = 0{,}95$ (или на заданном уровне значимости $\alpha = 0{,}05$) можно предполагать статистическую значимость для коэффициента корреляции – это $t_{\text{эмп}} \geq t_{\text{кр}}$. Следует учитывать при этом, что число степеней свободы ($\boldsymbol{k}$) будет иметь на одну константу меньше, чем при оценке точности одной выборки (1.4):

$$k = n - m - 1 = 10 - 1 - 1 = 8 \tag{1.4}$$

$$t_{\text{эмп}} = \frac{r_{xy} \cdot \sqrt{k}}{\sqrt{1-r_{xy}^2}} = \frac{0,79 \cdot \sqrt{8}}{\sqrt{1-0,62}} = 3,63 \tag{1.5}$$

$$t_{\text{кр}} = t_{\text{кр}}(k; \alpha) = \ t_{\text{кр}}(8; 0,05) = 2,31 \tag{1.6}$$

1.11. *Так как* $3,63 > 2,31$ *(соблюдено неравенство* $t_{\text{эмп}} > t_{\text{кр}}$ *), нулевая гипотеза* H_0 *отклоняется, и, следовательно, с вероятностью* $P = 0,95$ *можно полагать, что коэффициент корреляции исследуемых массивов данных статистически значим.*

1.12. Для дальнейшей оценки точности выборочного коэффициента корреляции следует воспользоваться интервальным методом, т.е. построить **доверительный интервал** для истинного значения коэффициента корреляции или коэффициента корреляции двух генеральных совокупностей ($\boldsymbol{\rho_{XY}}$).

1.13. Первый способ построения интервала подразумевает задействование рассмотренного выше критического значения t-распределения Стьюдента с $k = 8$ степенями свободы. В случае, если значение нижней границы будет составлять меньше минус единицы, а верхней – больше единицы, то они будут доведены в ту или иную сторону до границ коэффициента корреляции, который, как упоминалось ранее, изменяется в пределах $[-1; 1]$. Формула построения интервала в двух видах (1.7, 1.8), вычисления (1.9, 1.10), а также графическое представление на прямой с отложением выборочного коэффициента корреляции на ней (см. рис. 1.2) приведены ниже:

$$\rho_{XY} \in r_{xy} \pm t_{\text{кр}} \cdot \sqrt{\frac{1-r_{xy}^2}{k}} \tag{1.7}$$

$$r_{xy} - t_{\text{кр}} \cdot \sqrt{\frac{1-r_{xy}^2}{k}} < \rho_{XY} \ < r_{xy} + t_{\text{кр}} \cdot \sqrt{\frac{1-r_{xy}^2}{k}} \tag{1.8}$$

$$0,79 \pm 2,31 \cdot \sqrt{\frac{1-0,62}{8}} = 0,79 \pm 0,50 \tag{1.9}$$

$$0,29 < \rho_{XY} < 1 \tag{1.10}$$

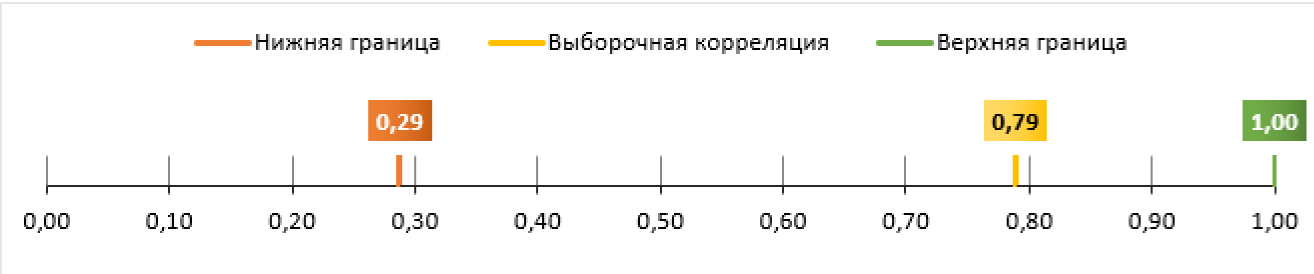

Рис. 1.2 Доверительный интервал для коэффициента корреляции (t-распределение; P=0,95).

1.14. *Следует предположить, что с* 95%*-ной вероятностью истинное значение коэффициента корреляции (или корреляция генеральных совокупностей) будет располагаться в пределах* (0,29; 1].

1.15. Второй способ интервальной оценки, преимущество которого в том, что он точнее первого (благодаря более узким границам интервала при той же доверительной вероятности), предполагает нахождение **нижней и верхней границ z-преобразования Фишера** ($\boldsymbol{z_L}$; $\boldsymbol{z_U}$) для выборочного коэффициента корреляции (1.11) и затем задействование их в построении доверительного интервала для корреляции генеральных совокупностей (1.12). Вычисление (1.13 – 1.16) и графическое представление интервала даны ниже (см. рис. 1.3):

$$z_L = \frac{1}{2} \cdot \ln\left(\frac{1+r_{xy}}{1-r_{xy}}\right) - \frac{1,96}{\sqrt{k-1}};\ z_U = \frac{1}{2} \cdot \ln\left(\frac{1+r_{xy}}{1-r_{xy}}\right) + \frac{1,96}{\sqrt{k-1}} \tag{1.11}$$

$$\frac{e^2 \cdot z_L - 1}{e^2 \cdot z_L + 1} < \rho_{XY} < \frac{e^2 \cdot z_U - 1}{e^2 \cdot z_U + 1} \tag{1.12}$$

$$z_L = \frac{1}{2} \cdot \ln\left(\frac{1+0,79}{1-0,79}\right) - \frac{1,96}{\sqrt{7}} = 0,33 \tag{1.13}$$

$$z_U = \frac{1}{2} \cdot \ln\left(\frac{1+0,79}{1-0,79}\right) + \frac{1,96}{\sqrt{7}} = 1,76 \tag{1.14}$$

$$\frac{7.39 \cdot 0,28 - 1}{7.39 \cdot 0,28 + 1} < \rho_{XY} < \frac{7.39 \cdot 1,76 - 1}{7.39 \cdot 1,76 + 1} \tag{1.15}$$

$$0,27 < \rho_{XY} < 0,94 \tag{1.16}$$

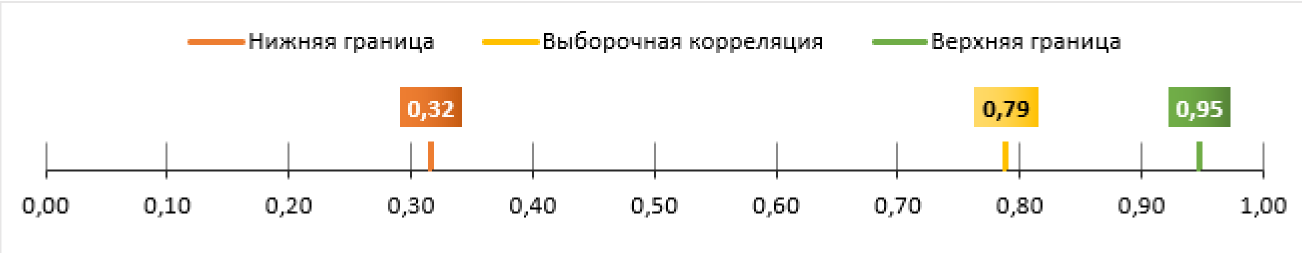

Рис. 1.3 Доверительный интервал для коэффициента корреляции (z-преобразование Фишера; P=0,95).

1.16. *Построенный интервал позволяет с вероятностью* $P = 0{,}95$ *предположить, что коэффициент корреляции генеральных совокупностей будет находится в пределах* $(0{,}27; 0{,}94)$. Все результаты вычислений в подразделе сведены в таблицу ниже (см. табл. 1.2):

Таблица 1.2

Точечная и интервальные оценки выборочного коэффициента корреляции.

Параметр/мера/оценка/показатель	Обозначение	Значение
Эмпирическое значение t-критерия	$t_{эмп}$	3,41
Критическое значение t-критерия	$t_{кр}$	2,31
Доверительный интервал для коэффициента корреляции (t-распределение Стьюдента; P=0,95).	ρ_{XY}	(0,25; 1]
Доверительный интервал для коэффициента корреляции (z-преобразование Фишера; P=0,95).	ρ_{XY}	(0,27; 0,94)

Раздел 2. Регрессионный анализ анализируемых массивов данных.

2.1. Под **парным регрессионным анализом** понимается метод моделирования и исследования зависимости **предиката** (результативного массива данных; y) от **предиктора** (иначе факторного массива; x).

2.2. Следует иметь в виду, что **регрессионные модели**, как и любые математические модели, являются абстрактными обобщениями или упрощенными представлениями о существующих в реальности процессах, явлениях или объектах. Следовательно, не следует ожидать, что регрессионное уравнение будет учитывать все существующие в природе нюансы и иметь поправку на каждое случайное возмущение.

2.2. **Линейная регрессия** – это простая и наиболее распространённая модель, используемая в регрессионном анализе. Её плюсы заключаются в легкости вычислительного процесса, простоте математических выводов, а также широкой применимости. А недостатки линейной регрессии проявляются в её высокой чувствительности к выбросам и в довольно вероятной ненадёжности прогнозов, составленных по ней (например, когда зависимость оказывается нелинейной).

2.1. Построение уравнения и нахождение параметров регрессии.

2.3. Зависимость предиката от предиктора в общих случаях следует считать вероятностной, а не функциональной, поэтому **теоретический** массив данных, т.е. массив из значений, вычисленных по уравнению регрессии, будет обозначатся $\boldsymbol{\hat{y}}$. Как видно из формулы построения выборочного уравнения парной линейной регрессии (2.1), необходимо для начала найти **параметры регрессии** ($\boldsymbol{b_0}$ и $\boldsymbol{b_1}$), которые, в свою очередь являются решением **системы нормальных уравнений** (2.2), выведенных **методом наименьших квадратов**. Вычисления и построенное уравнение также приведены ниже (2.3 – 2.5):

$$\hat{y} = b_0 + b_1 x \tag{2.1}$$

$$\begin{cases} b_0 \sum x_i + b_1 \sum x_i^2 = \sum x_i y_i \\ b_0 n + b_1 \sum x_i = \sum y_i \end{cases} \tag{2.2}$$

$$b_1 = \frac{cov(x,y)}{s_x^2} = \frac{11.87}{0.48} = 24{,}86 \quad (2.3)$$

$$b_0 = \bar{y} - b_1\bar{x} = 49{,}10 - 24{,}86 \cdot 2{,}32 = -8{,}60 \quad (2.4)$$

$$\hat{y} = -8{,}60 + 24{,}86 \cdot x_i \quad (2.5)$$

2.4. *Построенное выборочное уравнение регрессии позволяет предполагать, что при увеличении предиктора (x) на единицу, предикат (y) в среднем увеличится на 23,90, а прогнозируемый уровень предиката ($\hat{y}$) при нулевом значении предиктора ($x = 0$) примет значение –8,60.*

2.5. В случае необходимости приведения предиката и предиктора к одной условной единице (например, если они различаются единицами измерения) можно найти и использовать следующие **относительные параметры силы связи** – это, во-первых, **бета-коэффициент ($\boldsymbol{\beta}$)**, дающий понимание о том, на какую часть величины среднего квадратического изменится в среднем значение предиката при изменении предиктора на величину своего СКО, и, во-вторых, **средний коэффициент эластичности ($\overline{Э}$)**, показывающий, на сколько процентов в среднем меняется предикат при изменении предиктора на 1%. Формулы и вычисления даны ниже (2.6, 2.7):

$$\beta = b_1 \cdot \frac{s_x}{s_y} = 24{,}86 \cdot \frac{0{,}69}{19{,}38} = 0{,}79 \quad (2.6)$$

$$\overline{Э} = b_1 \cdot \frac{\bar{x}}{\bar{y}} = 24{,}86 \cdot \frac{2{,}32}{49{,}10} = 1{,}18\% \quad (2.7)$$

2.6. *Следовательно, увеличение предиктора на одну величину его СКО в среднем приводит к увеличению предиката на 0,79 его СКО. А относительный прирост предиката при увеличении предиктора на один процент в среднем составляет 1,18%.*

2.7. Найденную линейную зависимость предиката от предиктора графически можно отобразить в виде **корреляционного поля** (см. рис. 2.1), представляющего собой диаграмму рассеяния с отложенной линией регрессии. А результаты вычислений в подразделе сведены в таблицу ниже (см. табл. 2.1):

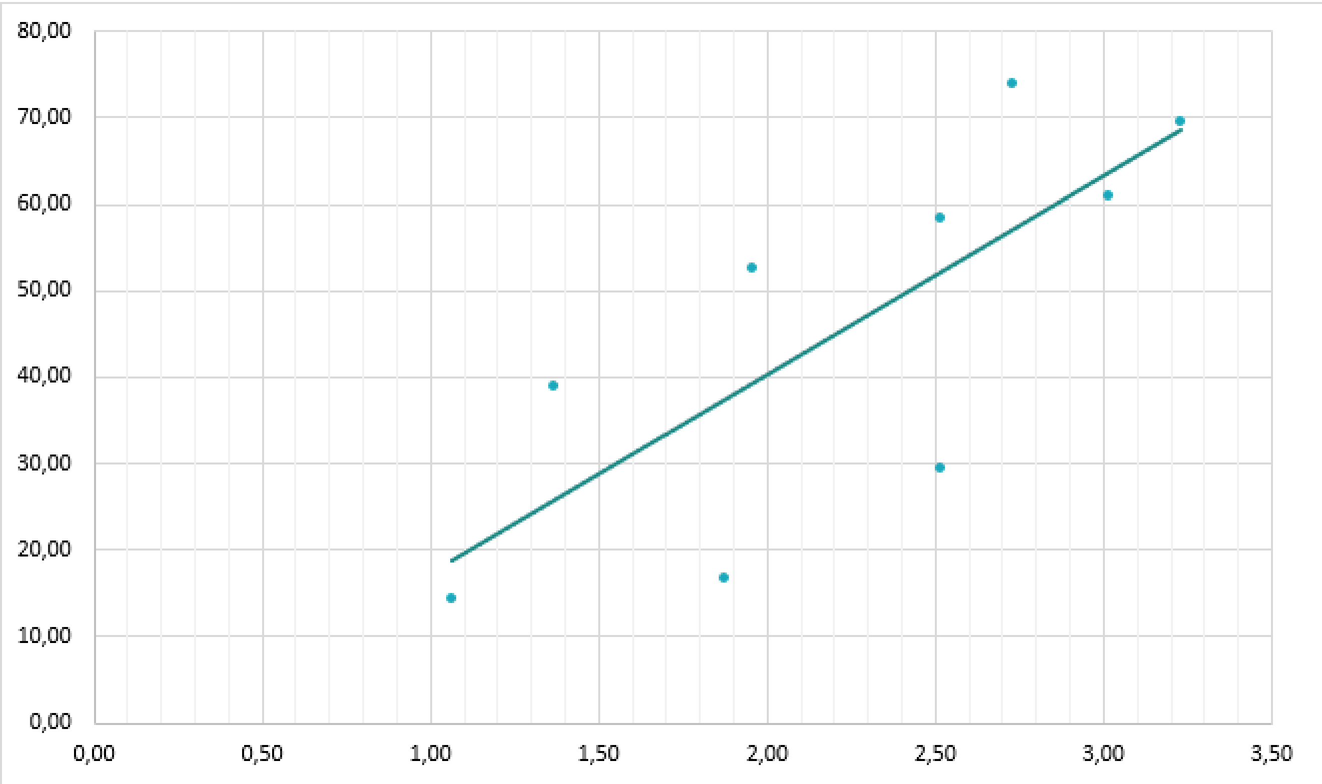

Рис. 2.1 Корреляционное поле для исследуемых массивов данных.

Таблица 2.1

Уравнение и параметры регрессионного анализа.

Уравнение парной линейной регрессии выборки	$\hat{y} = -8{,}60 + 24{,}86 \cdot x_i$	
Параметр/мера/оценка/показатель	Обозначение	Значение
Параметр b_1 уравнения регрессии	b_1	24,86
Параметр b_0 уравнения регрессии	b_0	–8,60
Бета-коэффициент	β	0,79
Средний коэффициент эластичности	$\overline{Э}$	1,18%

2.2. Точечные и интервальные оценки выборочного регрессионного уравнения и параметров регрессии.

2.8. При характеристике точности построенного уравнения линейной регрессии не обойтись без нахождения **остатков** – средних отклонений фактических значений предиката от вычисленных теоретических $(y_i - \hat{y}_i)$. Их значения и квадраты значений даны во вспомогательных таблицах, а графическое представление в виде гистограммы ниже (см. рис. 2.2):

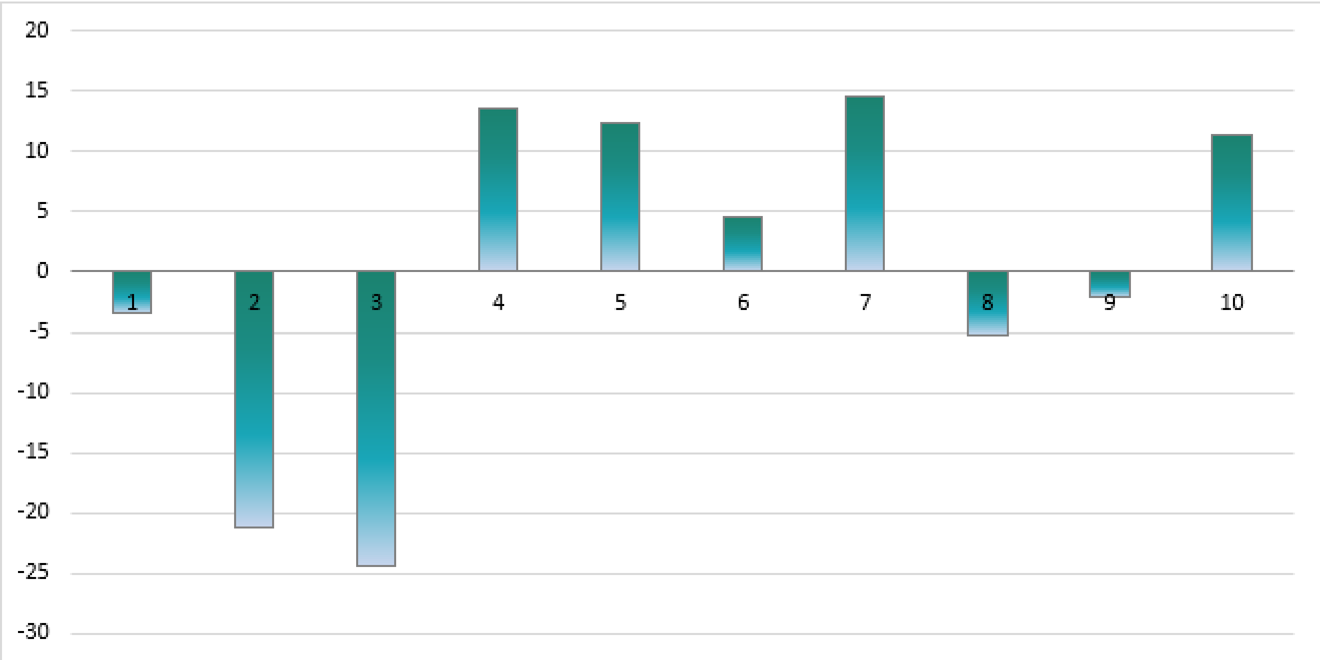

Рис. 2.2 Гистограмма остатков парной линейной регрессии.

2.9. Для характеристики качества подбора уравнения линейной регрессии следует рассмотреть несколько оценочных показателей, рассчитываемых по остаткам регрессии. Во-первых, **среднюю ошибку аппроксимации** $(\bar{A})$, представляющую собой среднее арифметическое модулей отношений остатков к фактическим значениям предикатного массива данных (2.8). Значения показателя, лежащие в промежутке $[0{,}05;\ 0{,}07]$, свидетельствуют о высокой точности подбора регрессионного уравнения, а максимально допустимым значением для него следует принять 0,15. Во-вторых, **исправленную остаточную дисперсию** $(\boldsymbol{S}_{\hat{y}}^{2})$, которая является несмещенной оценкой дисперсии остатков (2.9) по степеням свободы k.

Наконец, нахождением квадратного корня из неё будет получена наиболее полезная в дальнейшем анализе **стандартная ошибка регрессии** ($S_{\hat{y}}$; 2.10):

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y_i - \hat{y}_i}{y_i}\right| = \frac{3,46}{10} = 0,35 \quad (2.8)$$

$$S_{\hat{y}}^2 = \frac{1}{k}\sum(y_i - \hat{y}_i)^2 = \frac{1791,73}{8} = 223,97 \quad (2.9)$$

$$S_{\hat{y}} = \sqrt{S_{\hat{y}}^2} = \sqrt{223,97} = 14,97 \quad (2.10)$$

2.10. *Качество построенного уравнения регрессии оценивая средней ошибкой аппроксимации следует охарактеризовать как неудовлетворительное, т.к. значение ошибки более чем в два раза превышает допустимое* $(0,35 > 0,15)$. *Разброс фактических значений предикатного массива данных относительно теоретических в данной регрессионной модели составляет в среднем* $14,97$.

2.11. Для последующей оценки точности подбора уравнения регрессии и для ответа на вопрос о том, какая доля вариации предиката может быть объяснена изменчивостью предиктора, вычисляется **коэффициент детерминации,** рассчитываемый как отношение суммы квадратов отклонений теоретических, т.е. объясненных моделью значений, от среднего арифметического предиката к сумме квадратов отклонений предиката от средней (2.11), иными словами, отношение остаточной вариации к общей. Для упрощения формула нахождения сводится к квадрату вычисленного ранее выборочного коэффициента корреляции (2.12):

$$R_{xy}^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{2951,66}{4743,39} = 0,62 \quad (2.11)$$

$$R_{xy}^2 = r_{xy}^2 = 0,79^2 = 0,62 \quad (2.12)$$

2.12. *Точность подбора регрессионного уравнения по коэффициенту детерминации следует считать умеренной, т.к. вариация предиката на 62,23% может быть объяснена вариацией предиктора (доля независимой от предиктора вариации предиката составляет 37,77%).*

2.13. Коэффициент детерминации применяется также и в оценке статистической значимости уравнения линейной регрессии по эмпирическому и критическому **F-критериям**. Но сначала следует выдвинуть для него нулевую гипотезу (H_0), предполагающую случайную природу его формирования (2.13). Отклоняется она при соблюдении неравенства $F_{эмп} \geq F_{кр}$. Формулы и вычисления значений этих двух критериев, статистика которых имеет **F-распределение** (также называемое **распределением Фишера**), даны ниже (2.14, 2.15):

$$H_0\colon\ R_{xy}^2 = 0 \tag{2.13}$$

$$F_{эмп} = \frac{R_{xy}^2}{1-R_{xy}^2} \cdot \frac{k}{m} = \frac{0{,}59}{1-0{,}59} \cdot \frac{8}{1} = 11{,}64 \tag{2.14}$$

$$F_{кр} = F_{кр}(m; k; \alpha) = F_{кр}(1; 8; 0{,}05) = 5{,}32 \tag{2.15}$$

2.14. *В связи с тем, что* $11{,}64 > 5{,}32$ *(соблюдается неравенство* $F_{эмп} > F_{кр}$*), нулевая гипотеза отклоняется, и, следовательно, следует предполагать, что коэффициент детерминации и построенное выборочное уравнение регрессии статистически значимы с доверительной вероятностью* $P = 0{,}95$.

2.15. Для точечной оценки параметров b_1 и b_0 регрессионного уравнения вычисляются их **стандартные ошибки,** задействующие найденную ранее стандартную ошибку регрессии. Формулы и вычисления даны ниже (2.16, 2.17):

$$S_{b_1} = \frac{S_{\hat{y}}}{\sqrt{\sum(x_i-\bar{x}_i)^2}} = \frac{14{,}97}{\sqrt{4{,}78}} = 6{,}85 \tag{2.16}$$

$$S_{b_0} = \sqrt{\frac{S_{\hat{y}}^2 \cdot \sum x_i^2}{n\sum(x_i-\bar{x}_i)^2}} = \sqrt{\frac{46{,}89 \cdot 58{,}66}{47{,}76}} = 16{,}59 \tag{2.17}$$

2.16. Проверка статистической значимости параметра b_1 подразумевает выдвижение для него нулевой гипотезы (H_0; 2.18), затем нахождение эмпирического t-критерия Стьюдента (t_{b_1}), вычисляемого отношением модуля параметра к его стандартной ошибке (2.19), и наконец, сравнение полученного результата с найденным по таблице критическим значением t-критерия (2.20):

$$H_0\colon\ b_1 = 0 \tag{2.18}$$

$$t_{b_1} = \frac{|b_1|}{S_{b_1}} = \frac{24{,}86}{6{,}85} = 3{,}63 \tag{2.19}$$

$$t_{кр} = t_{кр}(k;\alpha) = t_{кр}(8;0{,}05) = 2{,}31 \tag{2.20}$$

2.17. *Так как* $3{,}63 > 2{,}31$ $(t_{b_1} > t_{кр})$, *нулевая гипотеза отклоняется, и с* 95%-*ной вероятностью можно полагать, что параметр* b_1 *выборочного уравнения регрессии статистически значим.*

2.18. Аналогичным образом следует выдвинуть и осуществить проверку нулевой гипотезы (H_0) для параметра b_0 (2.21, 2.22):

$$H_0\colon b_0 = 0 \tag{2.21}$$

$$t_{b_0} = \frac{|b_0|}{S_{b_0}} = \frac{8{,}60}{16{,}59} = 0{,}52 \tag{2.22}$$

2.19. *Исходя из полученного неравенства* $0{,}52 < 2{,}31$ *(* $t_{b_0} < t_{кр}$ *), нулевая гипотеза не отклоняется и подтверждается случайная природа формирования (статистическая незначимость) параметра* b_0 *выборочного уравнения регрессии с вероятностью* $P = 0{,}95$.

2.20. Интервальная оценка может не только подтвердить вышесказанное о параметрах b_1 и b_0, но и определить границы числового множества для параметров $\boldsymbol{\beta_1}$ и $\boldsymbol{\beta_0}$ регрессионного уравнения генеральных совокупностей ($\boldsymbol{\widehat{Y}}$; 2.23), в котором они будут располагаться с доверительной вероятностью $P = 0{,}95$.

$$\widehat{Y} = \beta_0 + \beta_1 X_1 \tag{2.23}$$

2.21. Формула построения доверительного интервала для параметра b_1 в двух видах (2.24, 2.25), графическое представление (см. рис. 2.3) и вычисления приведены ниже (2.26, 2,27):

$$\beta_1 \in b_1 \pm t_{кр} \cdot S_{b_1} \tag{2.24}$$

$$b_1 - t_{кр} \cdot S_{b_1} < \beta_1 < b_1 + t_{кр} \cdot S_{b_1} \tag{2.25}$$

$$24{,}86 \pm 2{,}31 \cdot 6{,}85 = 24{,}86 \pm 15{,}79 \qquad (2.\ 26)$$

$$9{,}07 < \beta_1 < 40{,}65 \qquad (2.\ 27)$$

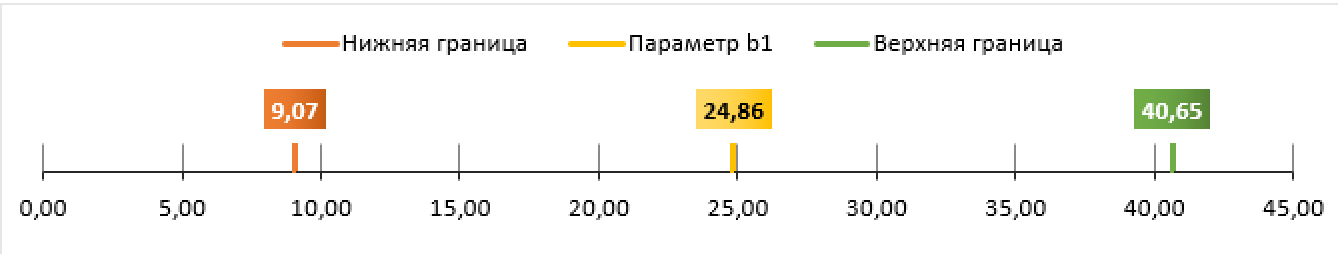

Рис. 2.3 Доверительный интервал для параметра регрессии β_1
(t-распределение Стьюдента; $P = 0{,}95$).

2.22. *Построенный интервал подтверждает значимость параметра* b_1 *выборочного уравнения регрессии и с доверительной вероятностью* $P = 0{,}95$ *выявляет границы для параметра* β_1 *уравнения регрессии генеральных совокупностей в открытом числовом множестве* $(9{,}06; 40{,}65)$.

2.23. Аналогичным образом ниже проведена интервальная оценка параметра b_0 выборочного уравнения регрессии (2.28 – 2.31; см. рис. 2.4):

$$\beta_0 \in b_0 \pm t_{кр} \cdot S_{b_0} \tag{2.28}$$

$$b_0 - t_{кр} \cdot S_{b_0} < \beta_0 < b_0 + t_{кр} \cdot S_{b_0} \tag{2.29}$$

$$-8{,}60 \pm 2{,}31 \cdot 16{,}59 = -8{,}60 \pm 38{,}25 \tag{2.30}$$

$$-46{,}85 < \beta_0 < 29{,}64 \tag{2.31}$$

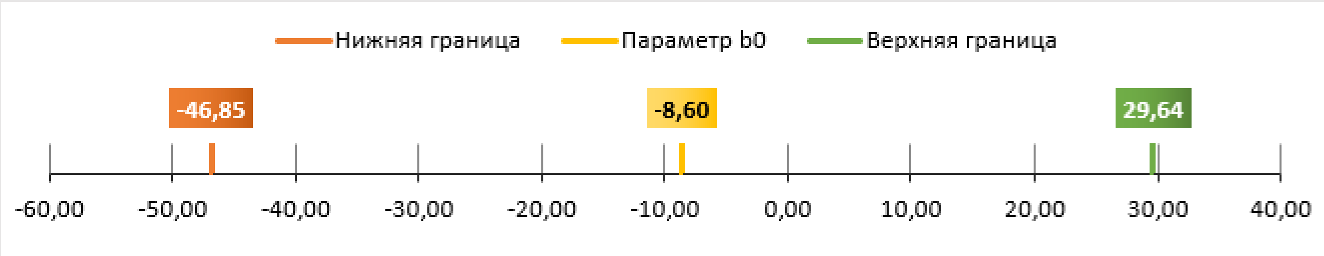

Рис. 2.4 Доверительный интервал для параметра регрессии β_0 (t-распределение Стьюдента; $P = 0{,}95$).

2.24. *Статистическая незначимость параметра* b_0 *выборочного уравнения регрессии подтверждена, так как точка* 0 *лежит внутри доверительного интервала, и параметр* β_0 *уравнения регрессии генеральной совокупности с* 95%*-ной вероятностью лежит в открытом числовом промежутке* $(-46{,}85; 29{,}64)$. *В заключение, следует сказать, что полученная выборочная регрессионная модель для анализируемых массивов данных недостаточно репрезентативна и нежелательна в использовании*. Все результаты вычислений в подразделе приведены ниже в табличном виде (см. табл. 2.2):

Таблица 2.2

Точечные и интервальные оценки выборочного уравнения регрессии.

Параметр/мера/оценка/показатель	Обозначение	Значение
Средняя ошибка аппроксимации	$\bar{A}$	0,35
Исправленная остаточная дисперсия	$S_{\hat{y}}^2$	223,97
Стандартная ошибка регрессии	$S_{\hat{y}}$	14,97
Коэффициент детерминации	R_{xy}^2	59,26%
Эмпирическое значение F-критерия	$F_{эмп}$	11,64
Критическое значение F-критерия	$F_{кр}$	5,32
Стандартная ошибка параметра b_1	S_{b_1}	6,85
Стандартная ошибка параметра b_0	S_{b_0}	16,59
Эмпирическое значение t-критерия для b_1	t_{b_1}	3,63
Эмпирическое значение t-критерия для b_0	t_{b_0}	0,52
Критическое значение t-критерия	$t_{кр}$	2,31
Доверительный интервал для параметра β_1	β_1	(9,07; 40,65)
Доверительный интервал для параметра β_0	β_0	(-46,85; 29,64)