ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ ИМПЕРАТОРА НИКОЛАЯ II»

КАФЕДРА «МАТЕМАТИКА»

М.В. Ишханян, Н.В. Карпенко

ЭКОНОМЕТРИКА

ЧАСТЬ I

ПАРНАЯ РЕГРЕССИЯ

Учебное пособие

Москва – 2016

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ ИМПЕРАТОРА НИКОЛАЯ II»

КАФЕДРА «МАТЕМАТИКА»

М.В. Ишханян, Н.В. Карпенко

ЭКОНОМЕТРИКА

ЧАСТЬ I

ПАРНАЯ РЕГРЕССИЯ

Рекомендовано редакционно-издательским советом
университета в качестве учебного пособия
для студентов направления 380301 «Экономика»

Москва – 2016

УДК – 330.43(076.5)
И – 97

Ишханян М.В., Карпенко Н.В. Эконометрика. Часть 1. Парная регрессия: Учебное пособие. – М.: МГУПС (МИИТ), 2016. – 117 с.

Учебное пособие предназначено для студентов направления «Экономика», изучающих дисциплину «Эконометрика». Содержит краткие методические указания, решения типовых задач, описание реализации на компьютере с помощью MS Excel, а также контрольные задания. Может быть использовано для проведения практических занятий, а так же для организации самостоятельной работы студентов.

Рецензенты:
В.А. Шведовский, д.соц.н, к.ф.-м.н., доцент кафедры «Прикладная математика» Высшей школы современных социальных наук, МГУ им. М.В. Ломоносова

О.А. Платонова, к.ф.-м.н., доцент, зав. кафедрой «Высшая и вычислительная математика» МИИТа;

Оглавление

Введение

Развитие экономики, усложнение взаимодействия экономических субъектов требуют тщательного и объективного анализа реально протекающих процессов. Эконометрика - один из основных инструментов решения этой задачи.

Экономико-математическое моделирование - основной метод исследования в эконометрике. Правильно построенная модель качественно и количественно оценивает величину изменения изучаемого явления или процесса в зависимости от изменения внешней среды.

Применение методов эконометрики позволяет выявить существующие связи между параметрами модели и внешними факторами, дать обоснованный прогноз развития в заданных условиях, проверить и численно оценить экономические последствия принимаемых управленческих решений.

Важными составляющими успешного моделирования являются экономически адекватная постановка задачи, подготовка качественных исходных данных, корректное применение математического аппарата, четкое понимание областей применимости используемых методов, правильная интерпретация полученных результатов. В настоящее время эконометрическое моделирование невозможно без использования средств вычислительной техники, современных статистических пакетов прикладных программ.

1. Парная корреляция

1.1. Понятие корреляционной зависимости

Экономические явления характеризуются множеством признаков, отражающих те или иные их свойства. Эти признаки изменяются (варьируются) во времени и пространстве. Нередко изменения признаков взаимозависимы и взаимообусловлены.

Различают два типа зависимости между признаками: функциональную и статистическую или стохастическую.

Статистическая зависимость - это связь, при которой каждому значению *независимой* переменной ***x*** (*факторного признака*) соответствует множество значений *зависимой* переменной ***y*** (*результативного признака*), причем неизвестно заранее, какое именно значение примет ***y***. Статистическая зависимость - свойство совокупности наблюдений в целом, а не отдельных ее единиц.

Частным случаем статистической зависимости является корреляционная зависимость.

Корреляционная зависимость или ***корреляция*** - это связь, при которой каждому значению независимой переменной ***x*** соответствует определенное математическое ожидание (среднее значение) зависимой переменной ***y***.

Корреляционная связь проявляется не в каждом отдельном случае, а только в средних величинах при достаточно большом числе статистических наблюдений.

Корреляционная связь бывает ***прямая*** (с увеличением (уменьшением) значений факторного признака ***x*** происходит увеличение (уменьшение) значений результативного признака ***y***) и ***обратная*** (с увеличением (уменьшением) значений факторного признака ***x*** происходит уменьшение (увеличение) значений результативного признака ***y***).

По аналитическому выражению корреляционная зависимость может быть прямолинейной ***(линейной)*** и криволинейной ***(нелинейной)***.

В случае линейной зависимости с возрастанием величины факторного признака происходит равномерное возрастание (или убывание) величин результативного признака (выражается уравнением прямой линии).

При криволинейной зависимости с возрастанием величины факторного признака возрастание (или убывание) результативного признака происходит неравномерно (выражается уравнениями кривых линий).

В зависимости от количества факторных признаков корреляционные связи делят на ***однофакторные*** (связь между одним признаком-фактором (***x***) и результативным признаком (***y***)) и ***многофакторные*** (связь между несколькими факторными признаками и результативным признаком).

Корреляционная зависимость исследуется с помощью методов корреляционного анализа.

Корреляционный анализ - это раздел математической статистики, посвященный изучению взаимосвязей между случайными величинами. Применяется тогда, когда данные наблюдений можно считать случайными и выбранными из генеральной совокупности, распределенной по нормальному закону.

Корреляционный анализ заключается в количественном определении тесноты связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

1.2. Коэффициент парной корреляции

Теснота связи между двумя признаками $\boldsymbol{x}$ и $\boldsymbol{y}$ количественно выражается величиной ***коэффициента парной корреляции*** $\boldsymbol{r_{xy}}$

$$r_{xy} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{n \cdot \sigma_x \cdot \sigma_y} \tag{1.1}$$

где $\boldsymbol{n}$ - объем исследуемой совокупности (объем выборки); $\bar{x}$, $\bar{y}$ - средние значения; $\boldsymbol{\sigma_x^2}$, $\boldsymbol{\sigma_y^2}$ – дисперсии; $\boldsymbol{\sigma_x}, \boldsymbol{\sigma_y}$ – среднеквадратические (стандартные) отклонения признаков $\boldsymbol{y}$, $\boldsymbol{x}$, определяются по формулам:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\ , \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\ ; \tag{1.2}$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2, \ \sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2, \tag{1.3}$$

$$\sigma_x = \sqrt{\sigma_x^2}\ , \ \sigma_y = \sqrt{\sigma_y^2}\ . \tag{1.4}$$

Коэффициент парной корреляции $\boldsymbol{r_{yx}}$ служит мерой ***линейной корреляционной зависимости между величи-***

нами y и x , при условии, что на формирование их значений оказывают влияние некоторые другие, неучтенные факторы.

Теснота ***нелинейной зависимости*** определяется с помощью ***индекса корреляции*** $\boldsymbol{\rho_{xy}}$ (корреляционного отношения).

Для расчета коэффициента парной корреляции также можно использовать следующие формулы:

$$r_{xy} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y}, \tag{1.5}$$

где $\overline{y \cdot x}$ - средняя арифметическая произведения двух признаков $\overline{y \cdot x} = \frac{1}{n}\sum_{i=1}^{n}(y_i \cdot x_i)$.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1.6}$$

1.3. Свойства коэффициента парной корреляции

1. $-1 \le \boldsymbol{r_{xy}} \le 1$.

2. $\boldsymbol{r_{xy}} = \boldsymbol{r_{yx}}$.

3. Если признаки $\boldsymbol{x}$ и $\boldsymbol{y}$ независимы, то $\boldsymbol{r_{xy}} = 0$.

4. Из равенства $\boldsymbol{r_{xy}} = 0$ не следует независимость величин $\boldsymbol{x}$ и $\boldsymbol{y}$, ***отрицается лишь их линейная корреляционная зависимость*** (может существовать нелинейная зависимость).

5. Значения $\boldsymbol{r_{xy}} = -1$, $\boldsymbol{r_{xy}} = 1$ соответствуют практически функциональной линейной связи между величинами $\boldsymbol{x}$ и $\boldsymbol{y}$.

6. При $\boldsymbol{r_{xy}} > 0$ величины $\boldsymbol{x}$ и $\boldsymbol{y}$ одновременно возрастают (прямая зависимость). При $\boldsymbol{r_{xy}} < 0$ с возрастанием величины $\boldsymbol{x}$ ($\boldsymbol{y}$) величина $\boldsymbol{y}$ ($\boldsymbol{x}$) убывает (обратная зависимость).

Для определения тесноты (силы) линейной корреляционной зависимости можно использовать шкалу Чеддока (табл. 1.1).

Таблица 1.1. Шкала Чеддока

Значение $\lvert r_{yx} \rvert$	**Характер линейной корреляционной связи между признаками y и x**
0 - 0,2	Связь практически отсутствует
0,2 - 0,5	Связь слабая
0,5 - 0,7	Связь средняя (умеренная)
0,7 - 0,95	Связь сильная (тесная)
0,95 - 1	Практически функциональная зависимость

1.4. Оценка статистической значимости коэффициента парной корреляции

Коэффициент парной корреляции $\boldsymbol{r_{xy}}$ как статистическую величину необходимо оценить на достоверность (***статистическую значимость***).

Для оценки значимости коэффициента парной корреляции используют ***t-критерий Стьюдента***, который применяется при ***t***-распределении, отличном от нормального, но приближающемся к нормальному при n > 30.

Находится расчетное значение (статистика) критерия

$$t_{расч} = \frac{r_{xy}}{\sqrt{1-r_{xy}^2}} \cdot \sqrt{n-m-1}, \qquad (1.7)$$

которое сравнивается с табличным (критическим) значением $\boldsymbol{t_{табл}}$, найденным по таблице критических значений распределения Стьюдента (табл. Приложения 2, двухсторонняя критическая область) по уровню значимости $\boldsymbol{\alpha}$ и числу степеней свободы $\boldsymbol{df} = \boldsymbol{n} - \boldsymbol{m} - 1$, или с помощью встроенной функции Excel «СТЬЮДРАСПОБР».

В экономических расчетах значение $\boldsymbol{\alpha}$ обычно принимается равным 0,05.

Если $|t_{расч}| > t_{табл}$, то ***коэффициент парной корреляции*** признается ***статистически значимым***. Если $|t_{расч}| < t_{табл}$, коэффициент корреляции считается незначимым, случайно отличающимся от нуля.

1.5. Интервальная оценка коэффициента парной корреляции

Для коэффициента парной корреляции $\boldsymbol{r_{xy}}$ находится его интервальная оценка, в качестве которой служит ***доверительный интервал***, определенный с надежностью (вероятностью) $\boldsymbol{\gamma} = 1 - \boldsymbol{\alpha}$:

$$r_{xy} - m_r \cdot t_{табл} < r < r_{xy} + m_r \cdot t_{табл}$$

или (1.8)

$$\left(r_{xy} - m_r \cdot t_{табл}\ ;\ r_{xy} + m_r \cdot t_{табл}\right).$$

Здесь $\boldsymbol{t_{табл}}$ - значение, найденное по таблице критических точек распределения Стъюдента по уровню значимости $\boldsymbol{\alpha}$ и числу степеней свободы $\boldsymbol{df} = \boldsymbol{n} - \boldsymbol{m} - 1$; $\boldsymbol{m_r}$ - стандартная ошибка коэффициента корреляции, равная

$$m_r = \sqrt{\frac{1 - r_{xy}^2}{n - m - 1}}\ . \tag{1.9}$$

С вероятностью γ доверительный интервал покрывает истинное значение коэффициента парной корреляции. Если границы доверительного интервала имеют разные знаки, $\boldsymbol{r_{xy}}$ статистически незначим.

Использование выражения (1.8) для получения интервальной оценки коэффициента парной корреляции целесообразно, когда выборка имеет достаточно большой объем ($\boldsymbol{n} \geq 30$) и значение $\boldsymbol{r_{xy}}$ велико ($|r_{xy}| \geq 0{,}7$).

Если эти условия не выполняются, для построения доверительного интервала используют ***z*-преобразование Фишера**

$$z(r) = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = \frac{1}{2}\left(\ln(1+r) - \ln(1-r)\right) \qquad (1.10)$$

Функция ***z*** нечетная, т.е. $z(-r) = -z(r)$.

Предварительно устанавливается интервальная оценка для ***z*** в виде

$$z' - t_{\gamma}\sqrt{\frac{1}{n-l-3}} \le z \le z' + t_{\gamma}\sqrt{\frac{1}{n-m-3}} \qquad (1.11)$$

где ***l*** - порядок коэффициента корреляции (для коэффициента парной корреляции ***l*** = 2); ***n*** - объем выборки, ***γ*** - вероятность выполнения неравенства; t_γ вычисляется по таблице значений интегральной функции Лапласа из условия $\Phi(t_{\gamma}) = \gamma$.

Значение ***z'*** определяется по таблице ***z***-преобразования (табл. Приложения 1) по найденному значению r_{xy} или с помощью встроенной функции Excel «ФИШЕР». Обратный переход от ***z*** к r_{xy} осуществляется также по таблице ***z***-преобразования либо с помощью встроенной функции Excel «ФИШЕРОБР», после чего получают интервальную оценку r_{xy} с надежностью $\gamma = 1 - \alpha$:

$$r_{min} \le r_{xy} \le r_{max} \qquad (1.12)$$

1.6. Коэффициент детерминации

На основе коэффициента парной корреляции рассчитывается ***коэффициент детерминации R²***:

$$R^2 = \left(r_{yx}\right)^2 \quad (0 \le R^2 \le 1) \tag{1.13}$$

Коэффициент детерминации оценивает долю вариации признака ***y*** , обусловленную изменением значений признака ***x***. Чем ближе значение $\boldsymbol{R^2}$ к единице, тем больше признак ***x*** участвует в формировании значений ***y***.

Проверка статистической значимости коэффициента детерминации (как в случае парной, так и множественной корреляции) производится с помощью ***F-критерия Фишера***.

Находится расчетное значение (статистика) критерия

$$F_{расч} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} \tag{1.14}$$

где ***n*** - объем выборки, ***m*** - число независимых переменных (в случае парной корреляции ***m*** = 1).

Значение $\boldsymbol{F_{расч}}$ сравнивается с табличным (критическим) значением $\boldsymbol{F_{табл}}$, найденным по таблице критических значений распределения Фишера-Снедекора (F-распределения) по уровню значимости $\boldsymbol{\alpha}$ и двум числам степеней свободы $\boldsymbol{df_1 = m}$ и $\boldsymbol{df_2 = n - m} - 1$ (табл. Приложения 3) или с помощью встроенной функции Excel «FРАСПОБР».

Если $\boldsymbol{F_{расч}} > \boldsymbol{F_{табл}}$, то коэффициент детерминации $\boldsymbol{R^2}$ признается статистически значимым. В противном случае ($\boldsymbol{F_{расч}} < \boldsymbol{F_{табл}}$) $\boldsymbol{R^2}$ статистически незначим.

1.7. Решение типовых задач

Задача 1.1. Имеются выборочные данные (табл. 1.2) показателей «Среднедушевой денежный доход» (x, руб.) и «Среднедушевой оборот розничной торговли» (y, руб.).

Таблица 1.2. Исходные данные

№ города	Среднедушевой денежный доход	Среднедушевой оборот розничной торговли	№ города	Среднедушевой денежный доход	Среднедушевой оборот розничной торговли
1	3357	2425	9	3724	2225
2	3135	2050	10	3416	1983
3	2842	1683	11	3340	1925
4	3991	2375	12	3089	1042
5	2293	1167	13	4372	2925
6	3563	2200	14	3022	2342
7	3219	1892	15	3383	2458
8	3308	2008	16	4267	2125

Требуется:

1) найти коэффициент парной корреляции между показателями;
2) найти коэффициент детерминации;
3) определить статистическую значимость коэффициентов парной корреляции и детерминации;
4) построить интервальную оценку коэффициента парной корреляции.

Решение

Объем статистической совокупности n = 16.

1) Найдем коэффициент парной корреляции.

1 способ. Для вычисления коэффициента парной корреляции воспользуемся формулой (1.6):

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

Составим расчетную таблицу (см. табл. 1.3) (расчеты выполнены в Excel). Средние значения признаков равны

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{54321}{16} = 3395{,}1; \; \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{32825}{16} = 2051{,}6.$$

Коэффициент парной корреляции равен

$$r_{xy} = \frac{2590081{,}4}{\sqrt{4070770{,}9 \cdot 3299597{,}9}} = 0{,}707.$$

Величина коэффициента корреляции, находящаяся в пределах от 0,7 до 0,95 согласно шкале Чеддока свидетельствует о ***тесной связи*** (близкой к умеренной) ***между среднедушевым денежным доходом и среднедушевым оборотом розничной торговли***. Поскольку значение коэффициента парной корреляции положительно, эта ***связь прямая***.

Таблица 1.3. Расчетная таблица коэффициента парной корреляции

№ п/п	x_i	y_i	$(x_i - \overline{x})$	$(y_i - \overline{y})$	$(x_i - \overline{x})(y_i - \overline{y})$	$(x_i - \overline{x})^2$	$(y_i - \overline{y})^2$
1,0	3357,0	2425,0	-38,1	373,4	-14214,0	1448,8	139455,6
2,0	3135,0	2050,0	-260,1	-1,6	406,3	67632,5	2,4
3,0	2842,0	1683,0	-553,1	-368,6	203838,1	305878,1	135838,3
4,0	3991,0	2375,0	595,9	323,4	192748,5	355141,5	104611,8
5,0	2293,0	1167,0	-1102,1	-884,6	974843,2	1214541,8	782450,8
6,0	3563,0	2200,0	167,9	148,4	24928,2	28203,0	22033,7
7,0	3219,0	1892,0	-176,1	-159,6	28093,0	30998,0	25460,2
8,0	3308,0	2008,0	-87,1	-43,6	3792,7	7579,9	1897,7
9,0	3724,0	2225,0	328,9	173,4	57050,1	108199,9	30080,6
10,0	3416,0	1983,0	20,9	-68,6	-1435,5	438,4	4700,8
11,0	3340,0	1925,0	-55,1	-126,6	6968,8	3031,9	16018,1
12,0	3089,0	1042,0	-306,1	-1009,6	308989,2	93674,3	1019216,4
13,0	4372,0	2925,0	976,9	873,4	853293,8	954406,9	762893,1
14,0	3022,0	2342,0	-373,1	290,4	-108351,3	139175,6	84353,9
15,0	3383,0	2458,0	-12,1	406,4	-4902,7	145,5	165191,4
16,0	4267,0	2125,0	871,9	73,4	64032,9	760275,0	5393,1
Сумма	**54321,0**	**32825,0**			**2590081,4**	**4070770,9**	**3299597,9**

2 способ. Для вычисления коэффициента парной корреляции используем формулу (1.5).

$$r_{xy} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y} = \frac{7127063{,}0 - 2051{,}6 \cdot 3395{,}1}{504{,}4 \cdot 454{,}1} = 0{,}707,$$

где среднеквадратические отклонения σ_x, σ_y переменных ***x***, ***y*** равны:

$$\sigma_x = \sqrt{\overline{x^2} - \left(\bar{x}\right)^2} = \sqrt{11780872{,}6 - (3395{,}1)^2} = 504{,}4$$

$$\sigma_y = \sqrt{\overline{y^2} - \left(\bar{y}\right)^2} = \sqrt{4415133{,}6 - (2051{,}6)^2} = 454{,}1$$

Необходимые расчеты приведены в табл. 1.4.

Таблица 1.4. Расчетная таблица коэффициента парной корреляции

№ п/п	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	3357	2425	11269449	5880625	8140725
2	3135	2050	9828225	4202500	6426750
3	2842	1683	8076964	2832489	4783086
4	3991	2375	15928081	5640625	9478625
5	2293	1167	5257849	1361889	2675931
6	3563	2200	12694969	4840000	7838600
7	3219	1892	10361961	3579664	6090348
8	3308	2008	10942864	4032064	6642464
9	3724	2225	13868176	4950625	8285900
10	3416	1983	11669056	3932289	6773928
11	3340	1925	11155600	3705625	6429500
12	3089	1042	9541921	1085764	3218738
13	4372	2925	19114384	8555625	12788100
14	3022	2342	9132484	5484964	7077524
15	3383	2458	11444689	6041764	8315414
16	4267	2125	18207289	4515625	9067375
Среднее	**3395,1**	**2051,6**	**11780872,6**	**4415133,6**	**7127063,0**

2) Найдем коэффициент детерминации

$$R^2 = \left(r_{yx}\right)^2 = (0{,}707)^2 = 0{,}4998.$$

Значение $\boldsymbol{R^2}$ показывает, что ***на 49,98% изменение оборота розничной торговли объясняется изменениями денежных доходов населения***. Оставшиеся 50,02% приходятся на другие факторы.

3) Статистическую значимость коэффициента парной корреляции $\boldsymbol{r_{xy}}$ проверим с помощью t-критерия Стьюдента. Найдем статистику

$$t_{расч} = \frac{r_{xy}}{\sqrt{1-r_{xy}^2}} \cdot \sqrt{n-m-1} = \frac{0{,}707}{\sqrt{1-0{,}4998}}\sqrt{16-1-1} = 3{,}738$$

(m = 1).

Табличное значение t-критерия Стьюдента при уровне значимости $\boldsymbol{\alpha}$ = 0,05 и числе степеней свободы $\boldsymbol{df} = \boldsymbol{n} - \boldsymbol{m} - 1$ = 16 - 1 - 1 = 14 составляет $\boldsymbol{t_{табл}}$ (0,05; 14) = 2,145. (табличное значение критерия найдено с помощью встроенной функции Excel «СТЬЮДРАСПОБР»).

Так как $|\boldsymbol{t_{расч}}| > \boldsymbol{t_{табл}}$ (3,738 > 2,145) , ***значение парного коэффициента корреляции статистически значимо***. Статистическая значимость коэффициента корреляции подтверждает тесную связь (близкую к умеренной) между среднедушевым денежным доходом и среднедушевым оборотом розничной торговли.

Статистическую значимость коэффициента детерминации проверим с помощью критерия Фишера. Найдем статистику

$$F_{расч} = \frac{R^2}{1-R^2} \cdot \frac{n \; - m \; -1}{m} = \frac{0{,}4998}{1-0{,}4998} \cdot \frac{16-1-1}{1} = 13{,}99$$

Табличное значение F-критерия Фишера при уровне значимости $\boldsymbol{\alpha}$ = 0,05 и числах степеней свободы $\boldsymbol{df}_1$ = $\boldsymbol{m}$ = 1 и $\boldsymbol{df}_2$ = $\boldsymbol{n}$ - $\boldsymbol{m}$ - 1 = 16 - 1 - 1 = 14 составляет $\boldsymbol{F_{табл}}$ (0,05; 1; 14) = 4,60. (табличное значение критерия найдено с помощью встроенной функции Excel «FРАСПОБР»).

Так как $\boldsymbol{F_{расч}} > \boldsymbol{F_{табл}}$ (13,99 > 4,60***), коэффициент детерминации признается статистически значимым***.

4) Построим интервальную оценку парного коэффициента корреляции (доверительный интервал)

$$r_{xy} - m_r \cdot t_{\text{табл}} < r < r_{xy} + m_r \cdot t_{\text{табл}}$$

Стандартная ошибка коэффициента корреляции равна

$$m_r = \sqrt{\frac{1-r_{xy}^2}{n-m-1}} = \sqrt{\frac{1-0{,}4998}{16-1-1}} = \sqrt{0{,}0357} = 0{,}189.$$

Табличное значение t-критерия Стьюдента при уровне значимости $\boldsymbol{\alpha}$ = 0,05 и числе степеней свободы $\boldsymbol{df}$ = $\boldsymbol{n}$ - $\boldsymbol{m}$ - 1 = 16 - 1 - 1 = 14 составляет $\boldsymbol{t_{табл}}$ (0,05; 14) = 2,145.

Доверительный интервал имеет вид

$$0{,}707 - 0{,}189 \cdot 2{,}145 < r < 0{,}707 + 0{,}189 \cdot 2{,}145$$

или

$$0{,}301 < r < 1{,}112$$

Т.е. с вероятностью γ = 1 - α = 1 - 0,05 = 0,95 значение коэффициента корреляции генеральной совокупности находится внутри интервала (0,301; 1,112).

Границы доверительного интервала имеют одинаковые знаки, что подтверждает вывод критерия Стьюдента о статистической значимости коэффициента парной корреляции.

Задача 1.2. (Пример применения ***z-преобразования Фишера*** для построения доверительного интервала для парного коэффициента корреляции).

По двумерной выборке значений признаков x и y объема $n = 8$ найден коэффициент корреляции $r_{xy} = -0,461$ ($l = 2$) и для $\alpha = 0,05$ r_{xy} незначим по критерию Стъюдента. Но поскольку $n = 8$ мало, необходимо подтвердить вывод о статистической незначимости построением доверительного интервала для r при $\gamma = 0,95$.

Найдем интервальную оценку для z. По таблице z-преобразования Фишера для $r_{xy} = -0,461$, учитывая, что $z'(-r) = -z'(r)$,будем иметь

$z'(-0,461) = -z'(0,461) = -0,499$ (табличное значение $z'(-0,461)$ найдено с помощью встроенной функции Excel «ФИШЕР»).

По таблице нормального закона распределения из условия $\Phi(t_\gamma) = 0,95$ найдем $t_\gamma = 1,96$. Тогда, согласно

$$z' - t_\gamma \sqrt{\frac{1}{n-l-3}} \le z \le z' + t_\gamma \sqrt{\frac{1}{n-l-3}}$$

интервальная оценка $\boldsymbol{z}$ имеет вид:

$$-0{,}499-1{,}96\cdot\sqrt{\frac{1}{8-2-3}}\le z\ \le -0{,}499+1{,}96\cdot\sqrt{\frac{1}{8-2-3}},$$

или

$$-1{,}630\ \le\ \boldsymbol{z}\ \le\ 0{,}633\ .$$

По таблице $\boldsymbol{z}$-преобразования для $\boldsymbol{z_{min}} = -1{,}630$ и $\boldsymbol{z_{max}} = 0{,}633$ найдем соответственно $\boldsymbol{r_{min}} = -0{,}926$ и $\boldsymbol{r_{max}} = 0{,}560$ (в данной работе значения $\boldsymbol{r_{min}}$ и $\boldsymbol{r_{max}}$ найдены с помощью встроенной функции Excel «ФИШЕРОБР»).

Доверительный интервал для парного коэффициента корреляции имеет вид

$$-0{,}926\ \le\ \boldsymbol{r}\ \le\ 0{,}560\ .$$

Полученная интервальная оценка подтверждает вывод о незначимости парного коэффициента корреляции $\boldsymbol{r}$, т.к. ноль находится внутри доверительного интервала (коэффициент корреляции не может одновременно принимать положительное и отрицательное значение).

2. Парная линейная регрессия

2.1. Построение уравнения парной линейной регрессии

Наиболее разработанной в эконометрике и наиболее простой с точки зрения понимания, интерпретации и техники расчетов является ***линейная форма*** регрессии, рассматривающая влияние вариации переменной $\boldsymbol{x}$ на переменную $\boldsymbol{y}$ и представляющая собой ***однофакторный регрессионный анализ***.

Линейная регрессионная модель имеет вид:

$$y = \hat{y} + \varepsilon \quad (2.1)$$

где $\hat{y} = f(x)$ - ***уравнение парной линейной (простой) регрессии***, **ε** - ***остаток*** (***отклонение***).

Уравнение парной линейной регрессии имеет вид

$$\hat{y} = b_0 + b_1 \cdot x \quad (2.2)$$

где $\boldsymbol{b_0}, \boldsymbol{b_1}$ – параметры уравнения.

$\boldsymbol{b_0}$ - ***свободный член*** регрессионного уравнения. Рассматривается как начальное значение результативного признака $\boldsymbol{y}$, если факторный признак $\boldsymbol{x}$ может принимать нулевое значение ($\boldsymbol{x}$ = 0).

$\boldsymbol{b_1}$ - ***коэффициент регрессии*** показывает, на какую величину в среднем изменится результативный признак $\boldsymbol{y}$, если факторный признак $\boldsymbol{x}$ увеличить на единицу своего измерения. Знак при коэффициенте регрессии показывает направление связи между $\boldsymbol{y}$ и $\boldsymbol{x}$: при $\boldsymbol{b_1} > 0$ - связь прямая; при $\boldsymbol{b_1} < 0$ - связь обратная.

Нахождение значений параметров $\boldsymbol{b_0}$, $\boldsymbol{b_1}$ производится на основе совокупности наблюдений (выборки, ***матрицы наблюдений***). Для разных выборок (даже одного объема) из генеральной совокупности будут найдены разные значения $\boldsymbol{b_0}$, $\boldsymbol{b_1}$. Поэтому их рассматривают как приближенные значения (***оценки***) истинных параметров регрессионного уравнения $\boldsymbol{\beta_0}$, $\boldsymbol{\beta_1}$. Сама процедура нахождения приближенных значений также называется ***оценкой параметров***.

Для получения оценок параметров линейного регрессионного уравнения используют ***метод наименьших квадратов*** (**МНК**), который минимизирует функцию

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 \to \min \qquad (2.3)$$

где $y_i, \hat{y}_i$, - соответственно фактическое (наблюдаемое) и теоретическое значение зависимой переменной y, найденное по уравнению регрессии; x_i - наблюдаемое значение независимой переменной; $\boldsymbol{n}$ - объем выборки.

Требования (условия) теоремы Гаусса-Маркова (предпосылки МНК):

А - «***истинная***» ***зависимость*** $\boldsymbol{y}$ от $\boldsymbol{x}$ имеет вид:

$$y = \beta_0 + \beta_1 \cdot x + \varepsilon;$$

В - $\boldsymbol{x}$ - ***неслучайна переменная*** (детерминированная);

С - ***столбцы матрицы наблюдений, с добавлением единичного столбца, линейно независимы*** (ранг матрицы равен 2);

D - ***остатки*** ε_i ***имеют нулевое математическое ожидание***

$$M(\varepsilon_i) = 0$$

и постоянную дисперсию σ_ε^2, не зависящую от номера наблюдения (свойство ***гомоскедастичности***), т.е.

$$D(\varepsilon_i) = \sigma_\varepsilon^2 = const;$$

E - для разных наблюдений ***остатки*** ε_i ***некоррелированы*** **(*независимы*)**, т.е. выполняется условие

$$\operatorname{cov}(\varepsilon_i, \varepsilon_l) = 0 \quad \text{, при } \boldsymbol{i \neq l}, \quad \boldsymbol{i}, \boldsymbol{l} = 1, 2, \ldots, \boldsymbol{n}\ .$$

Иногда вместо условия **D** рассматривают условие:

F - ***остатки*** ε_i ***подчиняются нормальному закону распределения***.

Теорема Гаусса-Маркова

В предположениях **A** - **E** оценки, полученные методом наименьших квадратов, являются несмещенными и обладают наименьшей дисперсией среди всех линейных несмещенных оценок параметров $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$.

Путем преобразований требование минимума суммы квадратов отклонений фактических значений результативного признака от его теоретических значений (2.3) сводится к системе нормальных уравнений вида:

$$\begin{cases} b_0 n + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \qquad (2.4)$$

После преобразований (2.4) получаем формулы для определения значения параметров парной линейной регрессии:

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} \quad \text{или} \quad b_1 = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} \tag{2.5}$$

$$b_0 = \bar{y} - b_1 \bar{x} \tag{2.6}$$

Здесь $\bar{x}, \bar{y}$ - средние значения наблюдаемых значений $\boldsymbol{x}$ и $\boldsymbol{y}$.

После построения регрессионной модели необходимо провести оценку ее качества, которая заключается в проверке качества уравнения регрессии и анализе остатков.

2.2. Основные понятия дисперсионного анализа

***Полная* (*общая*) *сумма квадратов* (*TSS*)** определяет разброс (дисперсию) зависимой переменной $\boldsymbol{y}$ относительно ее среднего значения $\bar{y}$

$$TSS = S_{\text{общ}}^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{2.7}$$

***Факторная* (*объясненная*) *сумма квадратов* (*RSS*)** определяет разброс значений $\hat{y}_i$ относительно среднего $\bar{y}$, объясненный регрессией,

$$RSS = S_{\text{факт}}^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{2.8}$$

Остаточная сума квадратов (***ESS***), т.е. сумма квадратов остатков представляет часть дисперсии ***y*** , не объясненную регрессией

$$ESS = S_{\text{ост}}^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} {\varepsilon_i}^2 \tag{2.9}$$

Основное тождество дисперсионного анализа имеет вид:

$$S_{общ}^2 = S_{факт}^2 + S_{ост}^2 \tag{2.10}$$

2.3. Характеристики линейной регрессионной модели

Построение уравнения парной линейной регрессии дополняется оценкой тесноты связи между зависимой и независимой переменными.

Коэффициент парной корреляции r_{yx}

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y} \tag{2.11}$$

или

$$r_{xy} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y} \tag{2.12}$$

служит мерой ***линейной корреляционно-регрессионной зависимости между факторами y и x*** , при условии, что на формирование их значений оказывают влияние некоторые другие, неучтенные факторы.

Коэффициент детерминации

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{2.13}$$

или

$$R^2 = \frac{S^2_{факт}}{S^2_{общ}} = 1 - \frac{S^2_{ост}}{S^2_{общ}}, \tag{2.14}$$

оценивает долю вариации признака $\boldsymbol{y}$, обусловленную изменением значений признака $\boldsymbol{x}$. Чем ближе значение R^2 $(0 < R^2 < 1)$ к единице, тем больше признак $\boldsymbol{x}$ участвует в формировании значений $\boldsymbol{y}$.

Для парной линейной регрессии справедлива формула

$$R^2 = \left(r_{xy}\right)^2, \tag{2.15}$$

2.4. Оценка качества уравнения линейной регрессии

1. Найденное уравнение парной линейной регрессии графически изображается прямой линией (рис. 2.1). На рисунке наблюдения изображены точками. Совокупность точек называется ***корреляционным полем*** (диаграммой рассеивания).

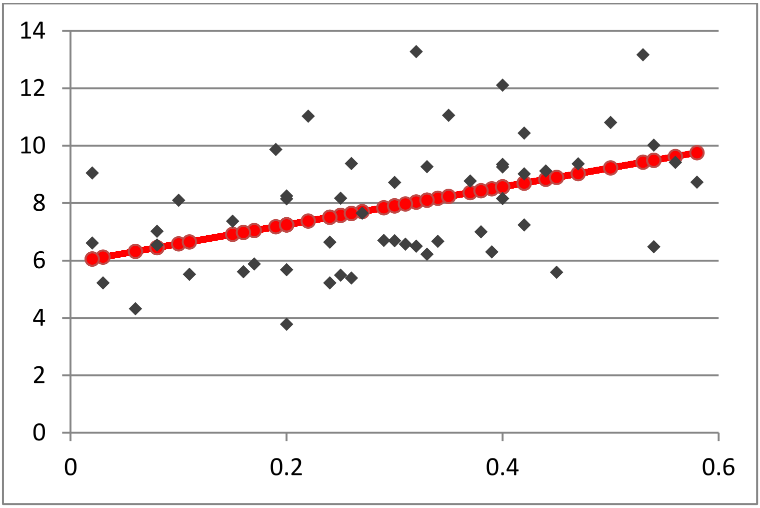


Рис. 2.1. Прямая линия регрессии на корреляционном поле

Линия регрессии занимает правильное положение на корреляционном поле, если число точек, лежащих выше и ниже линии регрессии примерно равно.

2. Если $\bar{y} \approx \bar{\hat{y}}_i$, арифметические расчеты выполнены верно.

3. Для оценки математической точности уравнения можно воспользоваться ***средней относительной ошибкой аппроксимации***

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\% \qquad (2.16)$$

Для принятия решения о точности уравнения пользуются табл.2.1

Таблица 2.1. Характеристика точности уравнения

Значение $\overline{A}$, %	**Точность уравнения**
менее 10	высокая
10 - 20	хорошая
20 - 50	удовлетворительная
более 50	неудовлетворительная

4. Проверка ***статистической значимости уравнения регрессии в целом*** осуществляется с помощью ***F-критерия Фишера***.

Находится ***расчетное значение*** (***статистика***) ***критерия***

$$F_{расч} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \cdot \frac{n - m - 1}{m} \qquad (2.17)$$

или

$$F_{расч} = \frac{S^2_{факт}}{S^2_{ост}} \cdot \frac{n - m - 1}{m} \qquad (2.18)$$

или

$$F_{расч} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} \qquad (2.19)$$

В случае парной линейной регрессии $\boldsymbol{m} = 1$.

Табличное (***критическое***) значение $\boldsymbol{F_{табл}}$, находится по таблице критических значений распределения Фишера-Снедекора (F-распределения) (табл. Приложения 3) по уровню значимости $\boldsymbol{\alpha}$ и двум числам степеней свободы $\boldsymbol{df_1} = \boldsymbol{m}$ и $\boldsymbol{df_2} = \boldsymbol{n} - \boldsymbol{m} - 1$, или с помощью встроенной функции Excel «FРАСПОБР».

Если $\boldsymbol{F_{расч}} > \boldsymbol{F_{табл}}$, то с вероятность ошибки $\boldsymbol{\alpha}$ уравнение регрессии признается ***в целом статистически значимым*** (адекватно описывающим исходные данные).

В противном случае ($\boldsymbol{F_{расч}} < \boldsymbol{F_{табл}}$) уравнение считается незначимым в целом.

5. Проверка ***статистической значимости оценок параметров*** $\boldsymbol{b}_0, \boldsymbol{b}_1$ производится с помощью ***t-критерия Стъюдента***.

Находятся ***расчетные значения критерия***

$$t_{b0} = \frac{b_0}{m_{b0}} \ , \ t_{b1} = \frac{b_1}{m_{b1}} \qquad (2.20)$$

где ***средние квадратические ошибки параметров*** равны:

$$m_{b0}=\sqrt{\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{(n-2)}\cdot\frac{\sum_{i=1}^{n}x_i^2}{n\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}} \qquad (2.21)$$

$$m_{b1}=\sqrt{\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{(n-2)\cdot\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}} \qquad (2.22)$$

Теоретическое значение критерия $\boldsymbol{t_{табл}}$ находится по таблице критических значений распределения Стъюдента (табл. Приложения 2) по уровню значимости $\boldsymbol{\alpha}$ и числу степеней свободы $\boldsymbol{df} = \boldsymbol{n} - \boldsymbol{m} - 1$, или с помощью встроенной функции Excel «СТЬЮДРАСПОБР».

Если $|\boldsymbol{t_{bj}}| > \boldsymbol{t_{табл}}$, то с вероятность ошибки $\boldsymbol{\alpha}$ оценка параметра уравнения регрессии $\boldsymbol{b_j}$ ($\boldsymbol{j}$ = 0, 1) признается статистически значимой.

В противном случае ($|\boldsymbol{t_{bj}}| < \boldsymbol{t_{табл}}$) - $\boldsymbol{b_j}$ статистически незначим.

6. ***Интервальные оценки*** (***доверительные интервалы***) параметров уравнения регрессии имеют вид:

$$b_0-m_{b0}\cdot t_{табл}<\beta_0<b_0+m_{b0}\cdot t_{табл}; \qquad (2.23)$$

$$b_1-m_{b1}\cdot t_{табл}<\beta_1<b_1+m_{b1}\cdot t_{табл}. \qquad (2.24)$$

С ***надежностью*** (вероятностью) $\boldsymbol{\gamma} = 1 - \boldsymbol{\alpha}$ они покрывают истинные параметры $\boldsymbol{\beta_j}$. Здесь $\boldsymbol{t_{табл}}$ - значение, найденное по таблице критических точек распределения Стьюдента по уровню значимости $\boldsymbol{\alpha}$ и числу степеней свободы $\boldsymbol{df} = \boldsymbol{n} - \boldsymbol{m} - 1$.

Если границы некоторого доверительного интервала имеют разные знаки, соответствующий параметр уравнения регрессии статистически незначим.

2.5. Анализ остатков

Согласно эконометрической модели (2.1)

$$y_i = \hat{y}_i + \varepsilon_i$$

остатки $\boldsymbol{\varepsilon_i}$ находятся как разность между фактическими (наблюдаемыми) и теоретическими значениями зависимой переменной:

$$\varepsilon_i = y_i - \hat{y}_i \qquad (2.25)$$

Остатки должны удовлетворять требованиям **D**, **E** теоремы Гаусса-Маркова.

1. Проверка требования **D**.

Числовой оценкой математического ожидания $M(\varepsilon_i)$ является среднее значение $\bar{\varepsilon}$. Для выполнения требования $M(\varepsilon_i) = 0$ необходимо, чтобы

$$\bar{\varepsilon} = \frac{\sum_{i=1}^{n} \varepsilon_i}{n} \approx 0 \qquad (2.26)$$

Дисперсия остатков ε_i должна быть одинаковой для всех значений x_i (свойство ***гомоскедастичности***). Если это условие не соблюдается, то имеет место ***гетероскедастичность*** остатков.

Эффективным методом предварительного анализа однородности остатков является визуальный анализ графиков остатков. Как правило, рассматривается график остатков в зависимости от номера наблюдений (рис. 2.2, 2.3).

В случае гомоскедастичности точки (остатки) равномерно располагаются внутри горизонтальной полосы, симметричной оси абсцисс (рис. 2.2).

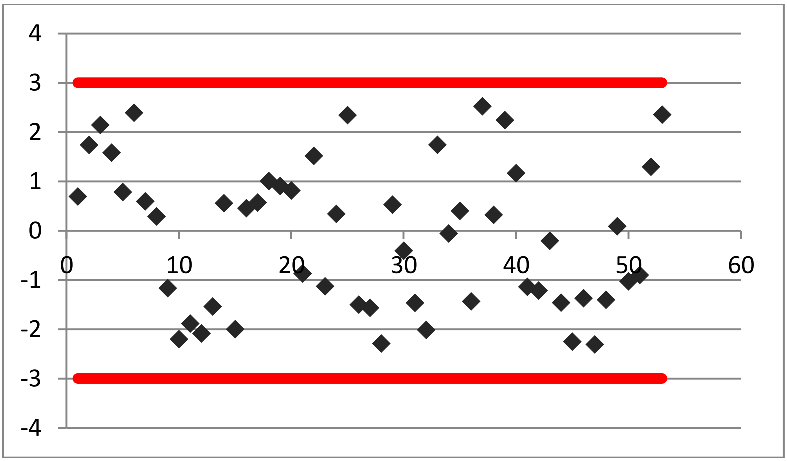


Рис. 2.2. Гомоскедастичность остатков

В случае гетероскедастичности точки (остатки), начиная с некоторого номера наблюдения, устойчиво выходят из горизонтальной полосы, симметричной оси абсцисс (рис. 2.3).

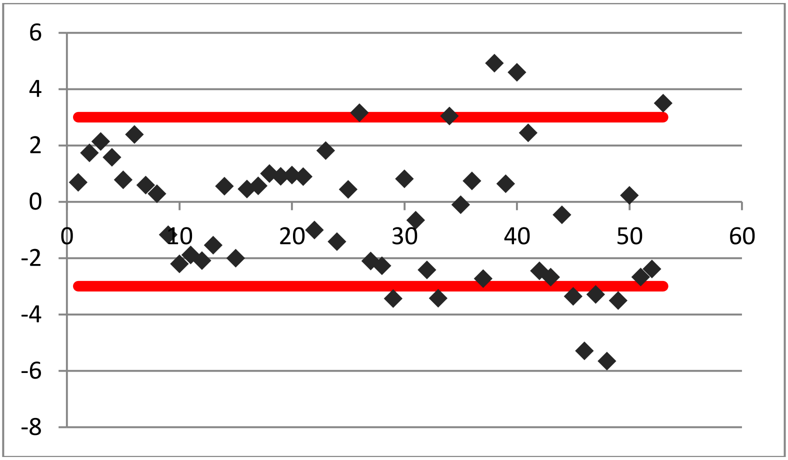


Рис. 2.3. Гетероскедастичность остатков

Для обнаружения эффекта гетероскедастичности используют тесты Уайта, Голдфелда-Квандта, Глейзера и др., но они достаточно трудоемки, и на практике наиболее часто ограничиваются визуальным анализом.

Точечная (числовая) оценка дисперсии остатков $\boldsymbol{\sigma_\varepsilon^2}$ находится по формуле:

$$s_\varepsilon^2 = \frac{1}{(n-2)}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \approx \sigma_\varepsilon^2 \qquad (2.27)$$

Интервальная оценка дисперсии остатков - доверительный интервал, который с вероятностью (надежностью) $\gamma = 1 - \alpha$ покрывает σ_ε^2, имеет вид:

$$\left(\frac{s_\varepsilon^2\cdot(n-2)}{\chi^2(k,\alpha_1)};\ \frac{s_\varepsilon^2\cdot(n-2)}{\chi^2(k,\alpha_2)}\right) \qquad (2.28)$$

или

$$\frac{s_{\varepsilon}^{2}\cdot(n-2)}{\chi^{2}(df,\alpha_{1})}<\sigma_{\varepsilon}^{2}<\frac{s_{\varepsilon}^{2}\cdot(n-2)}{\chi^{2}(df,\alpha_{2})}.$$

Здесь χ^2 - критические значения распределения Пирсона, найденные по табл. Приложения 4 по числу степеней свободы $df = n - 2$ и уровням значимости $\alpha_1 = 1 - \alpha/2$, $\alpha_2 = \alpha/2$.

2. Проверка требования **Е -** для разных наблюдений ***остатки ε_i некоррелированы (независимы).***

Наиболее распространенный метод проверки требования о независимости остатков - ***критерий Дарбина-Уотсона*** (о наличии в остатках автокорреляции первого порядка), в котором рассчитывается статистика

$$d_{расч}=\frac{\sum_{i=2}^{n}(\varepsilon_i-\varepsilon_{i-1})^2}{\sum_{i=1}^{n}\varepsilon_i^2} \qquad (2.29)$$

где $0 \le d_{расч} \le 4$.

Для парного уравнения регрессии теоретические значения критерия Дарбина-Уотсона d_U и d_L находятся по таблице критических значений (табл. Приложения 5) по объему выборки n, числу степеней свободы $df = 1$ и уровню значимости α. С помощью критических значений числовой промежуток (0; 4) разбивается на пять отрезков.

Правило принятия решения о зависимости остатков

Есть положительная автокорреляция остатков	Зона неопределенности	Автокорреляция остатков отсутствует	Зона неопределенности	Есть отрицательная автокорреляция остатков

0 $\quad d_L \quad d_U \quad$ **2** $\quad$ **4 -** $d_U \quad$ **4 -** $d_L \quad$ **4**

На практике в случае, когда статистика критерия $d_{расч}$ попадает в зону неопределенности, обычно признается наличие автокорреляции в остатках.

Замечания

1. Если уравнение регрессии признано качественным, а остатки удовлетворяют требованиям **D**, **E** теоремы Гаусса-Маркова, то регрессионная модель считается качественной, т.е. она адекватно описывает исходные данные.

2. Для получения качественной парной линейной регрессионной модели необходима выборка объема не меньше $\boldsymbol{n}$ = (6÷8)×2.

2.6. Практические приложения парной линейной регрессионной модели

Наиболее часто встречающиеся приложения регрессионной модели - оценка влияния объясняющих переменных на результативный признак и построение прогноза.

2.6.1. Эластичность

Для оценки влияния факторного признака ***x*** на результативный признак ***y*** вычисляют ***средний*** $\overline{Э}$ и ***частные*** $Э_i$ ***коэффициенты эластичности***.

Средний по совокупности коэффициент эластичности

$$\overline{Э} = b_1 \cdot \frac{\overline{x}}{\overline{y}} \quad (2.30)$$

показывает, на сколько процентов изменяется $\overline{y}$ при увеличении $\overline{x}$ на один процент. Средний коэффициент эластичности позволяет выявить общегрупповые закономерности.

Частный коэффициент эластичности

$$Э_i = b_1 \cdot \frac{x_i}{b_0 + b_1 \cdot x_i} \quad (2.31)$$

показывает, на сколько процентов изменится y_i , если x_i увеличится на 1%.

Здесь b_0, b_1 – параметры уравнения регрессии, x_i - ***i***-е наблюдение независимой переменной (i = 1, 2, …, n).

Частные коэффициенты эластичности позволяют выявить особенности, присущие отдельным объектам наблюдения.

2.6.2. Прогнозирование

В прогнозных расчетах предсказываемое значение $\boldsymbol{y}_{\text{пр}}$ определяется как точечный прогноз $\hat{y}_{\text{пр}}$ путем подстановки в уравнение регрессии соответствующего значения $\boldsymbol{x_{пр}}$. Точечный прогноз дополняется построением интервальной оценки (доверительного интервала).

Для парного линейного уравнения регрессии

$$y_{пр} \approx \hat{y}_{пр} = b_0 + b_1 \cdot x_{пр} \tag{2.32}$$

Интервальная оценка прогноза $y_{пр}$ имеет вид:

$$\hat{y}_{пр} - t_{табл} \cdot m_{\hat{y}пр} \le y_{пр} \le \hat{y}_{пр} + t_{табл} \cdot m_{\hat{y}пр} \tag{2.33}$$

Здесь: $\boldsymbol{t_{табл}}$ - критическое значение распределения Стьюдента, найденное по уровню значимости $\boldsymbol{\alpha}$ и числу степеней свободы $\boldsymbol{df} = \boldsymbol{n} - 2$.

Средняя стандартная ошибка прогноза равна:

$$m_{\hat{y}пр} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n-2} \cdot \left(1 + \frac{1}{n} + \frac{\left(x_{пр} - \bar{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right)} \tag{2.34}$$

2.7. Решение типовых задач

Задача 2.1. Имеются данные статистических наблюдений показателей «среднедушевой денежный доход» (***x***, руб.) и «среднедушевой оборот розничной торговли» (***y***, руб.) для 16 городов (см. табл. 2.2).

Таблица 2.2. Исходные данные

№ города	Среднедушевой денежный доход	Среднедушевой оборот розничной торговли	№ города	Среднедушевой денежный доход	Среднедушевой оборот розничной торговли
1	3357	2425	9	3724	2225
2	3135	2050	10	3416	1983
3	2842	1683	11	3340	1925
4	3991	2375	12	3089	1042
5	2293	1167	13	4372	2925
6	3563	2200	14	3022	2342
7	3219	1892	15	3383	2458
8	3308	2008	16	4267	2125

Требуется:

1) оценить тесноту линейной корреляционно-регрессионной зависимости;
2) найти уравнение парной линейной регрессии и построить его на корреляционном поле;
3) записать модель парной линейной регрессии;
4) оценить качество уравнения регрессии;
5) провести анализ остатков.

Решение

Объем выборки $\boldsymbol{n}$ = 16.

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

Коэффициент парной корреляции r_{xy} найдем по формуле (1.6). Составим расчетную таблицу (табл. 2.3).

Средние значения признаков равны

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{54321}{16} = 3395{,}1\,; \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{32825}{16} = 2051{,}6\,.$$

Коэффициент парной корреляции равен

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{2590081{,}4}{\sqrt{4070770{,}9 \cdot 3299597{,}9}} = 0{,}707$$

Статистическую значимость коэффициента парной корреляции проверим с помощью $\boldsymbol{t}$-критерия Стьюдента. Найдем статистику (1.7)

$$t_{расч} = \frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \cdot \sqrt{n - m - 1} = \frac{0{,}707}{\sqrt{1 - 0{,}4998}}\sqrt{16 - 1 - 1} = 3{,}74$$

.

Таблица 2.3. Расчетная таблица коэффициента парной корреляции

№ п/п	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1,0	3357,0	2425,0	-38,1	373,4	-14214,0	1448,8	139455,6
2,0	3135,0	2050,0	-260,1	-1,6	406,3	67632,5	2,4
3,0	2842,0	1683,0	-553,1	-368,6	203838,1	305878,1	135838,3
4,0	3991,0	2375,0	595,9	323,4	192748,5	355141,5	104611,8
5,0	2293,0	1167,0	-1102,1	-884,6	974843,2	1214541,8	782450,8
6,0	3563,0	2200,0	167,9	148,4	24928,2	28203,0	22033,7
7,0	3219,0	1892,0	-176,1	-159,6	28093,0	30998,0	25460,2
8,0	3308,0	2008,0	-87,1	-43,6	3792,7	7579,9	1897,7
9,0	3724,0	2225,0	328,9	173,4	57050,1	108199,9	30080,6
10,0	3416,0	1983,0	20,9	-68,6	-1435,5	438,4	4700,8
11,0	3340,0	1925,0	-55,1	-126,6	6968,8	3031,9	16018,1
12,0	3089,0	1042,0	-306,1	-1009,6	308989,2	93674,3	1019216,4
13,0	4372,0	2925,0	976,9	873,4	853293,8	954406,9	762893,1
14,0	3022,0	2342,0	-373,1	290,4	-108351,3	139175,6	84353,9
15,0	3383,0	2458,0	-12,1	406,4	-4902,7	145,5	165191,4
16,0	4267,0	2125,0	871,9	73,4	64032,9	760275,0	5393,1
Сумма	**54321,0**	**32825,0**			**2590081,4**	**4070770,9**	**3299597,9**
Среднее	**3395,1**	**2051,6**					

Табличное значение t-критерия при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $df = n - m - 1 = 16 - 1 - 1 = 14$ составляет $t_{табл}(0{,}05;\ 14) = 2{,}14$. (табличное значение найдено с помощью встроенной функции Excel «СТЬЮДРАСПОБР»).

Так как $|t_{расч}| > t_{табл}$ (3,74 > 2,14) , значение коэффициента парной корреляции статистически значимо.

Характер зависимости между факторами ***x***, ***y*** определим по шкале Чеддока (табл.1.1).

Величина статистически значимого коэффициента парной корреляции свидетельствует о прямой тесной линейной связи (близкой к умеренной) между среднедушевым денежным доходом и среднедушевым оборотом розничной торговли.

Коэффициент детерминации найдем по формуле (1.13)

$$R^2 = (r_{yx})^2 = (0{,}707)^2 = 0{,}4998.$$

Его значение показывает, что ***на 49,98% изменение оборота розничной торговли объясняется изменениями денежных доходов населения***. Оставшиеся 50,02% приходятся на другие факторы, не включенные в модель.

Значения коэффициентов парной корреляции и детерминации свидетельствуют о том, что между факторами «среднедушевой денежный доход» и «среднедушевой оборот розничной торговли» существует достаточно сильная регрессионная зависимость. Построение уравнения парной линейной регрессии теоретически обосновано.

2) Уравнение парной линейной регрессии будем искать в виде (2.2): $\hat{y} = b_0 + b_1 \cdot x$.

Его параметры $\boldsymbol{b_0}$, $\boldsymbol{b_1}$ найдем методом наименьших квадратов по формулам (2.5), (2.6), предварительно проведя необходимые расчеты (см. табл. 2.4-2.5):

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{114033008 - 54321 \cdot 32825 : 16}{188493961 - (54321)^2 : 16} = 0{,}64$$

$$b_0 = \bar{y} - b_1 \bar{x} = 2051{,}6 - 0{,}64 \cdot 3395{,}1 = -108{,}59$$

Уравнение парной линейной регрессии имеет вид:

$$\hat{y} = -108{,}59 + 0{,}64 \cdot x \ .$$

Коэффициент регрессии $\boldsymbol{b_1} = 0{,}64$ показывает, что с увеличением среднедушевого денежного дохода в среднем на 1 руб. среднедушевой розничный оборот в среднем возрастает на 0,64 руб.

Подставляя в полученное уравнение регрессии значения x_i определим ***теоретические значения*** $\hat{y}_i$ (см. табл. 2.4). По значениям $\hat{y}_i$ построим график линейной регрессии (см. рис. 2.4).

Таблица 2.4. Расчетная таблица линейного регрессионного уравнения

№ п/п	x_i	y_i	x_i^2	$x_i \cdot y_i$	$\hat{y}_i$
1	3357	2425	11269449	8140725	2027,3
2	3135	2050	9828225	6426750	1886,1
3	2842	1683	8076964	4783086	1699,7
4	3991	2375	15928081	9478625	2430,7
5	2293	1167	5257849	2675931	1350,4
6	3563	2200	12694969	7838600	2158,4

7	3219	1892	10361961	6090348	1939,5
8	3308	2008	10942864	6642464	1996,2
9	3724	2225	13868176	8285900	2260,9
10	3416	1983	11669056	6773928	2064,9
11	3340	1925	11155600	6429500	2016,5
12	3089	1042	9541921	3218738	1856,8
13	4372	2925	19114384	12788100	2673,2
14	3022	2342	9132484	7077524	1814,2
15	3383	2458	11444689	8315414	2043,9
16	4267	2125	18207289	9067375	2606,3
Сумма	**54321**	**32825**	**188493961**	**114033008**	**32825,0**
Среднее	**3395,1**	**2051,6**			**2051,6**

Таблица 2.5. Расчетная таблица линейного регрессионного уравнения

№ п/п	$\left\lvert \frac{y_i - \hat{y}_i}{y_i} \right\rvert$	$(\hat{y}_i - \bar{y})^2$	$(\hat{y}_i - y_i)^2$	$(x_i - \bar{x})^2$
1	0,164	586,5	158129,7	1448,8
2	0,080	27379,7	26865,1	67632,5
3	0,010	123828,9	277,9	305878,1
4	0,023	143772,2	3106,5	355141,5
5	0,157	491683,9	33621,2	1214541,8
6	0,019	11417,4	1729,3	28203,0
7	0,025	12548,9	2260,1	30998,0
8	0,006	3068,6	140,0	7579,9
9	0,016	43802,6	1285,5	108199,9
10	0,041	177,5	6705,0	438,4
11	0,048	1227,4	8377,4	3031,9
12	0,782	37922,2	663941,8	93674,3
13	0,086	386373,3	63427,5	954406,9
14	0,225	56342,6	278576,4	139175,6
15	0,168	58,9	171489,1	145,5
16	0,227	307782,7	231692,2	760275,0
Сумма	**2,078**	**1647973,3**	**1651624,6**	**4070770,9**
Среднее	**0,1299**			

3) ***Модель парной линейной регрессии*** (2.1) имеет вид

$$y = -108{,}59 + 0{,}64 \cdot x + \varepsilon$$

или

$$y_i = -108{,}59 + 0{,}64 \cdot x_i + \varepsilon_i$$

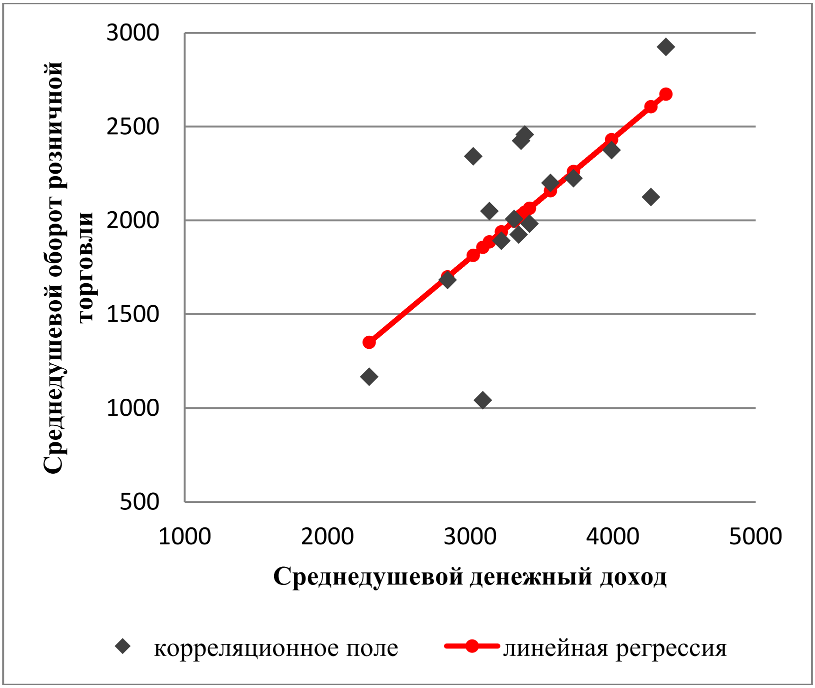


Рис. 2.4. Прямая линия регрессии на корреляционном поле

4) Оценим качество уравнения регрессии.

а) Найдем среднюю относительную ошибку аппроксимации

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\% = \frac{2{,}078}{16} \cdot 100\% = 12{,}99\% \ .$$

Так как 10% < $\overline{A}$ < 20% , ***уравнение имеет хорошую точность*** (см. табл. 2.1).

б) Проверим статистическую значимость уравнения регрессии в целом с помощью критерия Фишера. Для нахождения расчетного значения критерия воспользуемся формулой (2.17). Необходимые расчеты приведены в табл. 2.5.

$$F_{расч} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \cdot \frac{n-m-1}{m} = \frac{1647973{,}3}{1651624{,}6} \cdot \frac{16-1-1}{1} = 13{,}97 \ .$$

Табличное значение с $\boldsymbol{df_1 = m}$ = 1 и $\boldsymbol{df_2 = n - m} - 1$ = = 16 -1 - 1 = 14 степенями свободы при уровне значимости $\boldsymbol{\alpha}$ = 0,05 найдем с помощью встроенной функции Excel «FРАСПОБР». $\boldsymbol{F_{табл}}$(0,05; 1; 14) = 4,60 .

Поскольку $\boldsymbol{F_{расч} > F_{табл}}$ (13,97 > 4,60), ***уравнение парной линейной регрессии статистически значимо в целом***, оно адекватно описывает исходные данные.

в) Проверим статистическую значимость параметров уравнения регрессии с помощью ***t***-критерия Стьюдента. Необходимые расчеты приведены в табл. 2.4-2.5.

Расчетные значения критерия равны (2.20):

$$t_{b0} = \frac{b_0}{m_{b0}} = \frac{-108{,}59}{584{,}31} = -0{,}19 \ , \ t_{b1} = \frac{b_1}{m_{b1}} = \frac{0{,}64}{0{,}17} = 3{,}74 \ .$$

Средние квадратические ошибки параметров, найденные по формулам (2.21) и (2.22), равны

$$m_{b0} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n-2)} \cdot \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{16516246}{16-2} \cdot \frac{188493961}{16 \cdot 40707709}}$$

$$= 584{,}31$$

$$m_{b1} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n-2)\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{16516246}{(16-2)\cdot 40707709}} = 0{,}17 \ .$$

Табличное значение критерия Стьюдента $\boldsymbol{t_{табл}}$ (0,05; 14) = 2,14 .

Так как:

$|\boldsymbol{t_{b0}}| < \boldsymbol{t_{табл}}$ (0,19 < 2,14) - ***параметр*** $\boldsymbol{b_0}$ ***статистически незначим***;

$|\boldsymbol{t_{b1}}| > \boldsymbol{t_{табл}}$ (3,74 > 2,14) - ***параметр*** $\boldsymbol{b_1}$ ***статистически значим***.

г) Найдем интервальные оценки параметров $\boldsymbol{\beta_0}$, $\boldsymbol{\beta_1}$ по формулам (2.23), (2.24). Табличное значение критерия Стьюдента $\boldsymbol{t_{табл}}$ (0,05; 14) = 2,14 .

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

$$\left(-108{,}59-584{,}31\cdot 2{,}14;-108{,}59+584{,}31\cdot 2{,}14\right);$$

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

$$\left(0{,}64-0{,}17\cdot 2{,}14;0{,}64+0{,}17\cdot 2{,}14\right).$$

Интервальная оценка β_0 : $\left(-1361{,}81\ \ ;\ \ 1144{,}63\right)$.

Интервальная оценка β_1 : $\left(0{,}27\ \ ;\ \ 1{,}00\right)$.

5) Проведем анализ остатков.

Остатки найдем по формуле (2.25) (расчеты приведены в табл. 2.6) $\varepsilon_i = y_i - \hat{y}_i$

Проверим требование **D** теоремы Гаусса-Маркова.

1. ***Среднее значение остатков равно нулю*** (2.26):

$$\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i \approx 0$$

Первое условие требования **D** выполняется.

2. На графике (рис. 2.5) точки не расположены внутри горизонтальной полосы, симметричной оси абсцисс. ***Дисперсия остатков не постоянна*** (остатки гетероскедастичны). Второе условие требования **D** не выполняется.

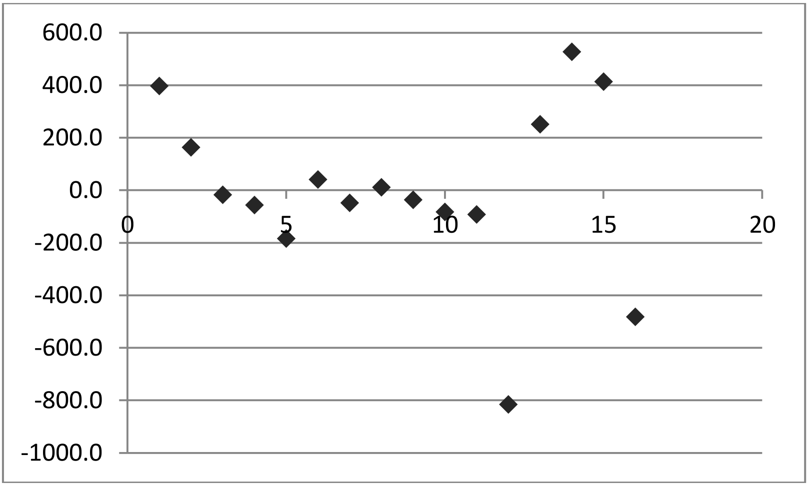


Рис. 2.5. График остатков

Точечная оценка дисперсии остатков σ_ε^2 (2.27) равна:

$$\sigma_\varepsilon^2 \approx s_\varepsilon^2 = \frac{1}{(n-2)}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 =$$

$$\frac{1}{(n-2)}\sum_{i=1}^{n}\varepsilon_i^2 = \frac{1}{16-2}\cdot 1651624{,}63 = 117973{,}2.$$

Интервальная оценка дисперсии остатков находится по формуле (2.28).

Критические значения распределения Пирсона χ^2 найдем по числу степеней свободы $df = n - 2 = 16 - 2 = 14$ и уровням значимости $\alpha_1 = 1 - \alpha/2 = 0{,}975$, $\alpha_2 = \alpha/2 = 0{,}025$ с помощью встроенной функции Excel «ХИ2ОБР».

$\chi^2_{табл}(0{,}975;\ 14) = 26{,}12$, $\chi^2_{табл}(0{,}025;\ 14) = 5{,}63$.

Тогда ***доверительный интервал дисперсии остатков***

$$\left(\frac{s_{\varepsilon}^{2} \cdot (n-2)}{\chi^{2}(df, \alpha_{1})} ; \frac{s_{\varepsilon}^{2} \cdot (n-2)}{\chi^{2}(df, \alpha_{2})} \right)$$

имеет вид

$$\left(\frac{117973{,}2 \cdot (16-2)}{26{,}12} ; \frac{117973{,}2 \cdot (16-2)}{5{,}63} \right)$$

или

$$\left(63234{,}73 \; ; \; 293427{,}78 \right).$$

Проверим требование **Е** теоремы Гаусса-Маркова - для разных наблюдений остатки $\boldsymbol{\varepsilon_i}$ независимы.

Воспользуемся критерием Дарбина-Уотсона. Вычислим статистику по формуле (2.29):

$$d_{расч} = \frac{\sum_{i=2}^{n} (\varepsilon_{i} - \varepsilon_{i-1})^{2}}{\sum_{i=1}^{n} \varepsilon_{i}^{2}} = \frac{2723489{,}06}{1651624{,}63} = 1{,}65$$

Таблица 2.6. Расчетная таблица анализа остатков линейной регрессионной модели

№ п/п	x_i	y_i	$\hat{y}_i$	$\varepsilon_i = y_i - \hat{y}_i$	ε_{i-1}	$(\varepsilon_i - \varepsilon_{i-1})^2$	ε_i^2
1	3357	2425	2027,34	397,66	-	-	158129,71
2	3135	2050	1886,09	163,91	397,66	54638,869	26865,07
3	2842	1683	1699,67	-16,67	163,91	32607,296	277,86
4	3991	2375	2430,74	-55,74	-16,67	1526,1782	3106,45
5	2293	1167	1350,36	-183,36	-55,74	16288,195	33621,17
6	3563	2200	2158,41	41,59	-183,36	50600,627	1729,32
7	3219	1892	1939,54	-47,54	41,59	7943,3521	2260,09
8	3308	2008	1996,17	11,83	-47,54	3525,1035	140,00
9	3724	2225	2260,85	-35,85	11,83	2273,9031	1285,46
10	3416	1983	2064,88	-81,88	-35,85	2118,8491	6705,03
11	3340	1925	2016,53	-91,53	-81,88	93,006784	8377,42
12	3089	1042	1856,83	-814,83	-91,53	523159,93	663941,76
13	4372	2925	2673,15	251,85	-814,83	1137794,3	63427,51
14	3022	2342	1814,20	527,80	251,85	76151,283	278576,44
15	3383	2458	2043,89	414,11	527,80	12925,641	171489,10
16	4267	2125	2606,34	-481,34	414,11	801842,53	231692,22
Сумма	**54321**	**32825**	**32825,00**			**2723489,06**	**1651624,6**
Среднее	**3395,1**	**2051,6**	**2051,6**	**0,0000000**			

Для уравнения парной линейной регрессии теоретические значения критерия Дарбина-Уотсона найдем по таблице критических значений (табл. Приложения 5) по объему выборки $\boldsymbol{n}$ **=** 16, числу степеней свободы $\boldsymbol{df}$ = 1 и уровню значимости $\boldsymbol{\alpha}$ = 0,05. $\boldsymbol{d_L}$ = 1,10 и $\boldsymbol{d_U}$ = 1,37.

Поскольку $\boldsymbol{d_{расч}}$ = 1,65 попадает в интервал $\left(1{,}37 < d_{расч} < 2{,}63\right)$ (см. рис. 2.6), в котором отсутствует автокорреляция, остатки независимы. Требование **Е** выполняется.

Есть положительная автокорреляция остатков	**Зона неопределенности**	**Автокорреляция остатков отсутствует**	**Зона неопределенности**	**Есть отрицательная автокорреляция остатков**

0 1,10 1,37 2,63 2,90 4

Рис. 2.6. Решающее правило критерия Дарбина-Уотсона

Вывод

Уравнение парной линейной регрессии имеет хорошую точность, статистически значимо в целом, имеет статистически незначимый свободный член $\boldsymbol{b_0}$ и статистически значимый коэффициент регрессии $\boldsymbol{b_1}$. Регрессионное уравнение считается некачественным.

Остатки независимы, имеют нулевое математическое ожидание и непостоянную дисперсию. Не все требования теоремы Гаусса-Маркова к остаткам выполняются.

В целом модель парной линейной регрессии считается некачественной.

Задача 2.2. В задаче 2.1. для данных статистических наблюдений показателей «среднедушевой денежный доход» (***x***, руб.) и «среднедушевой оборот розничной торговли» (***y***, руб.) (табл. 2.2) построено уравнение парной линейной регрессии

$$\hat{y} = -108{,}59 + 0{,}64 \cdot x\ .$$

Требуется:

1) вычислить прогнозное значение показателя «среднедушевой оборот розничной торговли», если значение показателя «среднедушевой денежный доход» составит 107% от среднего уровня;
2) оценить точность прогноза;
3) найти средние и частные коэффициенты эластичности.

Решение.

1) Найдем прогнозное значение признака $\boldsymbol{x_{пр}}$:

$$x_{пр} = \bar{x} \cdot 1{,}07 = 3395{,}1 \cdot 1{,}07 = 3632{,}72$$

Прогнозное значение $\hat{y}_{\text{пр}}$ найдем по формуле (2.33):

$$\hat{y}_{пр} = b_0 + b_1 \cdot x_{пр} = -108{,}59 + 0{,}64 \cdot 3632{,}72 = 2202{,}77.$$

2) Построим доверительный интервал прогноза вида (2.33) при уровне значимости $\boldsymbol{\alpha}$ = 0,05:

$$\hat{y}_{пр} - t_{табл} \cdot m_{\hat{y}пр} \le y_{пр} \le \hat{y}_{пр} + t_{табл} \cdot m_{\hat{y}пр}\ .$$

Средняя стандартная ошибка прогноза равна (см. формулу (2.34)):

$$m_{\hat{y}пр} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n-2} \cdot \left(1 + \frac{1}{n} + \frac{\left(x_{пр} - \bar{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right)} =$$

$$= \sqrt{\frac{1651624{,}6}{16-2} \cdot \left(1 + \frac{1}{16} + \frac{\left(3632{,}72 - 3395{,}1\right)^2}{4070770{,}9}\right)} = 356{,}35\,.$$

Критическое значение распределения Стъюдента $\boldsymbol{t_{табл}}$ найдем по уровню значимости $\boldsymbol{\alpha} = 0{,}05$ и числу степеней свободы $\boldsymbol{df} = \boldsymbol{n} - 2 = 16 - 2 = 14$: $\boldsymbol{t_{табл}}(0{,}05;\ 14) = 2{,}14$.

Найденные значения подставим в выражение (2.34):

$$2202{,}77 - 2{,}14 \cdot 356{,}35 \le y_{пр} \le 2202{,}77 + 2{,}14 \cdot 356{,}35$$

или

$$1440{,}18 \le y_{пр} \le 2965{,}36.$$

Таким образом, если среднедушевой денежный доход будет равен 3632,72 руб., то среднедушевой оборот розничной торговли составит около 2202,77 руб. и с вероятностью $1 - \alpha = 1 - 0{,}05 = 0{,}95$ будет находиться в пределах от 1440,18 до 2965,36 руб. (расчеты приведены в табл. 2.7).

3) Средний и частные коэффициенты эластичности в парной линейной регрессионной модели найдем по формулам (2.30) и (2.31):

$$\overline{Э} = b_1 \cdot \frac{\overline{x}}{\overline{y}} = 0,64 \cdot \frac{3395,1}{2051,6} = 1,05.$$

Средний коэффициент эластичности показывает, что ***при увеличении среднего по совокупности среднедушевого денежного дохода на 1% среднедушевой оборот розничной торговли увеличится в среднем на 1,05%.***

Частные коэффициенты эластичности

$$Э_i = b_1 \cdot \frac{x_i}{b_0 + b_1 \cdot x_i} = 0,64 \cdot \frac{x_i}{-108,59 + 0,64 \cdot x_i}$$

вычислены в табл. 2.7.

Анализ значений частных коэффициентов эластичности показывает, что ***для 5-го города рост среднедушевого денежного дохода на 1% приводит к наибольшему по группе городов росту оборота розничной торговли*** ($Э_5$ = 1,08).

Для 4, 13 и 16-го города увеличение среднедушевого денежного дохода на 1% приводит к наименьшему росту оборота розничной торговли, чем в целом по группе городов ($Э_4$ = 1,04, $Э_{13}$ = 1,04, $Э_{16}$ = 1,04).

Для 1, 6, 8, 9, 10, 11 и 15-го города увеличение среднедушевого денежного дохода на 1% приводит к такому же росту оборота розничной торговли, как в среднем по группе городов (соответствующие частные коэффициенты эластичности равны среднему коэффициенту эластичности $\overline{Э}$ = 1,05).

Таблица 2.7. Расчетная таблица

№ п/п	x_i	y_i	$(x_i - \overline{x})^2$	$Э_i$
1	3357	2425	1448,8	1,05
2	3135	2050	67632,5	1,06
3	2842	1683	305878,1	1,06
4	3991	2375	355141,5	***1,04***
5	2293	1167	1214541,8	**1,08**
6	3563	2200	28203,0	1,05
7	3219	1892	30998,0	1,06
8	3308	2008	7579,9	1,05
9	3724	2225	108199,9	1,05
10	3416	1983	438,4	1,05
11	3340	1925	3031,9	1,05
12	3089	1042	93674,3	1,06
13	4372	2925	954406,9	***1,04***
14	3022	2342	139175,6	1,06
15	3383	2458	145,5	1,05
16	4267	2125	760275,0	***1,04***
Сумма			**4070770,9**	
Среднее	**3395,1**	**2051,6**		

3. Парная нелинейная регрессия

3.1. Построение уравнения парной нелинейной регрессии

В силу многообразия и сложности экономических процессов, многие экономические зависимости не являются линейными по своей сути и не могут быть описаны линейными моделями. Существующие нелинейные соотношения выражаются с помощью соответствующих нелинейных функций.

Под ***формой нелинейной регрессионной зависимости*** (видом связи) понимают вид уравнения регрессии. Наиболее часто используются следующие виды нелинейных зависимостей.

Нелинейная регрессионная модель имеет вид:

$$\boldsymbol{y = \hat{y} + \varepsilon},$$

где $\boldsymbol{\hat{y} = f(x)}$ - ***уравнение нелинейной регрессии***, **ε** - ***остаток*** (***отклонение***).

Регрессии, ***нелинейные по оцениваемым параметрам:***

1. ***степенное*** уравнение $\hat{y} = b_0 \cdot x^{b1}$; (3.1)

Параметр $\boldsymbol{b}_1$ степенного уравнения является коэффициентом эластичности и показывает, на сколько процентов изменится $\bar{y}$ при увеличении независимой переменной ***x*** на один процент.

2. ***показательное*** уравнение $\hat{y} = b_0 \cdot b_1^x$; (3.2)

3. ***экспоненциальное*** уравнение $\hat{y} = b_0 \cdot e^{b1x}$ (3.3)

Регрессии, ***нелинейные по объясняющей переменной:***

4. ***гиперболическое*** уравнение $\hat{y} = b_0 + b_1 \dfrac{1}{x}$; (3.4)

5. ***полиномиальные*** уравнения степени ***m*** $\hat{y} = b_0 + b_1 \cdot x + b_2 \cdot x^2 + \dots + b_m \cdot x^m$; (3.5)

6. полином степени m = 2 – ***парабола*** $\hat{y} = b_0 + b_1 \cdot x + b_2 \cdot x^2$ (3.5.b)

Оценка параметров нелинейной регрессии $\boldsymbol{b_0}$, $\boldsymbol{b_1}$ (нахождение их приближенных значений) наиболее часто осуществляться с помощью метода наименьших квадратов (МНК) с предварительной линеаризацией уравнения регрессии (приведения уравнения к линейному виду).

3.1.1. Степенная регрессия

Для определения методом МНК параметров степенной функции (3.1)

$$\hat{y} = b_0 \cdot x^{b1} ,$$

приведем ее к линейному виду путем логарифмирования обеих частей уравнения, после чего получим

$$\ln \hat{y} = \ln b_0 + b_1 \ln x \cdot$$

Обозначим $Y = \ln \hat{y}$, $X = \ln x$, $B_0 = \ln b_0$. Тогда уравнение примет вид, линейный относительно ***X, Y***,

$$Y = B_0 + b_1 \cdot X \qquad (3.6)$$

Параметры уравнения (3.6) находятся по формулам:

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n} \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} X_i \right)^2} \text{ или } b_1 = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\overline{X^2} - \left(\overline{X}\right)^2} \qquad (3.7)$$

$$B_0 = \overline{Y} - b_1 \overline{X} \qquad (3.8)$$

Обратный переход к параметру $\boldsymbol{b_0}$ осуществляется по формуле:

$$b_0 = e^{B0} \qquad (3.9)$$

3.1.2. Показательная регрессия

Для определения методом МНК параметров показательной функции (3.2)

$$\hat{y} = b_0 \cdot b_1^x,$$

приведем ее к линейному виду путем логарифмирования обеих частей уравнения. После чего получим,

$$\ln \hat{y} = \ln b_0 + x \ln b_1.$$

Обозначим $Y = \ln \hat{y}$, $B_0 = \ln b_0$, $B_1 = \ln b_1$, тогда уравнение (3.2) примет линейный вид

$$Y = B_0 + B_1 \cdot x \tag{3.10}$$

Параметры уравнения (3.10) находятся по формулам:

$$B_1 = \frac{\sum_{i=1}^{n} x_i Y_i - \frac{1}{n} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} x_i \right)^2} \text{ или } B_1 = \frac{\overline{xY} - \overline{x} \cdot \overline{Y}}{\overline{x^2} - \left(\overline{x}\right)^2} \tag{3.11}$$

$$B_0 = \overline{Y} - B_1 \overline{x} \tag{3.12}$$

Обратный переход к параметрам $\boldsymbol{b_0}$, $\boldsymbol{b_1}$ осуществляется по формулам:

$$b_0 = e^{B0}\ ,\ b_1 = e^{B1} \tag{3.13}$$

3.1.3. Гиперболическая регрессия

Для нахождения методом МНК параметров гиперболической функции (3.4)

$$\hat{y} = b_0 + b_1 \frac{1}{x}\ ,$$

введем новую переменную $X = \frac{1}{x}$. В результате получим линейное уравнение

$$\hat{y} = b_0 + b_1 \cdot X\ , \tag{3.14}$$

параметры которого определяются по формулам:

$$b_1 = \frac{\sum_{i=1}^{n} X_i y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2}$$

или (3.15)

$$b_1 = \frac{\overline{Xy} - \overline{X} \cdot \overline{y}}{\overline{X^2} - \left(\overline{X}\right)^2}$$

$$b_0 = \overline{y} - b_1 \overline{X} \tag{3.16}$$

3.2. Оценка качества уравнения нелинейной регрессии

1. Оценка математической точности уравнения производится с помощью ***средней относительной ошибки аппроксимации***

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\% \qquad (3.17)$$

где y_i и $\hat{y}_i$ - соответственно фактические и теоретические значения переменной $\boldsymbol{y}$. Для принятия решения о точности уравнения можно воспользоваться табл. 2.1.

2. Проверка ***статистической значимости уравнения регрессии в целом*** производится с помощью ***F-критерия Фишера***.

Выдвигается гипотеза $\boldsymbol{H_0}$: уравнение регрессии статистически незначимо, при конкурирующей гипотезе $\boldsymbol{H_1}$: уравнение регрессии статистически значимо. Находится ***расчетное значение*** (***статистика***) ***критерия***

$$F_{расч} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \cdot \frac{n - m - 1}{m} \qquad (3.18)$$

где $\boldsymbol{y_i}$, $\boldsymbol{\hat{y}_i}$, $\boldsymbol{\overline{y}}$ – соответственно фактическое (наблюдаемое), теоретическое и среднее значения $\boldsymbol{y}$; $\boldsymbol{n}$ - объем выборки, $\boldsymbol{m}$ - число параметров уравнения регрессии при независимых переменных.

Табличное (***критическое***) значение $F_{табл}$, находится по таблице критических значений распределения Фишера-Снедекора (***F-распределения***) по уровню значимости α и двум числам степеней свободы $df_1 = m$ и $df_2 = n - m - 1$.

Если $F_{расч} > F_{табл}$, то гипотеза H_0 отвергается с вероятность ошибки α , т.е. уравнение регрессии признается в целом статистически значимым (***адекватно описывающим исходные данные***). В противном случае ($F_{расч} < F_{табл}$) уравнение считается незначимым.

Замечания

1. Для нелинейных регрессионных уравнений критерий Стьюдента не проверяется.

2. Для нелинейных регрессионных уравнений анализ остатков не производится. Проводится анализ остатков соответствующих линеаризованных уравнений.

3.3. Характеристики нелинейной регрессионной модели

Сила (теснота) нелинейной регрессионной связи (зависимости) между переменными ***x*** и ***y*** может быть измерена с помощью ***индекса корреляции*** (***корреляционного отношения***)

$$\rho = \sqrt{1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} \qquad (3.19)$$

Индекс корреляции принимает значения от 0 до 1. Чем больше значение индекса корреляции, тем ближе расчетные (теоретические) значения результативного признака $\hat{y}_i$ к фактическим y_i , тем сильнее нелинейная связь между переменными.

Индекс корреляции используется при любой форме нелинейной связи.

Вклад независимой переменной ***x*** в формирование значений ***y*** оценивается с помощью ***коэффициента (индекса) детерминации R²***

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{3.20}$$

Коэффициент детерминации принимает значение от 0 до 1.Чем ближе значение ***R²*** к единице, тем лучше ***x*** описывает изменчивость ***y***.

Для заданной матрицы наблюдений (таблицы исходных данных) можно построить несколько статистически значимых регрессионных уравнений. Выбор уравнения, наилучшим образом описывающего исходные данные, можно производить по наибольшему значению ***скорректированного (нормированного) коэффициента детерминации $R^2_{скорр}$***

$$R^2_{скорр} = 1 - \left(1 - R^2\right)\frac{n-1}{n-m-1} \tag{3.21}$$

Скорректированный коэффициент детерминации обладает свойством

$$0 < R^2_{скорр} < 1 \ .$$

Однако на практике, если построенное нелинейное уравнение не подходит для описания существующей в исходных данных нелинейной зависимости, коэффициент детерминации принимает значение $\boldsymbol{R^2} < 0{,}1$, а скорректированный коэффициент детерминации становится отрицательным ($\boldsymbol{R^2_{скорр}} < 0$) .

Нелинейное регрессионное уравнение должно быть по возможности более простым, чтобы сущность изучаемой зависимости между переменными проявлялась достаточно четко, а параметры уравнения поддавались определенному экономическому толкованию. Вопрос выбора соответствующего уравнения нелинейной связи решается в каждом случае отдельно.

3.4. Приложения нелинейных регрессионных моделей

3.4.1. Эластичность

Наиболее часто встречающееся приложение нелинейных регрессионных моделей - оценка силы влияния объясняющей переменной $\boldsymbol{x}$ на результативный признак $\boldsymbol{y}$.

Для этого используются ***средние*** $\overline{Э}$ и ***частные*** $Э_i$ ***коэффициенты эластичности***.

В общем случае коэффициент эластичности для парного уравнения регрессии $\hat{y} = f(x)$ находится по формуле

$$Э = f'(x)\,\frac{x}{y} \tag{3.22}$$

После дифференцирования, согласно (3.22), правых частей парных нелинейных уравнений регрессии получим следующие формулы (см. табл. 3.1.).

Таблица 3.1. Формулы определения коэффициентов эластичности

Вид уравнения регрессии	**Средний коэффициент эластичности** $\bar{Э}$	**Частные коэффициенты эластичности** $Э_i$
Степенная функция $\hat{y} = b_0 \cdot x^{b1}$	$\bar{Э} = b_1$	$Э_i = b_1$
Показательная функция $\hat{y} = b_0 \cdot b_1^x$	$\bar{Э} = \bar{x} \cdot \ln b_1$	$Э_i = x_i \cdot \ln b_1$
Гипербола $\hat{y} = b_0 + b_1 \frac{1}{x}$	$\bar{Э} = -\frac{b_1}{b_0 \cdot \bar{x} + b_1}$	$Э_i = -\frac{b_1}{b_0 \cdot x_i + b_1}$

Средний по совокупности наблюдений коэффициент эластичности $\bar{Э}$ показывает, на сколько процентов изменится среднее значение зависимой переменной $\bar{y}$ при увеличении среднего значения $\bar{x}$ на 1%. Средний коэффициент эластичности позволяет выявить общегрупповые закономерности.

Частный коэффициент эластичности $Э_i$ показывает, на сколько процентов изменится значение зависимой переменной для $\boldsymbol{i}$-го наблюдения ($\boldsymbol{y_i}$) при увеличении значения фактора для $\boldsymbol{i}$-го наблюдения ($\boldsymbol{x_i}$) на 1%. Частные коэффициенты эластичности выявляют особенности объектов наблюдений.

3.4.2. Прогнозирование

В прогнозных расчетах предсказываемое значение $\boldsymbol{y_{пр}}$ определяется как точечный прогноз $\hat{y}_{пр}$ путем подстановки в уравнение регрессии соответствующего значения факторной переменной $\boldsymbol{x_{пр}}$.

Наибольшая точность прогноза достигается в тех случаях, когда $\boldsymbol{x_{пр}}$ находятся в центре области наблюдений $\boldsymbol{x}$, используемых при построении уравнения регрессии (незначительно отличаются от $\bar{x}$). При удалении $\boldsymbol{x_{пр}}$ от $\bar{x}$ средняя ошибка прогноза растет. В случае, когда $\boldsymbol{x_{пр}}$ оказывается за пределами области наблюдаемых значений, нельзя говорить о надежности $\hat{y}_{пр}$.

3.5. Решение типовых задач

Задача 3.1. Имеются выборочные данные показателей «Трудовой стаж работника» ($\boldsymbol{x}$, год) и «Процент выполнения плана» ($\boldsymbol{y}$, %) (табл. 3.2).

Таблица 3.2. Исходные данные

№ наблюдения	**Процент выполнения плана**	**Трудовой стаж работника**	**№ наблюдения**	**Процент выполнения плана**	**Трудовой стаж работника**
1	84	1	11	110	6
2	92	2	12	102	5
3	80	0,5	13	108	7,5
4	85	2	14	112	8,5
5	94	2,5	15	113	7
6	89	3	16	115	8
7	113	9	17	105	7
8	118	9,5	18	116	11
9	111	3,5	19	121	11
10	102	4	20	122	12

Требуется:

1) Построить степенное $\hat{y} = b_0 \cdot x^{b_1}$ регрессионное уравнение зависимости процента выполнения плана от трудового стажа работника.

Для построенного уравнения регрессии:

2) оценить тесноту нелинейной связи;

3) оценить качество уравнения;

4) найти средние и частные коэффициенты эластичности.

Решение

Объем выборки $\boldsymbol{n}$ = 20, число независимых переменных (факторов) $\boldsymbol{m}$ = 1.

1) Для нахождения параметров $\boldsymbol{b_0}$, $\boldsymbol{b_1}$ уравнения степенной регрессии $\hat{y} = b_0 \cdot x^{b_1}$ приведем уравнение к линейному виду. Введем новые переменные

$$Y = \ln \hat{y},\, X = \ln x,\, B_0 = \ln b_0.$$

Тогда уравнение регрессии примет вид

$$Y = B_0 + b_1 \cdot X$$

Параметры уравнения определим по формулам (3.7 - 3.8) (необходимые расчеты приведены в табл. 3.3):

$$b_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{1}{n} \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2} = \frac{144{,}07 - \frac{1}{20} \cdot 30{,}60 \cdot 92{,}85}{60{,}89 - \frac{1}{20}(30{,}60)^2} = 0{,}14$$

или

$$b_1 = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\overline{X^2} - (\bar{X})^2} = \frac{7{,}20 - 4{,}64 \cdot 1{,}53}{3{,}04 - 1{,}53^2} = 0{,}14$$

$$B_0 = \bar{Y} - b_1 \bar{X} = 4{,}64 - 0{,}14 \cdot 1{,}53 = 4{,}43.$$

Обратный переход к параметру b_0 осуществим по формуле (3.9)

$$b_0 = e^{B_0} = e^{4{,}43} = 83{,}57.$$

Таблица 3.3. Расчетная таблица оценки параметров уравнения степенной регрессии

№ п/п	x_i	y_i	$X_i = \ln x_i$	$Y_i = \ln y_i$	X_iY_i	X_i^2	$\hat{y}_i$
1	1	84	0,00	4,43	0,00	0,00	83,57
2	2	92	0,69	4,52	3,13	0,48	92,19
3	0,5	80	-0,69	4,38	-3,04	0,48	75,76
4	2	85	0,69	4,44	3,08	0,48	92,19
5	2,5	94	0,92	4,54	4,16	0,84	95,15
6	3	89	1,10	4,49	4,93	1,21	97,64
7	9	113	2,20	4,73	10,39	4,83	114,07
8	9,5	118	2,25	4,77	10,74	5,07	114,95
9	3,5	111	1,25	4,71	5,90	1,57	99,79
10	4	102	1,39	4,62	6,41	1,92	101,70
11	6	110	1,79	4,70	8,42	3,21	107,71
12	5	102	1,61	4,62	7,44	2,59	104,96
13	7,5	108	2,01	4,68	9,43	4,06	111,17
14	8,5	112	2,14	4,72	10,10	4,58	113,15
15	7	113	1,95	4,73	9,20	3,79	110,09
16	8	115	2,08	4,74	9,87	4,32	112,19
17	7	105	1,95	4,65	9,06	3,79	110,09
18	11	116	2,40	4,75	11,40	5,75	117,36
19	11	121	2,40	4,80	11,50	5,75	117,36
20	12	122	2,48	4,80	11,94	6,17	118,82
Сумма	**120**	**2092**	**30,60**	**92,85**	**144,07**	**60,89**	**2089,89**
Среднее	**6**	**104,6**	**1,53**	**4,64**			**104,49**

Уравнение степенной регрессии имеет вид

$$\hat{y} = 83{,}57 \cdot x^{0{,}14}.$$

Степенная регрессионная модель имеет вид

$$y = 83{,}57 \cdot x^{0{,}14} + \varepsilon$$

или

$$y_i = 83{,}57 \cdot x_i^{0{,}14} + \varepsilon_i.$$

Коэффициент регрессии b₁ = 0,14 является средним коэффициентом эластичности.

Подставляя в полученное уравнение регрессии значения x_i можно определить ***теоретические значения*** $\hat{y}_i$ и построить линию регрессии на корреляционном поле (рис. 3.1).

Линия степенной регрессии проходит внутри корреляционного поля. Кроме того, число точек корреляционного поля (11), лежащих выше линии регрессии, примерно равно числу точек (9), лежащих ниже линии регрессии. Следовательно, ***линия степенной регрессии занимает правильное положение***.

2) Теснота нелинейной регрессионной зависимости оценим с помощью индекса корреляции (3.19) (необходимые здесь и далее расчеты приведены в табл. 3.4):

$$\rho = \sqrt{1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} = \sqrt{1 - \frac{374{,}87}{3208{,}80}} = 0{,}94$$

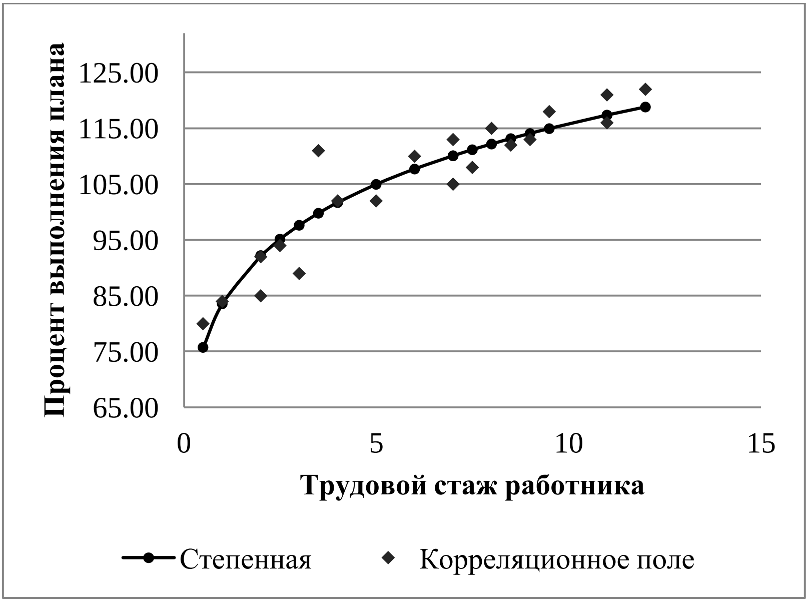


Рис. 3.1. Линия степенной регрессии на корреляционном поле

Значение ρ близко к 1, следовательно, ***степенная связь между процентом выполнения плана и трудовым стажем работников сильная.***

Коэффициент детерминации найдем по формуле (3.20)

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{2720{,}36}{3208{,}80} = 0{,}8478.$$

Полученное значение $\boldsymbol{R^2}$ показывает, что в степенной модели ***формирование значений показателя «Процент выполнения плана» на 84,78% объясняется влиянием фактора «Трудовой стаж работника»***. Оставшиеся

15,22% приходятся на другие факторы, не включенные в модель.

Скорректированный коэффициент детерминации равен (см. формулу (3.21))

$$R^2_{скорр} = 1 - \left(1 - R^2\right)\frac{n-1}{n-m-1} = 1 - (1 - 0{,}8478) \cdot \frac{20-1}{20-1-1} = 0{,}8393$$

Таблица 3.4. Расчетная таблица характеристик степенной модели

№ п/п	x_i	y_i	$\hat{y}_i$	$\left\|\frac{y_i - \hat{y}_i}{y_i}\right\|$	$(\hat{y}_i - \bar{y})^2$	$(\hat{y}_i - y_i)^2$	$(y_i - \bar{y})^2$	$Э_i$
1	1	84	83,57	0,0051	442,31	0,19	424,36	0,14
2	2	92	92,19	0,0020	154,05	0,04	158,76	0,14
3	0,5	80	75,76	0,0531	832,03	18,02	605,16	0,14
4	2	85	92,19	0,0846	154,05	51,67	384,16	0,14
5	2,5	94	95,15	0,0122	89,33	1,32	112,36	0,14
6	3	89	97,64	0,0970	48,48	74,60	243,36	0,14
7	9	113	114,07	0,0095	89,75	1,15	70,56	0,14
8	9,5	118	114,95	0,0258	107,13	9,30	179,56	0,14
9	3,5	111	99,79	0,1010	23,12	125,62	40,96	0,14
10	4	102	101,70	0,0030	8,43	0,09	6,76	0,14
11	6	110	107,71	0,0208	9,66	5,25	29,16	0,14
12	5	102	104,96	0,0290	0,13	8,78	6,76	0,14
13	7,5	108	111,17	0,0293	43,11	10,02	11,56	0,14
14	8,5	112	113,15	0,0103	73,17	1,33	54,76	0,14
15	7	113	110,09	0,0258	30,09	8,50	70,56	0,14
16	8	115	112,19	0,0245	57,56	7,91	108,16	0,14
17	7	105	110,09	0,0484	30,09	25,86	0,16	0,14
18	11	116	117,36	0,0117	162,87	1,86	129,96	0,14
19	11	121	117,36	0,0301	162,87	13,23	268,96	0,14
20	12	122	118,82	0,0261	202,13	10,13	302,76	0,14
Сумма	**120**	**2092**	**2089,89**	**0,6494**	**2720,36**	**374,87**	**3208,80**	
Среднее	**6**	**104,6**	**104,49**	**0,0325**				

3) Оценим качество степенного уравнения регрессии.

Поскольку $\bar{y} = 104{,}6 \approx \bar{\hat{y}} = 104{,}49$ ***расчет параметров b₀, b₁ проведен верно***.

Найдем среднюю относительную ошибку аппроксимации по формуле (3.17)

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\% = \frac{0{,}6494}{20} \cdot 100\% = 3{,}25\% \ .$$

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

Проверим статистическую значимость уравнения регрессии в целом с помощью критерия Фишера. Расчетное значение (статистика) критерия Фишера (см. формулу (3.18)) равно

$$F_{расч} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \cdot \frac{n-m-1}{m} = \frac{2720{,}36}{374{,}87} \cdot \frac{20-1-1}{1} = 130{,}62$$

Табличное значение критерия Фишера с $\boldsymbol{df_1} = \boldsymbol{m} = 1$ и $\boldsymbol{df_2} = \boldsymbol{n} - \boldsymbol{m} - 1 = 20 - 1 - 1 = 18$ степенями свободы при уровне значимости $\boldsymbol{\alpha} = 0{,}05$ найдем по табл. Приложения 3 или с помощью встроенной функции Excel «FРАСПОБР». $\boldsymbol{F_{табл}}(0{,}05;\ 1;\ 18) = 4{,}41$.

Поскольку $\boldsymbol{F_{расч}} > \boldsymbol{F_{табл}}$, ***уравнение степенной регрессии статистически значимо в целом***, т.е. адекватно описывает исходные данные.

4) Средний и частные коэффициенты эластичности в степенной модели (см. формулы табл. 3.1)

$$\overline{Э} = b_1 \text{ , } Э_i = b_1 \text{ ,}$$

постоянны и равны $b_1 = 0{,}14$, т.е. ***при увеличении среднего трудового стажа работника на 1% процент выполнения плана увеличится на 0,14%.***

Задача 3.2. В условиях задачи 3.1.

Требуется:

1) построить показательное $\hat{y} = b_0 \cdot b_1^x$ регрессионное уравнение зависимости процента выполнения плана от трудового стажа работника.

Для построенного уравнения регрессии:

2) оценить тесноту нелинейной связи;

3) оценить качество уравнения;

4) найти средние и частные коэффициенты эластичности.

Решение

Объем выборки $\boldsymbol{n} = 20$, число независимых переменных (факторов) $\boldsymbol{m} = 1$.

1) Для нахождения параметров $\boldsymbol{b_0}$***,*** $\boldsymbol{b_1}$ уравнения показательной регрессии $\hat{y} = b_0 \cdot b_1^x$ приведем уравнение к линейному виду.

Введем новые переменные

$$Y = \ln \hat{y}, \; B_0 = \ln b_0, \; B_1 = \ln b_1 .$$

Тогда уравнение регрессии примет вид

$$Y = B_0 + B_1 \cdot x .$$

Параметры уравнения определим по формулам (3.11 – 3.12) (необходимые расчеты приведены в табл. 3.5):

$$B_1 = \frac{\sum_{i=1}^{n} x_i Y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{565,10 - \frac{1}{20} \cdot 120 \cdot 92,85}{961,50 - \frac{1}{20} 120^2} = 0,033$$

или

$$B_1 = \frac{\overline{xY} - \bar{x} \cdot \bar{Y}}{\overline{x^2} - (\bar{x})^2} = \frac{28,26 - 6 \cdot 4,64}{48,07 - 6^2} = 0,033$$

$$B_0 = \bar{Y} - B_1 \bar{x} = 4,64 - 0,033 \cdot 6 = 4,44.$$

Обратный переход к параметру b_0 осуществим по формулам (3.13):

$$b_0 = e^{B_0} = e^{4,44} = 85,06, \; b_1 = e^{B_1} = e^{0,033} = 1,03 .$$

Уравнение показательной регрессии имеет вид:

$$\hat{y} = 85,06 \cdot 1,03^x .$$

Показательная регрессионная модель имеет вид

$$y = 85{,}06 \cdot 1{,}03^{x} + \varepsilon$$

или

$$y_i = 85{,}06 \cdot 1{,}03^{xi} + \varepsilon_i$$

Подставляя в полученное уравнение регрессии значения x_i можно определить ***теоретические значения*** $\hat{y}_i$ и построить линию регрессии на корреляционном поле (рис. 3.2).

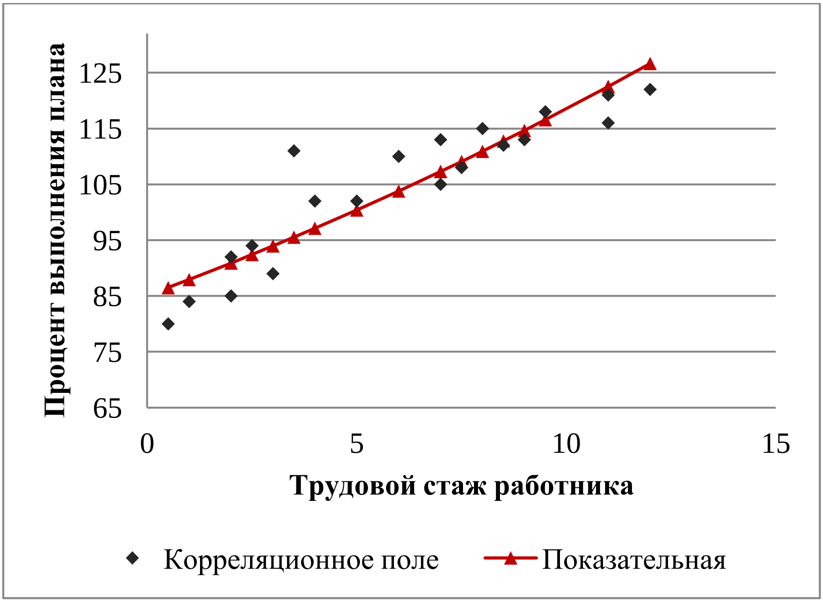


Рис. 3.2. Линия показательной регрессии на корреляционном поле

Линия показательной регрессии проходит внутри корреляционного поля. Кроме того, число точек корреляционного поля (9), лежащих выше линии регрессии, примерно равно числу точек (11), лежащих ниже линии регрессии. Следовательно, ***линия показательной регрессии занимает правильное положение***.

Таблица 3.5. Расчетная таблица параметров уравнения показательной регрессии

№ п/п	x_i	y_i	$Y_i = \ln y_i$	$x_i \cdot Y_i$	x_i^2	$\hat{y}_i$
1	1	84	4,43	4,43	1,00	87,93
2	2	92	4,52	9,04	4,00	90,89
3	0,5	80	4,38	2,19	0,25	86,48
4	2	85	4,44	8,89	4,00	90,89
5	2,5	94	4,54	11,36	6,25	92,41
6	3	89	4,49	13,47	9,00	93,96
7	9	113	4,73	42,55	81,00	114,65
8	9,5	118	4,77	45,32	90,25	116,57
9	3,5	111	4,71	16,48	12,25	95,53
10	4	102	4,62	18,50	16,00	97,13
11	6	110	4,70	28,20	36,00	103,79
12	5	102	4,62	23,12	25,00	100,40
13	7,5	108	4,68	35,12	56,25	109,09
14	8,5	112	4,72	40,11	72,25	112,77
15	7	113	4,73	33,09	49,00	107,29
16	8	115	4,74	37,96	64,00	110,91
17	7	105	4,65	32,58	49,00	107,29
18	11	116	4,75	52,29	121,00	122,52
19	11	121	4,80	52,75	121,00	122,52
20	12	122	4,80	57,65	144,00	126,65
Сумма	**120**	**2092**	**92,85**	**565,10**	**961,50**	2089,67
Среднее	**6**	**104,6**	**4,64**			**104,48**

2) Тесноту показательной регрессионной зависимости оценим с помощью индекса корреляции (3.19) (необходимые здесь и далее расчеты приведены в табл. 3.6):

$$\rho = \sqrt{1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} = \sqrt{1 - \frac{552{,}14}{3208{,}80}} = 0{,}91 \ .$$

Значение ρ близко к 1, следовательно, ***показательная связь между процентом выполнения плана и трудовым стажем работников сильная.***

Коэффициент детерминации (3.20)

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{2914{,}20}{3208{,}80} = 0{,}9082$$

показывает, что в показательной модели ***формирование значений показателя «Процент выполнения плана» на 90,82% объясняется влиянием фактора «Трудовой стаж работника»***. Оставшиеся 9,18% приходятся на другие факторы, не включенные в модель.

Скорректированный коэффициент детерминации равен (3.21)

$$R^2_{скорр} = 1 - \left(1 - R^2\right)\frac{n-1}{n-m-1} = 1 - (1 - 0{,}9082) \cdot \frac{20-1}{20-1-1} = 0{,}9031$$

Таблица 3.6. Расчетная таблица характеристик показательной модели

№ п/п	x_i	y_i	$\hat{y}_i$	$\left\lvert\frac{y_i-\hat{y}_i}{y_i}\right\rvert$	$(\hat{y}_i-\bar{y})^2$	$(\hat{y}_i-y_i)^2$	$(y_i-\bar{y})^2$	$Э_i$
1	1	84	87,93	0,0467	278,02	15,41	424,36	***0,03***
2	2	92	90,89	0,0120	187,91	1,23	158,76	0,07
3	0,5	80	86,48	0,0810	328,35	41,98	605,16	***0,02***
4	2	85	90,89	0,0693	187,91	34,72	384,16	0,07
5	2,5	94	92,41	0,0169	148,54	2,52	112,36	0,08
6	3	89	93,96	0,0557	113,25	24,58	243,36	0,10
7	9	113	114,65	0,0146	101,05	2,73	70,56	0,30
8	9,5	118	116,57	0,0121	143,29	2,04	179,56	0,32
9	3,5	111	95,53	0,1394	82,27	239,33	40,96	0,12
10	4	102	97,13	0,0478	55,84	23,74	6,76	0,13
11	6	110	103,79	0,0564	0,65	38,55	29,16	0,20
12	5	102	100,40	0,0156	17,61	2,55	6,76	0,17
13	7,5	108	109,09	0,0101	20,13	1,18	11,56	0,25
14	8,5	112	112,77	0,0068	66,69	0,59	54,76	0,28
15	7	113	107,29	0,0505	7,25	32,58	70,56	0,23
16	8	115	110,91	0,0356	39,83	16,72	108,16	0,27
17	7	105	107,29	0,0218	7,25	5,25	0,16	0,23
18	11	116	122,52	0,0562	321,06	42,49	129,96	**0,36**
19	11	121	122,52	0,0125	321,06	2,30	268,96	**0,36**
20	12	122	126,65	0,0381	486,25	21,63	302,76	**0,40**
Сумма	**120**	**2092**	**2089,67**	**0,7993**	**2914,2**	**552,14**	**3208,8**	
Среднее	**6**	**104,6**	**104,48**	**0,0400**				**0,199**

3) Оценим качество показательного уравнения регрессии.

Поскольку $\bar{y} = 104{,}6 \approx \bar{\hat{y}} = 104{,}48$ ***расчет параметров b_0, b_1 проведен верно***.

Найдем среднюю относительную ошибку аппроксимации (3.17)

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\% = \frac{0,7993}{20} \cdot 100\% = 4,00\% \ .$$

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

Проверим статистическую значимость уравнения регрессии в целом с помощью критерия Фишера. Расчетное значение (статистика) критерия Фишера (3.18) равно

$$F_{расч} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2} \cdot \frac{n-m-1}{m} = \frac{2914,20}{552,14} \cdot \frac{20-1-1}{1} = 95,00 .$$

Табличное значение критерия Фишера с $\boldsymbol{df_1} = \boldsymbol{m} = 1$ и $\boldsymbol{df_2} = \boldsymbol{n} - \boldsymbol{m} - 1 = 20 - 1 - 1 = 18$ степенями свободы при уровне значимости $\boldsymbol{\alpha} = 0,05$ найдем по табл. Приложения 5 или с помощью встроенной функции Excel «FРАСПОБР». $\boldsymbol{F_{табл}}(0,05; 1; 18) = 4,41$.

Поскольку $\boldsymbol{F_{расч}} > \boldsymbol{F_{табл}}$, ***уравнение показательной регрессии статистически значимо в целом***, т.е. адекватно описывает исходные данные.

4) Средний и частные коэффициенты эластичности в показательной модели найдем по формулам табл.3.2

$$\overline{Э} = \overline{x} \ln b_1 = 6 \cdot 0,033 = 0,199058 \approx 0,20$$

$$Э_i = x_i \cdot \ln b_1 = 0,033 \cdot x_i$$

Значения частных коэффициентов эластичности приведены в табл. 3.6.

Необходимо отметить, что среднее значение, вычисленное по столбцу частных коэффициентов эластичности, как правило, не совпадает со значением среднего коэффициента эластичности. Для анализа необходимо использовать средний коэффициент эластичности, вычисленный по формуле.

Средний коэффициент эластичности, полученный для показательной модели, показывает, что ***при увеличении среднего трудового стажа работников на 1% процент выполнения плана в среднем увеличится на 0,20%.***

Анализ значений частных коэффициентов эластичности показывает, что ***для 1-го и 3-го наблюдения показатель «трудовой стаж работника» практически не влияет на показатель «процент выполнения плана»*** ($Э_1$=0,03, $Э_3$=0,02) (для данных наблюдений влияние х на у наименьшее).

Для 18-го, 19-го и 20-го наблюдений увеличение показателя «трудовой стаж» на 1% приводит к наибольшему увеличению процента выполнения плана, чем в целом по группе наблюдений ($Э_{18}$=0,36, $Э_{19}$=0,36, $Э_{20}$=0,40) (для данных наблюдений влияние х на у наибольшее).

Другими словами, в рамках построенной показательной модели, у работников, имеющих небольшой стаж работы, менее трех лет, увеличение стажа практически не приводит к росту производительности труда. В тоже время у опытных работников, имеющих стаж работы более десяти лет, увеличение стажа приводит к значительно большему росту производительности труда.

Задача 3.3. В условиях задачи 3.1.

Требуется:

1) Построить гиперболическое $\hat{y} = b_0 + b_1 \frac{1}{x}$ регрессионное уравнение зависимости процента выполнения плана от трудового стажа работника.

Для построенного уравнения регрессии:

2) оценить тесноту нелинейной связи;

3) оценить качество уравнения;

4) найти средние и частные коэффициенты эластичности.

Решение

Объем выборки $\boldsymbol{n}$ = 20, число независимых переменных (факторов) $\boldsymbol{m}$ = 1.

1) Для нахождения параметров $\boldsymbol{b_0}$, $\boldsymbol{b_1}$ уравнения гиперболической регрессии $\hat{y} = b_0 + b_1 \frac{1}{x}$ приведем уравнение к линейному виду.

Введем новую переменную

$$X = \frac{1}{x}.$$

Тогда уравнение регрессии примет вид

$$\hat{y} = b_0 + b_1 \cdot X.$$

Параметры уравнения определим по формулам (3.15 - 3.16) (необходимые расчеты приведены в табл. 3.7):

$$b_1 = \frac{\sum_{i=1}^{n} X_i y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}(\sum_{i=1}^{n} X_i)^2} = \frac{625{,}50 - \frac{1}{20}\cdot 6{,}78 \cdot 2092}{612 - \frac{1}{20}6{,}78^2} = -21{,}88$$

или

$$b_1 = \frac{\overline{Xy} - \bar{X}\cdot\bar{y}}{\overline{X^2} - (\bar{X})^2} = \frac{31{,}32 - 0{,}34\cdot 104{,}6}{0{,}31 - 0{,}34^2} = -21{,}88$$

$$b_0 = \bar{y} - b_1\bar{X} = 2092 - (-21{,}88)\cdot 0{,}34 = 112{,}02.,$$

Уравнение гиперболической регрессии имеет вид

$$\hat{y} = 112{,}02 - 21{,}88\cdot\frac{1}{x}.$$

Гиперболическая регрессионная модель имеет вид

$$y \quad = 112{,}02 - 21{,}88\cdot\frac{1}{x} + \varepsilon$$

или

$$y_i = 112{,}02 - 21{,}88\cdot\frac{1}{x_i} + \varepsilon_i$$

Подставляя в полученное уравнение регрессии значения x_i можно определить ***теоретические значения*** $\hat{y}_i$ и построить линию регрессии на корреляционном поле (рис. 3.3).

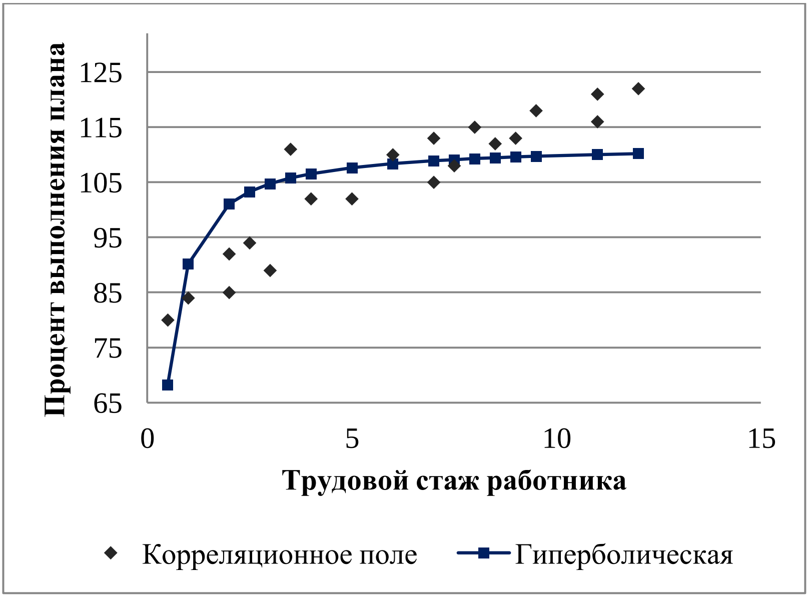


Рис. 3.3. Линия гиперболической регрессии на корреляционном поле

Линия показательной регрессии проходит внутри корреляционного поля. Кроме того, число точек корреляционного поля (10), лежащих выше линии регрессии, примерно равно числу точек (10), лежащих ниже линии регрессии. Следовательно, линия регрессии занимает правильное положение.

2) Тесноту гиперболической регрессионной зависимости оценим с помощью индекса корреляции (3.19) (необходимые здесь и далее расчеты приведены в табл. 3.8):

$$\rho = \sqrt{1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}} = \sqrt{1 - \frac{1380{,}27}{3208{,}80}} = 0{,}75 .$$

Значение ρ одинаково близко к 1 и к 0,5, следовательно, ***гиперболическая связь между процентом выполнения плана и трудовым стажем работников достаточно сильная, но близка к средней.***

Таблица 3.7. Расчетная таблица параметров уравнения гиперболической регрессии

№ п/п	x_i	y_i	$X_i = \frac{1}{x_i}$	$X_i \cdot y_i$	X_i^2	$\hat{y}_i$
1	1	84	1,00	84,00	1,00	90,14
2	2	92	0,50	46,00	0,25	101,08
3	0,5	80	2,00	160,00	4,00	68,26
4	2	85	0,50	42,50	0,25	101,08
5	2,5	94	0,40	37,60	0,16	103,26
6	3	89	0,33	29,67	0,11	104,72
7	9	113	0,11	12,56	0,01	109,58
8	9,5	118	0,11	12,42	0,01	109,71
9	3,5	111	0,29	31,71	0,08	105,76
10	4	102	0,25	25,50	0,06	106,55
11	6	110	0,17	18,33	0,03	108,37
12	5	102	0,20	20,40	0,04	107,64
13	7,5	108	0,13	14,40	0,02	109,10
14	8,5	112	0,12	13,18	0,01	109,44
15	7	113	0,14	16,14	0,02	108,89
16	8	115	0,13	14,38	0,02	109,28
17	7	105	0,14	15,00	0,02	108,89
18	11	116	0,09	10,55	0,01	110,03
19	11	121	0,09	11,00	0,01	110,03
20	12	122	0,08	10,17	0,01	110,19
Сумма	120	**2092**	**6,78**	**625,50**	**6,12**	2092
Среднее	6	**104,6**	**0,34**			**104,6**

Коэффициент детерминации (3.20)

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{1828{,}53}{3208{,}80} = 0{,}5698$$

показывает, что в гиперболической модели ***формирование значений показателя «Процент выполнения плана» на 56,98% объясняется влиянием фактора «Трудовой стаж работника»***. Оставшиеся 43,02% приходятся на другие факторы, не включенные в модель.

Скорректированный коэффициент детерминации равен (3.21)

$$R^2_{скорр} = 1 - \left(1 - R^2\right)\frac{n-1}{n-m-1} = 1 - (1 - 0{,}5698)\cdot\frac{20-1}{20-1-1} = 0{,}5460$$

3) Оценим качество гиперболического уравнения регрессии.

Поскольку $\bar{y} = 104{,}6 \approx \bar{\hat{y}} = 104{,}6$ ***расчет параметров b_0, b_1 проведен верно***.

Найдем среднюю относительную ошибку аппроксимации (3.17)

$$\bar{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|\cdot 100\% = \frac{1{,}4403}{20}\cdot 100\% = 7{,}20\% \ .$$

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

Таблица 3.8. Расчетная таблица характеристик гиперболической модели

№ п/п	x_i	y_i	$\hat{y}_i$	$\left\|\frac{y_i - \hat{y}_i}{y_i}\right\|$	$(\hat{y}_i - \bar{y})^2$	$(\hat{y}_i - y_i)^2$	$(y_i - \bar{y})^2$	$Э_i$
1	1	84	90,14	0,0731	209,16	37,67	424,36	0,24
2	2	92	101,08	0,0987	12,41	82,38	158,76	0,11
3	0,5	80	68,26	0,1468	1320,60	137,83	605,16	***0,64***
4	2	85	101,08	0,1891	12,41	258,45	384,16	0,11
5	2,5	94	103,26	0,0986	1,78	85,83	112,36	0,08
6	3	89	104,72	0,1767	0,02	247,21	243,36	0,07
7	9	113	109,58	0,0302	24,85	11,67	70,56	**0,02**
8	9,5	118	109,71	0,0702	26,14	68,68	179,56	**0,02**
9	3,5	111	105,76	0,0472	1,36	27,41	40,96	0,06
10	4	102	106,55	0,0446	3,79	20,67	6,76	0,05
11	6	110	108,37	0,0148	14,21	2,66	29,16	0,03
12	5	102	107,64	0,0553	9,24	31,81	6,76	0,04
13	7,5	108	109,10	0,0102	20,24	1,21	11,56	0,03
14	8,5	112	109,44	0,0228	23,44	6,55	54,76	**0,02**
15	7	113	108,89	0,0364	18,40	16,89	70,56	0,03
16	8	115	109,28	0,0497	21,91	32,71	108,16	0,03
17	**7**	105	108,89	0,0370	18,40	15,13	0,16	0,03
18	11	116	110,03	0,0515	29,45	35,68	129,96	**0,02**
19	11	121	110,03	0,0907	29,45	120,42	268,96	**0,02**
20	12	122	110,19	0,0968	31,27	139,42	302,76	**0,02**
Сумма	**120**	**2092**	2092	**1,4403**	**1828,53**	**1380,27**	**3208,80**	
Среднее	**6**	**104,6**	**104,6**	**0,0720**				**0,0834**

Проверим статистическую значимость уравнения регрессии в целом с помощью критерия Фишера. Расчетное значение (статистика) критерия Фишера (3.18) равно

$$F_{расч} = \frac{\sum\limits_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum\limits_{i=1}^{n}(\hat{y}_i - y_i)^2} \cdot \frac{n-m-1}{m} = \frac{1828{,}53}{1380{,}27} \cdot \frac{20-1-1}{1} = 23{,}85 \cdot$$

Табличное значение критерия Фишера с $\boldsymbol{df_1} = \boldsymbol{m} = 1$ и $\boldsymbol{df_2} = \boldsymbol{n} - \boldsymbol{m} - 1 = 20 - 1 - 1 = 18$ степенями свободы при уровне значимости $\boldsymbol{\alpha} = 0{,}05$ найдем по табл. Приложения 5 или с помощью встроенной функции Excel «FРАСПОБР». $\boldsymbol{F_{табл}}(0{,}05;\ 1;\ 18) = 4{,}41$.

Поскольку $\boldsymbol{F_{расч}} > \boldsymbol{F_{табл}}$, ***уравнение показательной регрессии статистически значимо в целом***, т.е. адекватно описывает исходные данные.

4) Средний и частные коэффициенты эластичности в гиперболической модели найдем по формулам табл. 3.1.

$$\overline{Э} = -\frac{b_1}{b_0 \cdot \overline{x} + b_1} = -\frac{-21{,}88}{112{,}02 \cdot 6 + (-21{,}88)} = 0{,}03365 \approx 0{,}03$$

$$Э_i = -\frac{b_1}{b_0 \cdot x_i + b_1} = -\frac{-21{,}88}{112{,}02 \cdot x_i + (-21{,}88)}$$

Значения частных коэффициентов эластичности приведены в табл. 3.8.

Необходимо отметить, что среднее значение, вычисленное по столбцу частных коэффициентов эластичности, не совпадает со значением среднего коэффициента эластичности ($0{,}08 \neq 0{,}03$). Для анализа необходимо использовать средний коэффициент эластичности, вычисленный по формуле.

Средний коэффициент эластичности, найденный для гиперболической модели, показывает, что ***при увеличении среднего трудового стажа работников на 1% процент выполнения плана в среднем увеличится на 0,03% (влияние трудового стажа на рост производительности***

труда в целом по группе наблюдений практически отсутствует).

Анализ значений частных коэффициентов эластичности показывает, что ***для 7, 8, 14, 18, 19 и 20-го наблюдений показатель «трудовой стаж работника» имеет наименьшее влияние на показатель «процент выполнения плана»*** (все частные коэффициенты эластичности равны 0,02).

Для 3-го наблюдения увеличение показателя «трудовой стаж» на 1% приводит к наибольшему увеличению процента выполнения плана, чем в целом по группе наблюдений ($Э_3$=0,64).

Другими словами, в рамках построенной гиперболической модели, у работников, имеющих небольшой стаж работы (два года и менее), увеличение стажа приводит к быстрому росту производительности труда. В тоже время у опытных работников, имеющих стаж работы шесть и более лет, увеличение стажа практически не приводит к росту производительности труда.

Задача 3.4. В задачах 3.1. - 3.3. для моделирования зависимости показателя «Процент выполнения плана» (***y***, %) от фактора «Трудовой стаж работника» (***x***, год) построены нелинейные уравнения регрессии:

степенное - $\hat{y} = 83{,}57 \cdot x^{0{,}14}$;

показательное - $\hat{y} = 85{,}06 \cdot 1{,}03^{x}$;

гиперболическое - $\hat{y} = 112{,}02 - 21{,}88 \cdot \dfrac{1}{x}$.

Требуется

выбрать уравнение регрессии, наилучшим образом описывающее данные наблюдений (см. табл. 3.2).

Решение

Постоим линии нелинейной регрессии на одном корреляционном поле

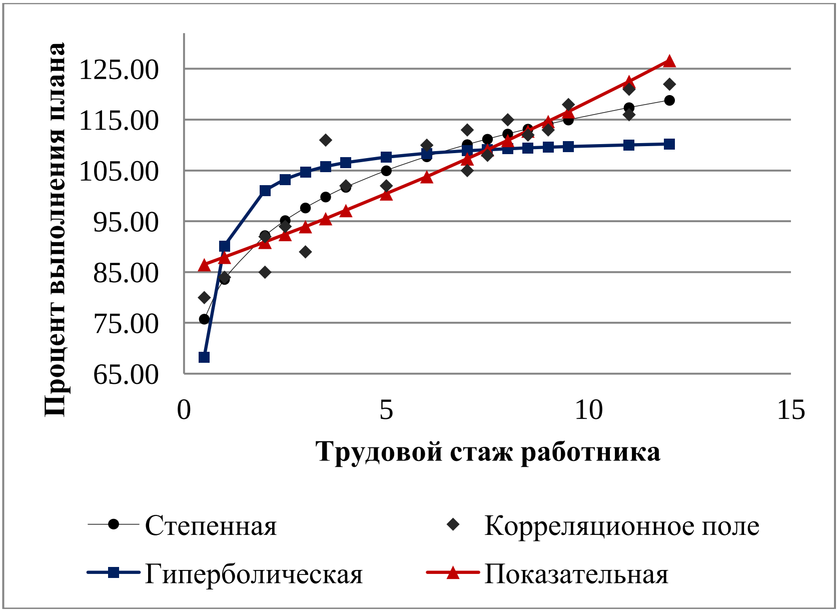


Рис. 3.4. Линии регрессии на корреляционном поле

Результаты расчетов характеристик нелинейных моделей, полученные в задачах 3.1. – 3.3., сведем в табл. 3.9.

Наилучшее регрессионное уравнение выбирается по наибольшему значению скорректированного коэффициента детерминации.

Степенная модель имеет наилучшие значения модельных характеристик: наименьшую среднюю ошибку аппроксимации (наилучшая математическая точность); наибольший индекс корреляции (наиболее сильная нелинейная связь); наибольшее расчетное значение критерия Фишера (наиболее адекватное описание исходных данных). В тоже время скорректированный коэффициент детерминации для нее меньше, чем для показательной модели.

Таблица 3.9. Характеристики нелинейных регрессионных моделей

Вид уравнения регрессии	**Скорректированный коэффициент детерминации,** $R^2_{\text{скорр}}$	**F-статистика,** $F_{\text{расч}}$	**Средняя относительная ошибка аппроксимации,** $\bar{A}$**, (%)**	**Индекс корреляции,** $\boldsymbol{\rho}$
степенное	0,84	**130,62**	**3,25**	**0,94**
показательное	**0,90**	95,00	4,00	0,91
гиперболическое	0,55	23,85	7,20	0,75

Несмотря на то, что формально наилучшая модель выбирается по наибольшему значению скорректированного коэффициента детерминации, в данном случае ***будем считать наилучшей степенную модель***, поскольку она правильнее отражает закономерность в исходных данных, отраженную на корреляционном поле (см. рис. 3.4)

Для практического применения следует использовать выводы о степени влияния трудового стажа на процент выполнения плана, сделанные по коэффициентам эластичности, которые вычислены для степенной модели.

Замечание

При решении практических задач, необходимо корректировать выводы, полученные на основе результатов расчетов, с учетом особенностей исследуемой производственной деятельности.

4. Задания для самостоятельной работы

4.1 Упражнения

Упражнение 1. Рассчитанный по выборочным данным коэффициент корреляции оказался равным а) -1; б) 0,95; в) 1,2. Это означает, что (для каждого случая выберите один ответ):

1) между изучаемыми переменными есть слабая отрицательная линейная связь;

2) между изучаемыми переменными есть связь, но она не является линейной;

3) между изучаемыми переменными есть функциональная линейная отрицательная связь;

4) между изучаемыми переменными отсутствует связь;

5) полученное число никак не интерпретируется, допущена ошибка в вычислениях;

6) между изучаемыми переменными есть функциональная положительная зависимость;

7) между изучаемыми переменными достаточно сильная прямая линейная зависимость;

8) ни один из предложенных ответов не является правильным.

Упражнение 2. Студенты решали задачи на исследование корреляционной зависимости между двумя экономическими показателями и получили следующие результаты:

Студент 1	коэффициент корреляции равен 0,3 корреляционное отношение равно 0,8
Студент 2	коэффициент корреляции равен -1,2 корреляционное отношение равно -0,9
Студент 3	коэффициент корреляции равен -0,8 корреляционное отношение равно -0,8
Студент 4	коэффициент корреляции равен 0,9 корреляционное отношение равно 0,8
Студент 5	коэффициент корреляции равен 0,6 корреляционное отношение равно -0,4
Студент 6	коэффициент корреляции равен -0,3 корреляционное отношение равно 0,8

Какой вывод должен сделать каждый студент:

1) между показателями существует прямая линейная корреляционная зависимость;

2) между показателями существует обратная линейная корреляционная зависимость;

3) между показателями существует криволинейная корреляционная зависимость;

4) ни один из предложенных вариантов не подходит, ошибка в расчетах.

Упражнение 3. При исследовании взаимосвязи между числом заявок на недельные ссуды и текущей ставкой процента по закладным из 200 недель последних 5 лет случайным образом были выбраны 15. На основе отобранных данных была построена линейная модель регрессии и определены следующие параметры модели: $b_0 = 152{,}399$, $b_1 = -6{,}811$, $R^2 = 0{,}945$.

Объясните менеджеру, не имеющему статистической подготовки, экономический смысл полученных показателей.

Упражнение 4. Построена модель регрессии, отражающая влияние рекламы на спрос:

$$\hat{y} = 3{,}5 - 1{,}2 \cdot x \quad , \quad R^2 = 0{,}9.$$

Выбрать правильный вариант ответа. Значение коэффициента регрессии можно интерпретировать следующим образом:

1) реклама оказывает стимулирующее воздействие на спрос;

2) реклама оказывает очень слабое влияние на спрос;

3) реклама вызвала негативное отношение покупателей к данному товару;

4) по приведенным данным нельзя вообще сделать никаких выводов о влиянии рекламы.

Упражнение 5. Построена модель регрессии, отражающая зависимость спроса на путевки (в шт.) от изменения цены на них (в тыс. руб.):

$\hat{y} = 2{,}8 \cdot x^{-1{,}0}$, где у - спрос на путевки; х - цена на путевки.

Выбрать правильный вариант ответа. Значение коэффициента регрессии можно интерпретировать следующим образом:

1) при изменении цены спрос на путевки не изменяется;

2) при увеличении цены на 1 руб. спрос уменьшится в среднем на 1 путевку;

3) при увеличении цены на 1000 руб. спрос в среднем уменьшится на 1 путевку;

4) при увеличении цены на 1% спрос в среднем увеличится на 1%;

5) при увеличении цены на 1% спрос в среднем увеличится на 1 путевку.

Упражнение 6. Зависимость спроса на товар К от его цены характеризуется по 30 наблюдениям уравнением:

$$lg\ y = 1{,}5 - 0{,}85 \cdot lg\ x\ .$$

Доля остаточной дисперсии в общей составила 28%:

1) записать модель в исходном виде;

2) дать интерпретацию коэффициента регрессии.

Упражнение 7. По 20 наблюдениям было получено уравнение регрессии:

$$\hat{y} = 8{,}4 - 4{,}5 \cdot x \quad , \text{r} = -0{,}81.$$

Дать интерпретацию параметров.

Упражнение 8. По 12 фермам, производящих молочную продукцию, получено уравнение регрессии для себестоимости y единицы продукции (тыс. руб.) от уровня x технической оснащенности (тыс. руб.):

$$\hat{y} = 10 + \frac{600}{x} \ .$$

Дать интерпретацию параметров.

4.2. Задачи

Задача 4.1. Имеются данные о среднегодовых темпах роста выпуска валовой продукции по десяти отраслям и темпах роста производительности труда на одного работника. Определить в целом степень тесноты связи между производительностью труда и выпуском валовой продукции. Исходные данные представлены в табл. 4.1.

Таблица 4.1

Отрасль	**Среднегодовой темп роста (%)**	
	Валовая продукция	**Производительность одного работника**
Электроэнергетика	16,5	11,9
Черная металлургия	13,6	8,1
Машиностроение	18,8	13,2
Химическая	22,3	11,3
Строительство	15,6	1,4
Лесозаготовки	11,1	7,7
Бумажно-целлюлозная	11,6	5,8
Текстильная	9,3	5,7
Пищевая	8,9	5,6
Полиграфическая	14,0	8,4

Задача 4.2. Основываясь на данных табл. 4.2, проанализировать наличие и тесноту корреляционной связи между парой признаков, проведя полный корреляционный анализ (построить корреляционное поле, рассчитать парный коэффициент корреляции, оценить его статистическую значимость, построить доверительный интервал).

Таблица 4.2

X	2	3,5	1	2	4	3,5	1	3	4	5
Y	5	7	5	5	7,5	7	5,5	6	7	8

Задача 4.3. Компания владеет 12 магазинами. Финансовый директор рассматривает возможность слияния мелких магазинов для увеличения прибыльности кампании. Для принятия решения ему необходимо установить связь между прибылью и оборотом. Данные для каждого магазина за последний финансовый год в табл. 4.3.

Таблица 4.3

Магазины	**Месячная прибыль (тыс. руб.)**	**Оборот (тыс. руб.)**
1	2	50
2	4	60
3	11	85
4	17	85
5	18	100
6	28	120
7	34	140
8	36	155
9	48	180
10	55	210
11	71	250
12	85	300

Задача 4.4. Туристическую фирму крупного курортного города интересует связь между числом отпускников, останавливающихся в отелях, и расходами на рекламу отелей. Взято случайное число отелей 8, сходных по размерам (см. табл. 4.4).

Таблица 4.4

Отель	1	2	3	4	5	6	7	8
Реклама (тыс. долл.)	9	6	10	8	7	4	11	9
Число гостей	1100	1200	1600	1300	1100	800	1200	1500

Построить модель для объяснения влияния рекламы на выбор отеля. Прокомментировать результаты анализа.

Задача 4.5. Для данных табл. 4.5

1. Построить уравнение линейной регрессии, описывающие зависимость прибыли банка (y) от объема межбанковских кредитов и депозитов (x). Пояснить экономический смысл коэффициента регрессии.

2. Оценить степень тесноты связи между переменными с помощью коэффициента корреляции. Определить объясняемость значений зависимой переменной за счет влияния независимой переменной.

3. Проверить адекватность модели с помощью F-критерия Фишера на уровне значимости 0,05.

4. Рассчитать:

- стандартную ошибку регрессии;

- стандартные ошибки оценок коэффициентов уравнения регрессии.

Проверить значимость коэффициентов уравнения регрессии и построить их интервальные оценки на уровне значимости 0,05. Для этого рассчитать значения t-статистик для оценок коэффициентов уравнения регрессии.

4. Найти стандартную ошибку и доверительный интервал для прогнозного значения $y_{пр}$ при $x_{пр} = 3$.

Таблица 4.5

№	1	2	3	4	5	6	7	8
X	2	2,1	2,3	2,4	2,9	3,3	3,8	4,6
Y	30,78	31,71	32,70	30,19	34,69	37,17	40,38	44,33

Задача 4.6. На крупном промышленном предприятии проводятся курсы технической подготовки, предназначенные для всех принятых работников рабочих специальностей. В табл. 4.6 приведен возраст 11 работников, выбранных произвольно, а также время, необходимое для выработки у них навыков в определенной области.

Таблица 4.6

Номер работника	**Возраст (лет)**	**Время подготовки (часов)**
1	18	4
2	19	3
3	20	4
4	21	6
5	22	5
6	23	6
7	27	6
8	29	7
9	32	7
10	35	8
11	38	7

1) Построить нелинейные уравнения зависимости между возрастом работника (x) и временем, необходимым для освоения определенных навыков и умений (y):

степенное $\hat{y} = b_0 \cdot x^{b1}$;

показательное $\hat{y} = b_0 \cdot b_1^x$;

гиперболическое $\hat{y} = b_0 + b_1 \dfrac{1}{x}$.

2) Для построенных уравнений регрессии:

- оценить тесноту нелинейной связи;

- оценить качество уравнения;

- найти средние и частные коэффициенты эластичности, сделать вывод.

3) Среди построенных уравнений выбрать уравнение наилучшим образом описывающее зависимость времени, необходимого для освоения определенных навыков и умений, от возраста работника.

4.3. Вопросы и задания для самоконтроля

1. Записать модель парной линейной регрессии.

2. В чем различие между уравнением регрессии и выборочным уравнением регрессии?

3. Что понимается под спецификацией модели?

4. Какие преимущества имеет линейный вид зависимости по отношению к нелинейным?

5. Сформулируйте предпосылки МНК.

6. Запишите систему нормальных уравнений.

7. Запишите формулу нахождения коэффициента регрессии.

8. Каков смысл коэффициента регрессии?

9. Запишите формулу нахождения коэффициента корреляции.

10. Для чего применяют коэффициент корреляции?

11. Каковы основные свойства коэффициента корреляции?

12. Как проверяется значимость коэффициента корреляции?

13. Каковы основные свойства оценок модели линейной регрессии? Дайте их определение.

14. Сформулируйте основные гипотезы, накладываемые на переменные линейной регрессионной модели.

15. Дайте формулировку теоремы Гаусса-Маркова.

16. Какими свойствами обладают оценки параметров уравнения линейной регрессии?

17. Когда модель будет называться классической нормальной регрессионной моделью?

18. Что служит оценкой дисперсии случайных остатков?

19.Запишите формулу расчета стандартной ошибки.

20. Дайте определение понятию «интервальная оценка».

21. Запишите формулы расчета средней ошибки прогноза.

22. По какой формуле рассчитывается доверительный интервал для оценки прогноза?

23. Для чего служит коэффициент детерминации? Как он рассчитывается?

24. Для чего применяют F-статистику? Как на ее основе сделать выводы по модели?

25. Для чего применяют t-статистику? Как на ее основе сделать выводы по модели?

26. Записать модели парной нелинейной регрессии: степенной, показательной, гиперболической.

27. В чем заключается линеаризация нелинейных уравнений регрессии? Укажите способы линеаризации.

28. Запишите формулы нахождения коэффициентов нелинейных уравнений регрессии: степенного, показательного, гиперболического.

29. Запишите формулу нахождения индекса корреляции.

30. Для чего применяют индекс корреляции?

31. Каковы свойства индекса корреляции?

32. Как проверяется статистическая значимость индекса корреляции?

33. Для чего используется скорректированный коэффициент детерминации? По какой формуле он рассчитывается?

34. Что характеризуют средние и частные коэффициенты эластичности?

35. Как с помощью коэффициентов эластичности оценить силу влияния независимой переменной на зависимую переменную?

36. Запишите формулы, с помощью которых находятся средние и частные коэффициенты эластичности для линейной, степенной, показательной, гиперболической модели.

4.4. Ответы

Упражнения

1. а) – 3); б) – 7); в) – 5).

2. С1 – 3); С2 – 4); С3 – 2); С4 – 1); С5 – 4); С6 – 3).

3. Объем заявок на недельные ссуды на 94,5% объясняется величиной текущей ставки процента. При увеличении ставки процентов на 1% объем заявок снизится в среднем на 6,811. При этом минимально возможный объем заявок составляет в среднем 152, 399.

4. 3). 5. 3).

6. 1) $\hat{y} = 10^{1,5} \cdot x^{-0,85}$; 2) коэффициент регрессии -0,85 является коэффициентом эластичности показывает, что при возрастании цены на 1% спрос не товар снижается на 0,85%.

7. Между переменными у и х существует тесная обратная линейная регрессионная зависимость. При увеличении х в среднем на 1 единицу среднее значение у уменьшится на 4,5 единиц. При нулевом значении х у составит 8,4.

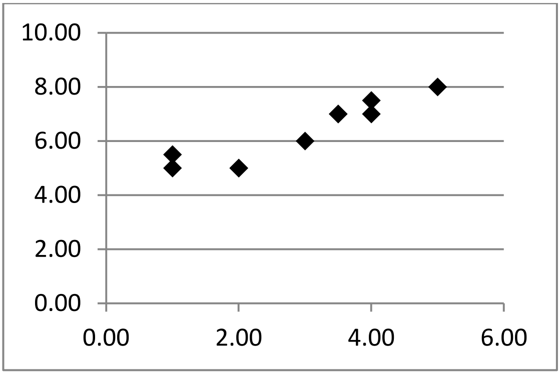
8. При увеличении значений факторного признака х значения результативного признака уменьшаются, причем это уменьшение все время замедляется, и при х → ∞ средняя величина признака у будет равна 10.

Задачи

4.1. $r_{xy} = 0{,}60; m_r = 0{,}28; t_r = 2{,}10; t_{\text{крит}} = 2{,}31$

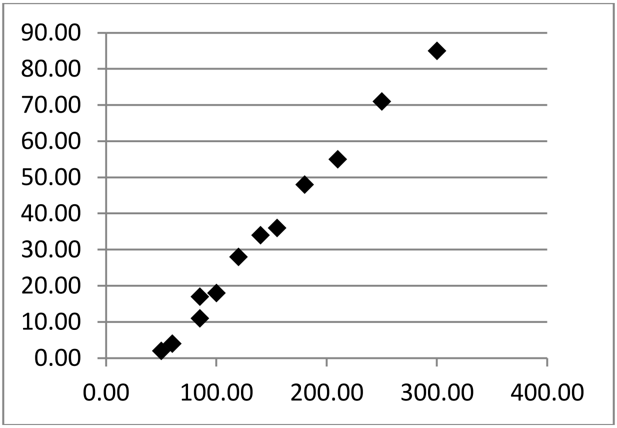
4.2. $r_{xy} = 0{,}93; m_r = 0{,}13;\ t_r = 7{,}05; t_{\text{крит}} = 2{,}31$

ДИ: $(0{,}62; 1{,}23)$



4.3. $r_{xy} = 1{,}00; m_r = 0{,}02;\ t_r = 43{,}38; t_{\text{крит}} = 2{,}23$

ДИ: $(0{,}95; 1{,}05)$



4.4. $r_{xy} = 0{,}68$; $m_r = 0{,}30;\ t_r = 2{,}29; t_{\text{крит}} = 2{,}45; R^2 = 0{,}47; F_{\text{расч}} = 5{,}23; F_{\text{крит}} = 5{,}99.$

4.5. $r_{xy} = 0{,}98; m_r = 0{,}08;\ t_r = 12{,}00; t_{\text{крит}} = 2{,}45;$

$R^2 = 0{,}96; F_{\text{расч}} = 144{,}03; F_{\text{крит}} = 5{,}9; S_e = 1{,}09;$

$\hat{y}_x = 19{,}55 + 5{,}37x; m_{b_0} = 1{,}36; m_{b_1} = 0{,}45;$

$t_{b_0} = 14{,}34; t_{b_1} = 12{,}00; t_{\text{крит}} = 2{,}45;$

$16,21 \leq \beta_0 \leq 22,88; 4,27 \leq \beta_1 \leq 6,46;$

$x_{\text{пр}} = 3,00; \hat{y}_{\text{пр}} = 35,65; m_{\hat{y}_{\text{пр}}} = 1,15; 32,82 \leq y_{\text{пр}} \leq 38,47;$

4.6. $\hat{y}_x = 0,22 \cdot x^{0,99}; \hat{y}_x = 2,17 \cdot 1,04^x; \hat{y}_x = 11,44 - \frac{138,95}{x}$

Вид уравнения регрессии	Индекс детерминации	$F_{\text{расч}}$	$\bar{A}$, (%)	ρ	$\overline{Э}$
степенное	0,9318	122,90	11,42	0,9653	0,99
показательное	0,9067	87,50	13,35	0,9522	0,93
гиперболическое	0,8153	39,73	10,74	0,9030	0,89

Приложения

Приложение 1

Таблица значений z-преобразования Фишера

r	**0**	**1**	**2**	**3**	**4**
0,0	0	0,0101	0,02	0,03	0,04
0,1	0,1003	0,1104	0,1206	0,1308	0,1409
0,2	0,2027	0,2132	0,2237	0,2342	0,2448
0,3	0,3095	0,3205	0,3316	0,3428	0,3541
0,4	0,4236	0,4356	0,4477	0,4599	0,4722
0,5	0,5493	0,5627	0,5764	0,5901	0,6042
0,6	0,6932	0,7089	0,725	0,7414	0,7582
0,7	0,8673	0,8872	0,9077	0,9287	0,9505
0,8	1,0986	1,127	1,1568	1,1881	1,2212
0,9	1,4722	1,5275	1,589	1,6584	1,7381
0,99	2,6466	2,6996	2,7587	2,8257	2,9031

r	**5**	**6**	**7**	**8**	**9**
0,0	0,0501	0,0601	0,0701	0,0802	0,0902
0,1	0,1511	0,1614	0,1717	0,182	0,1923
0,2	0,2554	0,2661	0,2769	0,2877	0,2986
0,3	0,3654	0,3767	0,3884	0,4001	0,4118
0,4	0,4847	0,4973	0,5101	0,523	0,5361
0,5	0,6184	0,6328	0,6475	0,6625	0,6777
0,6	0,7753	0,7928	0,8107	0,8291	0,848
0,7	0,973	0,9962	1,0203	1,0454	1,0714
0,8	1,2562	1,2933	1,3331	1,3758	1,4219
0,9	1,8318	1,9459	2,0923	2,2976	2,6467
0,99	2,9945	3,1063	3,2504	3,4534	3,8002

В качестве примера (для понимания, как получилась таблица) посчитаем значение z для r=0,997 (число из строки 0,99 и столбца 7): z(0,997)=3,2504.

Приложение 2

Критические значения t-критерия Стьюдента

Число степеней свободы d.f.	**Уровень значимости α (двусторонняя критическая область)**			**Число степеней свободы d.f.**	**Уровень значимости α (двусторонняя критическая область)**		
	0,1	**0,05**	**0,01**		**0,1**	**0,05**	**0,01**
1	6,3138	12,706	63,657	**18**	1,7341	2,1009	2,8784
2	2.9200	4,3027	9,9248	**19**	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	**20**	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	**21**	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	**22**	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	**23**	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	**24**	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	**25**	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	**26**	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	**27**	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	**28**	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	**29**	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	**30**	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	**40**	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	**60**	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	**120**	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758
Число степеней свободы d.f.	**0,05**	**0,025**	**0,005**	**Число степеней свободы d.f.**	**0,05**	**0,025**	**0,005**
	Уровень значимости α (односторонняя критическая область)				**Уровень значимости α (односторонняя критическая область)**		

Приложение 3

Критические значения распределения Фишера-Снедекора при уровне значимости α =0,05

k1 / k2	1	2	3	4	5	6	8	12	24	∞
1	161,4	199,5	215,7	224,6	230,2	234	238,9	243,9	249	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62

Продолжение прилож.3

k1 / **k2**	**1**	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**24**	∞
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Приложение 4

Критические значения распределения Пирсона

Число степеней свободы k	Уровень значимости α					
	0,01	**0,025**	**0,05**	**0,95**	**0,975**	**0,99**
1	6,6349	5,0239	3,8415	0,0039	0,0010	0,0002
2	9,2103	7,3778	5,9915	0,1026	0,0506	0,0201
3	11,3449	9,3484	7,8147	0,3519	0,2158	0,1148
4	13,2767	11,1433	9,4877	0,7107	0,4844	0,2971
5	15,0863	12,8325	11,0705	1,1455	0,8312	0,5543
6	16,8119	14,4494	12,5916	1,6354	1,2373	0,8721
7	18,4753	16,0128	14,0671	2,1674	1,6899	1,2390
8	20,0902	17,5346	15,5073	2,7326	2,1797	1,6465
9	21,6660	19,0228	16,9190	3,3251	2,7004	2,0879
10	23,2093	20,4832	18,3070	3,9403	3,2470	2,5582
11	24,7250	21,9201	19,6751	4,5748	3,8158	3,0535
12	26,2170	23,3367	21,0261	5,2260	4,4038	3,5706
13	27,6883	24,7356	22,3620	5,8919	5,0088	4,1069
14	29,1412	26,1190	23,6848	6,5706	5,6287	4,6604
15	30,5779	27,4884	24,9958	7,2609	6,2621	5,2294
16	31,9999	28,8454	26,2962	7,9617	6,9077	5,8122
17	33,4087	30,1910	27,5871	8,6718	7,5642	6,4078
18	34,8053	31,5264	28,8693	9,3905	8,2308	7,0149
19	36,1909	32,8523	30,1435	10,1170	8,9065	7,6327
20	37,5662	34,1696	31,4104	10,8508	9,5908	8,2604
21	38,9322	35,4789	32,6706	11,5913	10,2829	8,8972
22	40,2894	36,7807	33,9244	12,3380	10,9823	9,5425
23	41,6384	38,0756	35,1725	13,0905	11,6886	10,1957
24	42,9798	39,3641	36,4150	13,8484	12,4012	10,8564
25	44,3141	40,6465	37,6525	14,6114	13,1197	11,5240
26	45,6417	41,9232	38,8851	15,3792	13,8439	12,1982
27	46,9629	43,1945	40,1133	16,1514	14,5734	12,8785
28	48,2782	44,4608	41,3371	16,9279	15,3079	13,5647
29	49,5879	45,7223	42,5570	17,7084	16,0471	14,2565
30	50,8922	46,9792	43,7730	18,4927	16,7908	14,9535

Продолжение прилож.4

Число степеней свободы k	**Уровень значимости α**					
	0,01	**0,025**	**0,05**	**0,95**	**0,975**	**0,99**
31	52,1914	48,2319	44,9853	19,2806	17,5387	15,6555
32	53,4858	49,4804	46,1943	20,0719	18,2908	16,3622
33	54,7755	50,7251	47,3999	20,8665	19,0467	17,0735
34	56,0609	51,9660	48,6024	21,6643	19,8063	17,7892
35	57,3421	53,2034	49,8019	22,4650	20,5694	18,5089
36	58,6192	54,4373	50,9985	23,2686	21,3359	19,2327
37	59,8925	55,6680	52,1923	24,0749	22,1056	19,9602
38	61,1621	56,8955	53,3835	24,8839	22,8785	20,6914
39	62,4281	58,1201	54,5722	25,6954	23,6543	21,4262
40	63,6907	59,3417	55,7585	26,5093	24,4330	22,1643
41	64,9501	60,5606	56,9424	27,3256	25,2145	22,9056
42	66,2062	61,7768	58,1240	28,1441	25,9987	23,6501
43	67,4594	62,9904	59,3035	28,9647	26,7854	24,3976
44	68,7095	64,2015	60,4809	29,7875	27,5746	25,1480
45	69,9568	65,4102	61,6562	30,6123	28,3662	25,9013
46	71,2014	66,6165	62,8296	31,4390	29,1601	26,6572
47	72,4433	67,8207	64,0011	32,2676	29,9562	27,4159
48	73,6826	69,0226	65,1708	33,0981	30,7545	28,1770
49	74,9195	70,2224	66,3387	33,9303	31,5549	28,9407
50	76,1539	71,4202	67,5048	34,7643	32,3574	29,7067

Приложение 5

Критические точки распределения Дарбина-Уотсона

n	k=1		k=2		k=3		k=4		k=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0.61	1.40	-	-	-	-	-	-	-	-
7	0.7	1.36	0.47	1.9	-	-	-	-	-	-
8	0.76	1.33	0.56	1.78	0.37	2.29	-	-	-	-
9	0.82	1.32	0.63	1.7	0.46	2.13	-	-	-	-
10	0.88	1.32	0.7	1.64	0.53	2.02	-	-	-	-
11	0.93	1.32	0.66	1.6	0.6	1.93	-	-	-	-
12	0.97	1.33	0.81	1.58	0.66	1.86	-	-	-	-
13	1.01	1.34	0.86	1.56	0.72	1.82	-	-	-	-
14	1.05	1.35	0.91	1.55	0.77	1.78	-	-	-	-
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,1	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,9	1,71	0,78	1,9	0,67	2,1
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,4	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,2	1,41	1,1	1,54	1	1,68	0,9	1,83	0,79	1,99
21	1,2	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,22	1,43	1,15	1,54	1,05	1,66	0,96	1,8	0,86	1,94
23	1,24	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,9	1,92
24	1,26	1,45	1,19	1,55	1,1	1,66	1,01	1,78	0,93	1,9
25	1,27	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,29	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,3	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,32	1,48	1,26	1,56	1,18	1,65	1,1	1,75	1,03	1,85
29	1,33	1,48	1,27	1,56	1,2	1,65	1,12	1,74	1,05	1,84
30	1,34	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,35	1,5	1,3	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,36	1,5	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,37	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,38	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,39	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,8
36	1,4	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,8
37	1,41	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,8

n	k=1		k=2		k=3		k=4		k=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
38	1,42	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,6	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,6	1,34	1,66	1,29	1,72	1,23	1,79
41	1,45	1,55	1,4	1,6	1,35	1,66	1,29	1,72	1,24	1,78
42	1,46	1,55	1,41	1,61	1,36	1,66	1.31	1.72	1,25	1,78
43	1,46	1,56	1,41	1,61	1,37	1,66	1,32	1,72	1,27	1,78
44	1,47	1,56	1,42	1,61	1,37	1,66	1,33	1,72	1,28	1,78
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
46	1,48	1,57	1,44	1,62	1,39	1,67	1,34	1,72	1,3	1,77
47	1,49	1,57	1,44	1,62	1,39	1,67	1,35	1,72	1,31	1,77
48	1,49	1,58	1,45	1,62	1.41	1,67	1,36	1,72	1,32	1,77
49	1,5	1,58	1,46	1,62	1,41	1,67	1,37	1,72	1,32	1,77
50	1,5	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
51	1,51	1,59	1,47	1,63	1,43	1,67	1,38	1,72	1,34	1,77
52	1,51	1,59	1,47	1,63	1,43	1,68	1,39	1,72	1,35	1,77
53	1,52	1,59	1,48	1,63	1,44	1,68	1,4	1,72	1,36	1,77
54	1,52	1,6	1,48	1,64	1,45	1,68	1,41	1,72	1,37	1,77
55	1,53	1,6	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
56	1,53	1,6	1,49	1,64	1,46	1,68	1,42	1,72	1,38	1,77
57	1,54	1,61	1,5	1,64	1,46	1,68	1,43	1,72	1,39	1,77
58	1,54	1,61	1,5	1,65	1,47	1,68	1,43	1,72	1,39	1,77
59	1,54	1,61	1,51	1,65	1,47	1,69	1,44	1,73	1,4	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,5	1,7	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,7	1,49	1,74	1,46	1,77
75	1,6	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,6	1,7	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,7	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,6	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

Литература

1. Эконометрика: Учебник / И.И. Елисеева, С.В. Курышева, Т.В. Костеева и др.; под ред. И.И. Елисеевой. - 2-е изд., перераб. и доп. - М.: Финансы и статистика, 2005.

2. Практикум по эконометрике: Учеб. Пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др.; Под ред. И.И. Елисеевой. - М.: Финансы и статистика, 2003.

3. Бородич С.А. Эконометрика / Учебное пособие для ВУЗов. - Мн.: Новое знание, 2004.

4. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. - М., Дело, 2004..

5. Доугерти К. Введение в эконометрику. / Пер. с англ. - М.: МГУ, ИНФРА-М, 2001.

6. Новак Эдвард. Введение в методы эконометрики. Сборник задач: Пер. с польск. / Под ред. И.И. Елисеевой. - М.: Финансы и статистика, 2004.

7. Эконометрика: практикум / Н.В. Сернова и др.; под ред. Н.В. Серновой. - М.: МГИМО-Университет, 2010.

8. Белько И.В. Эконометрика. Практикум: учеб. пособие / И.В. Белько, Е.А. Криштапович. - Минск: Изд-во Гревцова, 2011.

9. Кочнева Л.Ф. , Милевский А.С. Эконометрика. Часть 1. Парная регрессия. Учебное пособие. М.: МИИТ, 2005. – 32 с.

10. Милевский А.С. Высшая математика. Часть 7. Эконометрика: конспект лекций. М.: МИИТ, 2011. – 123 с.

Св. план 2016 г., поз.

Ишханян Маргарита Владимировна
Карпенко Надежда Викторовна

ЭКОНОМЕТРИКА
ЧАСТЬ I
ПАРНАЯ РЕГРЕССИЯ

Учебное пособие для направления 380301
«Экономика»

Подписано в печать Формат 60 Х 84 / 16

Заказ № Усл. - печ. л. -

Тираж - экз.