МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ имени академика С.П. КОРОЛЕВА»
(Самарский университет)

ЭКОНОМЕТРИКА.
ПАРНАЯ РЕГРЕССИЯ

Рекомендовано редакционно-издательским советом федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» в качестве методических указаний к лабораторным работам для студентов, обучающихся по программам высшего образования

Самара
Издательство Самарского университета
2016

УДК 33(075)
ББК 65в6

Составители: ***А.П. Котенко, О.А. Кузнецова***

Рецензент профессор кафедры организации производства
Самарского университета Д. Ю. Иванов

Эконометрика. Парная регрессия: метод. указания к лабораторным работам / сост. *А.П. Котенко, О.А. Кузнецова*. – Самара: Издательство Самарского университета, 2016. – 48 с.

Методические указания составлены применительно к учебному плану по направлениям «Экономика», «Менеджмент», «Бизнес-информатика». Учтены требования государственного образовательного стандарта высшего профессионального образования по вышеуказанным направлениям и стандарта организации СТО СГАУ 02068410-003–2016.

В методических указаниях приводятся методы решения эконометрических задач.

Предназначены для очной, очно-заочной и вечерней форм обучения.

Подготовлены на кафедре математических методов в экономике.

УДК 33(075)
ББК 65в6

ОБЩИЕ СВЕДЕНИЯ

Эконометрика – одна из базовых дисциплин экономического образования. В экономике в большинстве случаев между переменными величинами существуют зависимости, когда каждому значению одной переменной соответствует не какое-то определенное, а множество возможных значений другой переменной. Иначе говоря, каждому значению одной переменной соответствует определенное (условное) распределение другой переменной. Такая зависимость получила название статистической.

Задачами регрессионного анализа являются установление формы зависимости между переменными, оценка функции регрессии, оценка неизвестных значений (прогноз значений) зависимой переменной.

Переменные могут быть экзогенными (внешними, независимыми, объясняющими) – y, либо эндогенными (внутренними, зависимыми, объясняемыми) – x.

В эконометрике используются пространственные и временные переменные. Пространственные данные характеризуют разные объекты за один и тот же период времени (средняя заработная плата по регионам). Временные данные характеризуют данные по одному и тому же объекту за разные периоды времени (динамика продаж предприятия).

Число наблюдений должно как минимум в 7 раз превышать количество экзогенных переменных.

ПАРНАЯ РЕГРЕССИЯ

ОСНОВНЫЕ ПОНЯТИЯ

Парная регрессия – это уравнение $y = \tilde{f}(x) + \varepsilon$ связи переменных y и x, где y – **зависимая переменная** (**результативный признак**), x – **независимая объясняющая переменная** (**признак-фактор**, *регрессор*), ε – **случайный член регрессии** (*остаток*, *невязка*) в *аддитивной* форме записи, учитывающий случайные (стохастические, недетерминированные) воздействия.

Возможна **мультипликативная форма** записи регрессии $y = \tilde{f}(x) \cdot \varepsilon$, а также другие математические способы учёта стохастических влияний.

Идентификация уравнения регрессии сводится к оценке её параметров. Для *линейных по неизвестным параметрам* регрессий применим метод наименьших квадратов (МНК), позволяющий найти оценки, минимизирующие сумму квадратов отклонений **фактических** (**наблюдаемых**, **экспериментальных**) значений результативного признака y от **теоретических** (**гипотетических**, *регрессионных*, **сглаженных**) значений

$$\tilde{y}: \sum (y_x - \tilde{y}_x)^2 \to \min .$$

Для линейных, а также нелинейных, *приводимых к линейным по идентифицируемым параметрам*, регрессионных моделей решается следующая, линейная по искомым точечным оценкам идентифицируемых параметров, система алгебраических уравнений в нормальной форме:

$$\begin{cases} an + b\sum x = \sum y, \\ a\sum x + b\sum x^2 = \sum xy. \end{cases}$$

Её решения имеют вид $a = \bar{y} - b\bar{x}$, $b = \dfrac{\operatorname{cov}(x, y)}{\sigma_x^2} = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}$.

Интерпретация коэффициента регрессии b – изменение регрессора x на 1 единицу влечёт изменение исследуемого фактора y на b единиц. Т.о. коэффициент b – **экономически интерпретируемый** параметр (это иногда позволяет, к примеру, проверить его экспериментально).

Математическая интерпретация свободного члена регрессии a – значение y при нулевом значении регрессора x = 0. Однако, в большинстве экономических задач регрессор не может принять нулевое или даже близкое к 0 значение. Поэтому свободный член линейной регрессии a относят к **неинтерпретируемым параметрам**, которые почти никогда не удаётся проверить экспериментально.

При интерпретации параметров линейной регрессии отметим 3 факта.

1. Параметры a и b являются лишь выборочными (точечными) оценками параметров линейной регрессии $\tilde{y} = \alpha + \beta x$ генеральной совокупности, поэтому возможно сколь угодно большое их отклонение от «истинных» значений.

2. Уравнение регрессии (не только линейное) отражает лишь **общую тенденцию** выборки, когда каждое отдельное исследование подвержено воздействию сколь угодно больших случайностей.

3. Верность интерпретации зависит от правильности *спецификации* уравнения. «Наилучшая» **МНК-прямая** всегда существует, но не всегда является достаточно хорошей.

Методы наименьших квадратов и максимального правдоподобия

Пусть имеется n совместных наблюдений дискретных случайных величин X и Y. Представим их n точками (x_i, y_i) плоскости $\Re^2$. Спроектируем каждую точку параллельно оси Oy на прямую искомой линейной регрессии $\hat{y} = ax + b$. Получим **регрессионные** (**теоретические**, **сглаженные**, **гипотетические**) значения $\hat{y}_i = ax_i + b$.

Подберём теперь параметры a и b искомой регрессии из условия

$$S(a,b) := \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2 \to \min .$$

Для этого воспользуемся **необходимым условием экстремума** функции двух переменных: $\begin{cases} \partial S/\partial a = 0, \\ \partial S/\partial b = 0. \end{cases}$

Получим $\begin{cases} \partial S/\partial a = 2\sum_{i=1}^{2} (y_i - ax_i - b)(-x_i) = 0, \\ \partial S/\partial b = 2\sum_{i=1}^{2} (y_i - ax_i - b)(-1) = 0. \end{cases}$

Приведём эту системудвух линейных неоднородных уравнений с двумя неизвестными к **нормальному виду**, когда неизвестные находятся в левой, а известные слагаемые в правой части уравнений:

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i, \\ a\sum_{i=1}^{n} x_i + b\cdot n = \sum_{i=1}^{n} y_i. \end{cases}$$

Решить её можно известными методами исключения неизвестных, обратной матрицы или по формулам Крамера. К примеру, по формулам Крамера решение примет вид $a = \Delta_a/\Delta$, $b = \Delta_b/\Delta$, где **определитель системы**

$$\Delta := \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^0 \end{vmatrix} = \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{vmatrix} \neq 0,$$

$$\Delta_a := \begin{vmatrix} \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i & n \end{vmatrix}, \ \Delta_b := \begin{vmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i \end{vmatrix}.$$

Из первого уравнения СЛАУ (1) получим $a = \mathrm{Cov}(X,Y)/\mathrm{Var}(X)$. Тогда из второго уравнения $b = \bar{y} - a\bar{x}$. Последнее равенство означает, что МНК-прямая всегда проходит через **центр** $(\bar{x}, \bar{y})$ диаграммы рассеяния.

Линейная и нелинейная регрессия

Различают **линейные** и **нелинейные** регрессии.

Линейная регрессия: $y = a+bx+\varepsilon$.

Нелинейные регрессии делятся на регрессии, нелинейные относительно регрессоров, и регрессии, нелинейные по оцениваемым параметрам.

Регрессии, *нелинейные по регрессорам*:

полиномы $y = a + b_1x + b_2x^2 + \ldots + b_kx^k + \varepsilon$;

(данная зависимость **линеаризуется** сведением к множественной линейной регрессии путём введения **новых регрессоров** $X_1 := x$, $X_2 := x^2$, …, $X_k := x^k$);

равносторонняя гипербола $y = a + b/x + \varepsilon$;

(данная зависимость линеаризуется заменой $X := 1/x$) и т.п.

Регрессии, *нелинейные по оцениваемым параметрам*:

показательная зависимость $y = ab^x \cdot \varepsilon$;

(данная зависимость линеаризуется заменами

$Y: = \ln y$, $A: = \ln a$, $B: = \ln b$, $E: = \ln\varepsilon$);

степенная зависимость $y = ax^b \cdot \varepsilon$;

(данная зависимость линеаризуется заменами
$Y: = \ln y$, $X: = \ln x$, $A: = \ln a$, $E: = \ln\varepsilon$);

экспоненциальная $y = e^{a+bx} \cdot \varepsilon$;

(данная зависимость линеаризуется заменами $Y: = \ln y$, $E: = \ln\varepsilon$) и т.п.

Регрессии, которые можно свести к линейным, называют **нелинейными внутренне линейными**. К ним применим МНК.

Регрессии, которые не сводятся к линейным, называют **нелинейными внутренне нелинейными**. К ним МНК не применим, т.к. система уравнений для идентификации **МНК-параметров** является нелинейной и требует специальных приёмов нахождения точного или приближённого решения.

Коэффициент и индекс корреляции

Коэффициент парной линейной корреляции

$$r_{xy} := b\frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x, y)}{\sigma_x\sigma_y} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2}\cdot\sqrt{\overline{y^2} - \bar{y}^2}},$$

$$-1 \le r_{xy} \le 1,$$

оценивает тесноту *линейной* связи случайных величин, а **индекс корреляции**

$$\rho_{xy} := \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}} = \sqrt{1 - \frac{\sum(y_x - \tilde{y}_x)^2}{\sum(y_x - \bar{y})^2}},$$

$$0 \le \rho_{xy} \le 1,$$

оценивает тесноту *нелинейной* связи.

Коэффициент и индекс детерминации

Оценку **адекватности** построенной модели экспериментальным наблюдениям даёт коэффициент (индекс) детерминации, а также средняя ошибка аппроксимации.

Долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака y характеризует **коэффициент** (**индекс**) **детерминации**

$$R^2 := \frac{\mathrm{Var}(\tilde{y})}{\mathrm{Var}(y)} = \frac{\sum(\tilde{y}_x - \bar{y})^2}{\sum(y_x - \bar{y})^2}, \text{ или}$$

$$R^2 := 1 - \frac{\mathrm{Var}(\varepsilon)}{\mathrm{Var}(y)} = 1 - \frac{\sum(y_x - \tilde{y}_x)^2}{\sum(y_x - \bar{y})^2}.$$

Коэффициент (индекс) *детерминации* равен квадрату коэффициента (индекса) линейной парной *корреляции*.

Средняя ошибка аппроксимации

$$\bar{A} := \frac{1}{n}\sum\left|\frac{y_x - \tilde{y}_x}{y_x}\right| \cdot 100\%$$

оценивает **среднее отклонение** расчётных значений от фактических. Допустимый предел значений $\bar{A}$ – не более 8-10%.

F-критерий Фишера.
Расчёт F-критерия Фишера парной регрессии

Оценка качества уравнения регрессии состоит в проверке *нулевой гипотезы* H_0 о **статистической незначимости** уравнения регрессии и показателя тесноты связи. Для этого *фактическое* значение статистики $F_{\text{факт}}$ сравнивается с *критическим* (*табличным*) значением $F_{\text{табл}}$ критерия Фишера.

Фактическое значение статистики Фишера равно отношению удельных (рассчитанных на одну степень свободы) факторной и остаточной дисперсий:

$$F_{\text{факт}} := \frac{\sum(\tilde{y}_x - \bar{y})^2 / m}{\sum(y_x - \tilde{y}_x)^2 / (n - m - 1)} = \frac{r_{xy}^2}{1 - r_{xy}^2}(n-2),$$

где n – объём выборки, m – число параметров при регрессорах.

Табличное значение $F_{\text{табл}}$ – это максимально возможное значение F-статистики при действии случайных факторов при заданных уровне значимости α и числе степеней свободы.

F-статистика Фишера имеет 2 числа **степеней свободы**:

– **верхнее число** = числу объясняющих переменных (в парной регрессии =1);

– **нижнее число** = числу наблюдений в выборке минус число оцениваемых параметров (в парной регрессии = n – 2).

$$F_{\text{факт}} := \frac{ESS/p}{RSS/q},$$

где p – верхнее число степеней свободы, q – нижнее число степеней свободы.

Для *парной* регрессии $F_{\text{факт}} := \dfrac{ESS}{RSS} \cdot (n-2)$.

Уровень значимости α – вероятность *отвергнуть верную* гипотезу. Стандартное значение α = 0,05; часто табулируются значения 0,01 и 0,1.

При $F_{\text{табл}} < F_{\text{факт}}$ нулевая гипотеза H_0 о случайном характере оцениваемых характеристик отклоняется и принимается их **статистическая значимость и надёжность**. В противном случае $F_{\text{табл}} > F_{\text{факт}}$ гипотеза H_0 принимается, т.е. признаётся **статистическая незначимость и ненадёжность** найденного уравнения регрессии.

Если $1 \le F < +\infty$, следует рассмотреть F^{-1}.

Использование t-критерия Стьюдента

Для оценки **статистической значимости** коэффициентов регрессии и корреляции рассчитывается t-критерий Стьюдента и доверительные интервалы каждого показателя. *Нулевая гипотеза* H_0 говорит о **случайной природе показателей**, т.е. незначимом их отклонении от 0. С табличным значением сравнивается отношение значения оценки и её случайной ошибки

$$t_a := a/m_a, \quad t_b := b/m_b, \quad t_r := r/m_r,$$

где **случайные ошибки** точечных оценок параметров определяются формулами

$$m_a := \sqrt{\frac{\sum(y_x - \tilde{y}_x)^2}{n-2} \cdot \frac{\sum x^2}{n\sum(x-\bar{x})^2}} = \sqrt{S_{\text{ост}}^2 \frac{\sum x^2}{n^2\sigma_x^2}} = \frac{S_{\text{ост}}}{n\sigma_x}\sqrt{\sum x^2},$$

$$m_b := \sqrt{\frac{\sum(y_x - \tilde{y}_x)^2}{(n-2)\sum(x-\bar{x})^2}} = \sqrt{\frac{S_{\text{ост}}^2}{\sum(x-\bar{x})^2}} = \frac{S_{\text{ост}}}{\sigma_x\sqrt{n}},$$

$$m_{r_{xy}} := \sqrt{\frac{1-r_{xy}^2}{n-2}}.$$

Сравнивая *фактическое* $t_{\text{факт}}$ и *критическое* (табличное) $t_{\text{табл}}$ значения, принимаем или отвергаем гипотезу H_0:

при $t_{\text{табл}} < t_{\text{факт}}$ – отвергаем (т.е. a, b и r_{xy} под действием систематически действующего фактора x отличаются от 0 **не случайно**);

при $t_{\text{табл}} > t_{\text{факт}}$ – принимаем (т.е. a, b и r_{xy} формируются **случайно**).

Статистики Фишера и Стьюдента связаны соотношением $t_b^2 = t_r^2 = \sqrt{F}$.

Предельные ошибки для расчёта доверительных интервалов найдём по формулам $\Delta_a := t_{\text{табл}} m_a$, $\Delta_b := t_{\text{табл}} m_b$.

Для проверки **гипотезы о линейной зависимости** между СВ X и Y справедлива следующая процедура:

1. Вычислить t-статистику $t := r \cdot \sqrt{\dfrac{n-2}{1-r^2}}$.

2. По заданному уровню значимости α найти критическое значение t – статистики с $(n-2)$ степенями свободы.

3. Если величина t превосходит критическое значение в положительную или отрицательную сторону, то нулевая гипотеза $H_0 : r = 0$ отклоняется, т.е. принимается *альтернативна*я гипотеза H_1 о **наличии линейной положительной или отрицательной зависимости** соответственно.

Замечание: тест справедлив лишь для проверки нулевой гипотезы об отсутствии линейной зависимости; для проверки какой-либо другой гипотезы (например, о равенстве коэффициента корреляции заданному ненулевому числу) требуется иная, более сложная процедура.

Для *парного* регрессионного анализа t-критерий для нулевой гипотезы $H_0 : r_{xy} = 0$, F – критерий для гипотезы $H_0 : R^2 = 0$ и t – критерий для $H_0 : \beta = 0$ эквивалентны.

ЛАБОРАТОРНАЯ РАБОТА №1. ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Цель работы: научиться строить модель парной линейной регрессии и оценить её качество.

Исходные данные к работе:

В табл.1 приведена статистика распределения расходов на потребление продуктов питания и средней заработной платы по годам.

Таблица 1. Статистика распределения расходов на потребление продуктов питания и средней заработной платы

№	y	x
1	33,66	120,00
2	28,56	112,30
3	20,40	107,25
4	27,54	107,25
5	48,96	127,25
6	21,42	112,20
7	22,44	113,85
8	43,86	122,10
9	47,94	122,10
10	72,42	132,10
11	79,56	127,05
12	55,08	127,05
13	72,42	128,70
14	77,52	132,00
15	81,60	133,65
16	56,10	128,60
17	77,52	140,25
18	51,00	130,25
19	66,30	130,25
20	64,26	131,90
21	66,30	136,90

Данные для индивидуальных заданий рассчитываются по формуле $y = y + 2 \cdot N$, где «N» обозначен номер варианта работы, соответствующий номеру студента в списке группы.

Парная линейная регрессия

$$y = a + b \bullet x + \varepsilon.$$

Задание: определить коэффициенты парной линейной регрессии методами определителей и наименьших квадратов, оценить качество полученной модели.

Порядок выполнения работы:

1. Составить систему нормальных уравнений и найти параметры регрессии методом определителей

Согласно методу наименьших квадратов

$$\sum(y - y_{расч})^2 \rightarrow \min.$$

Система нормальных уравнений для определения параметров ***a*** и ***b*** линейной регрессии выглядит следующим образом:

$$\begin{cases} na + b\sum x = \sum y \\ a\sum x + b\sum x^2 = \sum xy. \end{cases}$$

где n – количество наблюдений.

Количество наблюдений должно по крайней мере в 7 раз превышать количество переменных в регрессионной модели.

Для подстановки числовых параметров в систему уравнений необходимо заполнить вспомогательную таблицу (табл. 2).

Из системы получаем матрицу

$$\left(\begin{array}{cc|c} n & \sum x & \sum y \\ \sum x & \sum x^2 & \sum xy \end{array}\right)$$

И считаем определители

$$\Delta = n \cdot \sum x^2 - \sum x \cdot \sum x,$$

$$\Delta a = \sum y \cdot \sum x^2 - \sum x \cdot \sum xy,$$

$$\Delta b = n \cdot \sum xy - \sum y \cdot \sum x$$

Δ – главный определитель матрицы

Δa – определитель матрицы a

Δb – определитель матрицы b

$$a = \frac{\Delta a}{\Delta},$$

$$b = \frac{\Delta b}{\Delta}.$$

Таблица 2. Вспомогательная таблица для нахождения коэффициентов парной линейной регрессии

№	y	x	$x \cdot y$	x^2	y^2	$\hat{y}$	ош.аппр.	$(y-\hat{y})^2$	$(x-\bar{x})^2$
1									
…									
21									
Сумма									
Среднее									

2. Найти параметры регрессии методом наименьших квадратов

$$b = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2},$$

$$a = \bar{y} - b \cdot \bar{x}.$$

3. Записать уравнение регрессии с найденными параметрами.

Параметр b называется коэффициентом регрессии. Его величина показывает среднее изменение результата с изменением фактора на одну единицу.

Знак при коэффициенте регрессии b показывает направление связи: при $b > 0$ – связь прямая, а при $b < 0$ – связь обратная.

Параметр a формально показывает значение y при x = 0. Если признак-фактор *x* не имеет и не может иметь нулевого значения, то трактовка свободного члена a не имеет смысла. При отрицательном значении параметр *a* может не иметь экономического содержания.

4. Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Качество построенной модели определяется с помощью показателей корреляции, детерминации, критериев Фишера и Стьюдента.

При использовании линейной регрессии используется линейный коэффициент корреляции r_{xy}. Линейный коэффициент корреляции находится в определенных пределах

$$(-1) <= r_{xy} <= (+1).$$

При этом чем ближе r_{xy} к нулю, тем слабее корреляция, чем ближе r_{xy} к (-1) или к (+1), тем сильнее корреляция, т.е. зависимость x и y близка к линейной

$$r_{xy} = b\frac{\sigma_x}{\sigma_y}$$

где σ_x – среднее квадратическое отклонение фактора,

σ_y – среднее квадратическое отклонение результата.

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 ,$$

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 ,$$

$$\sigma_x = \sqrt{\sigma_x^2},$$

$$\sigma_y = \sqrt{\sigma_y^2}.$$

О наличии существенной линейной связи между переменными y и x_i, можно говорить при значении $|r_{yxi}| > 0{,}5 - 0{,}6$.

Коэффициент детерминации

$$D = r_{xy}^{\ 2} .$$

Критерий Фишера позволяет оценить качество составления всей модели. Расчётное значение критерия должно быть больше табличного. Табличное значение критерия определяется с помощью параметров n и m, где m – число оцениваемых параметров уравнения регрессии (для парной регрессии m = 2), n – число наблюдений. Согласно основной идее дисперсионного анализа для парной регрессии число степеней свободы уравнения регрессии $k_1 = m - 1$, а число степеней свободы остаточной дисперсии $k_2 = n - m$.

$$F_{факт} = \frac{r_{xy}^2}{1 - r_{xy}^2}(n-2).$$

Критерий Стьюдента позволяет оценить качество параметров уравнения.

Оценка значимости коэффициентов регрессии и корреляции с помощью t-критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_a = \frac{a}{m_a}, \quad t_b = \frac{b}{m_b}, \quad t_r = \frac{r}{m_r}.$$

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}},$$

$$m_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}},$$

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

Расчётное значение t – статистики должно быть больше критического (табличного).

Средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\%.$$

Допустимый предел значений* A *– не более 8–10 %.

5. Сформулировать выводы по работе.

ЛАБОРАТОРНАЯ РАБОТА №2.
ПАРНАЯ СТЕПЕННАЯ РЕГРЕССИЯ

Цель работы: построение модели парной степенной регрессии и оценка её качества.

Исходные данные к работе: в табл.3 приведены исходные данные к работе.

Таблица 3. Исходные данные к работе

№	*x*	*y*
1	120,00	33,66
2	112,30	28,56
3	107,25	20,40
4	107,25	27,54
5	127,25	48,96
6	112,20	21,42
7	113,85	22,44
8	122,10	43,86
9	122,10	47,94
10	132,10	72,42
11	137,05	79,56
12	127,05	55,08
13	128,70	72,42
14	142,00	77,52
15	133,65	81,60
16	128,60	56,10
17	140,25	77,52
18	130,25	51,00
19	130,25	66,30
20	131,90	64,26
21	136,90	66,30

Данные для индивидуальных заданий рассчитываются по формуле $y = y + 2 * N$, $x = x + 5 \cdot N$, где «N» обозначен номер варианта работы, соответствующий номеру студента в списке группы.

Парная степенная регрессия

$y=ax^b\varepsilon$

Задание: определить коэффициенты парной линейной регрессии методами определителей и наименьших квадратов, оценить качество полученной модели.

Порядок выполнения работы:

1. Привести уравнение регрессии к линейному виду с помощью логарифмирования

$$Ln\, y = ln\, a + b\cdot ln\, x + ln\, \varepsilon,$$

далее уравнение можно решить как линейное.

2. Составить систему нормальных уравнений и найти методом определителей параметры регрессии

Метод наименьших квадратов

$$\sum(y\text{-}y_{расч})^2 \to \min.$$

Система нормальных уравнений для определения параметров ***a*** и ***b*** линейной регрессии выглядит следующим образом:

$$\begin{cases} n\cdot lna + b\cdot\sum lnx = \sum lny \\ a\cdot\sum lnx + b\cdot\sum lnx^2 = \sum lnx\cdot lny. \end{cases}$$

где n – количество наблюдений.

Количество наблюдений должно по крайней мере в 7 раз превышать количество переменных в регрессионной модели.

Для подстановки числовых параметров в систему уравнений необходимо заполнить вспомогательную таблицу (табл. 4)

Из системы получаем матрицу

$$\left[\begin{array}{cc|c} n & \sum lnx & \sum lny \\ \sum lnx & \sum lnx^2 & \sum lnx\cdot lny \end{array}\right]$$

И находим значения определителей

$$\Delta = n\cdot\sum lnx^2 - \sum lnx\cdot\sum lnx,$$

$$\Delta a = \sum lny\cdot\sum lnx^2 - \sum lnx\cdot\sum lnx\cdot lny,$$

$$\Delta b = n\cdot\sum lnx\cdot lny - \sum lny\cdot\sum x,$$

Таблица 4. Вспомогательная таблица для нахождения коэффициентов парной степенной регрессии

№	y	x	Lny	lnx	$Lnx \cdot lny$	lnx^2	lny^2
1							
…							
21							
Сумма							
Среднее							

$$lna = \frac{\Delta lna}{\Delta},$$

$$b = \frac{\Delta b}{\Delta},$$

где Δ – главный определитель матрицы,

Δlna – определитель матрицы a,

Δlnb – определитель матрицы b.

Чтобы найти значение параметра a, необходимо провести процедуру потенциирования.

$$a = e^{\ln a}.$$

3. Определить параметры парной регрессии в соответствии с методом наименьших квадратов

$$b = \frac{\overline{lnx \cdot lny} - \overline{lnx} \cdot \overline{lny}}{\overline{lnx^2} - \overline{lnx}^2},$$

$$lna = \overline{lny} - b \cdot \overline{lnx}.$$

Записать уравнение регрессии с найденными параметрами.

Параметр b показывает на сколько % изменится результат при изменении фактора на один %.

4. Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Качество построенной модели определяется с помощью показателей корреляции, детерминации, критериев Фишера и Стьюдента.

При использовании линейной регрессии используется линейный коэффициент корреляции r_{xy}. Линейный коэффициент корреляции находится в пределах

$$(-1) <= r_{xy} <= (+1).$$

При этом чем ближе r_{xy} к нулю, тем слабее корреляция, чем ближе r_{xy} к (-1) или к (+1), тем сильнее корреляция, т.е. зависимость x и y близка к линейной

$$r_{xy} = b\frac{\sigma_{lnx}}{\sigma_{lny}},$$

где σ_x – среднее квадратическое отклонение фактора,

σ_y – среднее квадратическое отклонение результата

$$\sigma^2_{lnx} = \overline{lnx^2} - \overline{lnx}^2,$$

$$\sigma^2_{lny} = \overline{lny^2} - \overline{lny}^2,$$

$$\sigma_{lnx} = \sqrt{\sigma^2_{lnx}},$$

$$\sigma_{lny} = \sqrt{\sigma^2_{lny}}.$$

О наличии существенной линейной связи между переменными y и x_i, можно говорить при значении $|r_{yxi}| > 0,5 - 0,6$.

Коэффициент детерминации

$$D = r_{xy}{}^2.$$

Критерий Фишера позволяет оценить качество составления всей модели. Расчётное значение критерия должно быть больше табличного. Табличное значение критерия определяется с помощью параметров n и m, где m – число оцениваемых параметров уравнения регрессии (для парной регрессии $m = 2$), n – число наблюдений. Согласно основной идее дисперсионного анализа для парной регрессии число степеней свободы уравнения регрессии $k_1 = m - 1$, а число степеней свободы остаточной дисперсии $k_2 = n - m$

$$F_{факт} = \frac{r^2_{xy}}{1 - r^2_{xy}}(n-2).$$

Критерий Стьюдента позволяет оценить качество параметров уравнения. Для его нахождения составляется вспомогательная таблица (табл. 5).

Таблица 5. Вспомогательная таблица для нахождения критерия Стьюдента

№	y	x	$\widehat{lny}$	ош. аппр.	$(lnx-\overline{lnx})^2$	$(lny-\widehat{lny})^2$
1						
…						
21						
Сумма						
Среднее						

$\widehat{lny}$ – расчётное значение функции

$$Ln\ \hat{y} = ln\,a + b \cdot ln\,x,$$

где lna, lnb – ранее найденные параметры регрессии.

Оценка значимости коэффициентов регрессии и корреляции с помощью t – критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_a = \frac{a}{m_a};\ t_b = \frac{b}{m_b};\ t_r = \frac{r}{m_r}.$$

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum(\ln y - \ln \hat{y}_x)^2 / (n-2)}{\sum(\overline{\ln x} - \ln x)^2}};$$

$$m_a = \sqrt{\frac{\sum(\ln y - \ln \hat{y}_x)^2}{(n-2)} \cdot \frac{\sum \ln x^2}{n\sum(\overline{\ln x - \ln x})^2}};$$

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

Расчётное значения t – статистики должно быть больше критического (табличного).

Средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических:

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y-\hat{y}}{y}\right| \cdot 100\%.$$

Допустимый предел значений A *– не более 8-10 %.*

5. Сформулировать выводы по работе.

ЛАБОРАТОРНАЯ РАБОТА №3. ПАРНАЯ ПОКАЗАТЕЛЬНАЯ И ПАРНАЯ ЭКСПОНЕНЦИАЛЬНАЯ РЕГРЕССИИ

Цель работы: построение моделей парной показательной регрессии и парной экспоненциальной регрессии и оценка их качества.

Исходные данные к работе: данные для ЛР 2.

Данные для индивидуальных заданий рассчитываются по формуле $y = y + 2 * N$, $x = x + 5 \cdot N$, где «N» обозначен номер варианта работы, соответствующий номеру студента в списке группы.

Парная показательная регрессия

$$\boldsymbol{y=ab^x\varepsilon}$$

Задание: определить коэффициенты парной показательной регрессии методами определителей и наименьших квадратов, оценить качество полученной модели.

Порядок выполнения работы:

1. Привести уравнение регрессии к линейному виду с помощью логарифмирования

$$Ln\, y = ln\, a + x \cdot ln\, b + ln\, \varepsilon,$$

далее уравнение можно решить как линейное.

2. Составить систему нормальных уравнений и найти методом определителей параметры регрессии.

Метод наименьших квадратов

$$\sum(y - y_{расч})^2 \rightarrow \min.$$

Система нормальных уравнений для определения параметров ***a*** и ***b*** линейной регрессии выглядит следующим образом:

$$\begin{cases} n \cdot lna + lnb \cdot \sum x = \sum lny \\ a\sum x + lnb\sum x^2 = \sum xlny \end{cases}$$

где n – количество наблюдений.

Количество наблюдений должно по крайней мере в 7 раз превышать количество переменных в регрессионной модели.

Для подстановки числовых параметров в систему уравнений необходимо заполнить вспомогательную таблицу (табл. 6).

Из системы получаем матрицу

$$\left[\begin{array}{cc|c} n & \sum x & \sum lny \\ \sum x & \sum x^2 & \sum xlny \end{array}\right]$$

Таблица 6. Вспомогательная таблица для нахождения коэффициентов парной линейной регрессии

№	y	x	lny	$x*lny$	x^2	lny^2	
1							
…							
21							
Сумма							
Среднее							

И считаем определители

$$\Delta = n \cdot \sum x^2 - \sum x \cdot \sum x,$$

$$\Delta a = \sum lny \cdot \sum x^2 - \sum x \cdot \sum xlny,$$

$$\Delta b = n \cdot \sum xlny - \sum lny \cdot \sum x,$$

$$lna = \frac{\Delta lna}{\Delta},$$

$$lnb = \frac{\Delta lnb}{\Delta},$$

где Δ – главный определитель матрицы,

Δlna – определитель матрицы *a,*

Δlnb – определитель матрицы *b.*

Чтобы найти значения параметров, необходимо провести процедуру потенциирования

$$a = e^{\ln a},$$

$$b = e^{\ln b}.$$

3. Найти параметры регрессии методом наименьших квадратов

$$lnb = \frac{\overline{x \cdot lny} - \bar{x} \cdot \overline{lny}}{\overline{x^2} - \bar{x}^2},$$

$$lna = \overline{lny} - lnb \cdot \bar{x},$$

$$a = e^{\ln a},$$

$$b = e^{\ln b}.$$

Записать уравнение регрессии с найденными параметрами.

4. Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Качество построенной модели определяется с помощью показателей корреляции, детерминации, критериев Фишера и Стьюдента.

При использовании линейной регрессии используется линейный коэффициент корреляции r_{xy}. Линейный коэффициент корреляции находится в определенных пределах

$$(-1) <= r_{xy} <= (+1).$$

При этом чем ближе r_{xy} к нулю, тем слабее корреляция, чем ближе r_{xy} к (-1) или к (+1), тем сильнее корреляция, т.е. зависимость x и y близка к линейной

$$r_{x\ln y} = \ln b \frac{\sigma_x}{\sigma_{\ln y}}$$

где σ_x – среднее квадратическое отклонение фактора,

σ_y – среднее квадратическое отклонение результата.

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2,$$

$$\sigma_{lny}^2 = \overline{lny^2} - \overline{lny}^2,$$

$$\sigma_x = \sqrt{\sigma_x^2},$$

$$\sigma_{lny} = \sqrt{\sigma_{lny}^2}.$$

О наличии существенной линейной связи между переменными y и x_i, можно говорить при значении $|r_{yxi}| > 0,5 - 0,6$.

Коэффициент детерминации

$$D = r_{x\ln y}^{\ 2}.$$

Критерий Фишера позволяет оценить качество составления всей модели. Расчётное значение критерия должно быть больше табличного. Табличное значение критерия определяется с помощью параметров n и m, где m – число оцениваемых параметров уравнения регрессии (для парной регрессии $m = 2$), n – число наблюдений.

Согласно основной идее дисперсионного анализа для парной регрессии число степеней свободы уравнения регрессии $k_1 = m - 1$, а число степеней свободы остаточной дисперсии $k_2 = n - m$

$$F_{факт} = \frac{r_{x\ln y}^2}{1 - r_{x\ln y}^2}(n-2).$$

Критерий Стьюдента позволяет оценить качество параметров уравнения.

Таблица 7. Вспомогательная таблица
для нахождения критерия Стьюдента

№	y	x	$\widehat{lny}$	ош.аппр.	$(x-\bar{x})^2$	$(lny-\widehat{lny})^2$
1						
…						
21						
Сумма						
Среднее						

Оценка значимости коэффициентов регрессии и корреляции с помощью t-критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_a = \frac{\ln a}{m_a}; \quad t_b = \frac{\ln b}{m_b}; \quad t_r = \frac{r}{m_r}.$$

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum(\ln y - \ln \hat{y}_x)^2 / (n-2)}{\sum(x-\bar{x})^2}};$$

$$m_a = \sqrt{\frac{\sum(\ln y - \ln \hat{y}_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n\sum(x-\bar{x})^2}};$$

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

Расчётное значения t – статистики должно быть больше критического (табличного).

Средняя ошибка аппроксимации - среднее отклонение расчетных значений от фактических:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{\ln y - \ln \hat{y}}{\ln y}\right| \cdot 100\%.$$

Допустимый предел значений А – не более 8–10 %.

5. Сформулировать выводы по работе.

Экспоненциальная модель

$$y = e^{a+bx+e}$$

Задание: определить коэффициенты экспоненциальной регрессии, оценить качество полученной модели.

Порядок выполнения работы:

1. Привести уравнение регрессии к линейному виду с помощью логарифмирования

$$Ln\, y = a + bx + \varepsilon,$$

далее уравнение можно решить как линейное.

2. Составить систему нормальных уравнений и найти методом определителей параметры регрессии

Метод наименьших квадратов

$$\sum(y - y_{расч})^2 \to \min.$$

Система нормальных уравнений для определения параметров ***a*** и ***b*** линейной регрессии выглядит следующим образом:

$$\begin{cases} n \cdot a + b \cdot \sum x = \sum \ln \cdot y \\ a\sum x + b\sum x^2 = \sum x \cdot \ln y \end{cases}$$

где n – количество наблюдений.

Количество наблюдений должно по крайней мере в 7 раз превышать количество переменных в регрессионной модели.

Для подстановки числовых параметров в систему уравнений необходимо заполнить вспомогательную таблицу (табл. 8).

Таблица 8. Вспомогательная таблица для нахождения коэффициентов регрессии

№	y	x	lny	$x \cdot lny$	x^2	lny^2
1						
…						
21						
Сумма						
Среднее						

Из системы получаем матрицу

$$\left[\begin{array}{cc|c} n & \sum x & \sum lny \\ \sum x & \sum x^2 & \sum xlny \end{array}\right]$$

И считаем определители

$$\Delta = n \cdot \sum x^2 - \sum x \cdot \sum x,$$
$$\Delta a = \sum lny \cdot \sum x^2 - \sum x \cdot \sum x \cdot lny,$$
$$\Delta b = n \cdot \sum x \cdot lny - \sum lny \cdot \sum x.$$

Δ – главный определитель матрицы,
Δa – определитель матрицы a,
Δb – определитель матрицы b.

$$a = \frac{\Delta a}{\Delta},$$
$$b = \frac{\Delta b}{\Delta}.$$

3. Найти параметры регрессии методом наименьших квадратов

$$b = \frac{\overline{x \cdot lny} - \bar{x} \cdot \overline{lny}}{\overline{x^2} - \bar{x}^2},$$

$$a = \overline{lny} - lnb \cdot \bar{x}.$$

Записать уравнение регрессии с найденными параметрами.

4. Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Качество построенной модели определяется с помощью показателей корреляции, детерминации, критериев Фишера и Стьюдента.

При использовании линейной регрессии используется линейный коэффициент корреляции r_{xy}. Линейный коэффициент корреляции находится в определенных пределах

$$(-1) <= r_{xy} <= (+1).$$

При этом чем ближе r_{xy} к нулю, тем слабее корреляция, чем ближе r_{xy} к (-1) или к (+1), тем сильнее корреляция, т.е. зависимость x и y близка к линейной

$$r_{xy} = b\frac{\sigma_x}{\sigma_y},$$

где σ_x – среднее квадратическое отклонение фактора,

σ_y – среднее квадратическое отклонение результата.

$$\sigma_x^2 = \overline{x^2} - \overline{x}^2,$$

$$\sigma_y^2 = \overline{y^2} - \overline{y}^2,$$

$$\sigma_x = \sqrt{\sigma_x^2},$$

$$\sigma_y = \sqrt{\sigma_y^2}.$$

О наличии существенной линейной связи между переменными y и x_i, можно говорить при значении $|r_{yxi}| > 0{,}5 - 0{,}6$.

Коэффициент детерминации

$$D = r_{xy}^{\ 2}.$$

Критерий Фишера позволяет оценить качество составления всей модели. Расчётное значение критерия должно быть больше табличного. Табличное значение критерия определяется с помощью параметров n и m. где m- число оцениваемых параметров уравнения регрессии (для парной регрессии $m=2$), n – число наблюдений. Согласно основной идее дисперсионного анализа для парной регрессии число степеней свободы уравнения регрессии $k_1 = m - 1$, а число степеней свободы остаточной дисперсии $k_2 = n - m$

$$F_{факт} = \frac{r_{xy}^2}{1 - r_{xy}^2}(n-2).$$

Критерий Стьюдента позволяет оценить качество параметров уравнения.

Оценка значимости коэффициентов регрессии и корреляции с помощью t-критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_a = \frac{a}{m_a}; \quad t_b = \frac{b}{m_b}; \quad t_r = \frac{r}{m_r}.$$

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum(\ln y - \ln \hat{y}_x)^2 / (n-2)}{\sum(x - \bar{x})^2}};$$

$$m_a = \sqrt{\frac{\sum(\ln y - \ln \hat{y}_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n\sum(x-\bar{x})^2}};$$

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

Таблица 9. Вспомогательная таблица для нахождения ошибок параметров функции

№	y	x	$ln\ y_{расч}$	ош.аппр.	$(ln\ y - ln\ y_{расч})^2$	$(x - x_{ср})^2$
1						
…						
21						
Сумма						
Среднее						

$$\text{Ln}\ \hat{y} = a + b\,x$$

где a, b – ранее найденные параметры регрессии.

Расчётное значения t-статистики должно быть больше критического (табличного).

Средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{\ln y - \ln \hat{y}}{\ln y}\right| \cdot 100\%.$$

Допустимый предел значений A – не более 8–10 %.

5. Сформулировать выводы по работе.

ЛАБОРАТОРНАЯ РАБОТА №4.
МОДЕЛЬ ФИЛИПСА И МОДЕЛЬ ЭНГЕЛЯ

Цель работы: построение моделей Филипса и Энгеля и оценка их качества.

Исходные данные к работе: исходные данные приведены в табл. 10 и 11.

Данные для индивидуальных заданий рассчитываются по формуле $y = y + 2 * N$, $x = x + 5 * N$, где «N» обозначен номер варианта работы, соответствующий номеру студента в списке группы.

Таблица 10. Исходные данные для построения кривой Филипса

№	y	x
1	116,51	5,4
2	113,22	5,6
3	95,625	6,8
4	83,934	8,3
5	87,552	7,6
6	76,666	9,2
7	439,31	1,2
8	237,86	2,3
9	143,91	4,1
10	108,41	5,7
11	80,726	8,4
12	82,392	8,2
13	76,232	9,3
14	64,954	12
15	41,981	25
16	46,749	20
17	43,886	23
18	49,514	17

19	68,639	11
20	65,501	12
21	48,137	18

Задание: определить коэффициенты парной линейной регрессии методами определителей и наименьших квадратов, оценить качество полученной модели.

Кривая Филипса характеризует нелинейное соотношение между нормой безработицы x и процентом прироста заработной платы y

$$y = a + \frac{b}{x} + e$$

x – норма безработицы, %

y – прирост заработной платы, %

Порядок выполнения работы:

1. Составить систему нормальных уравнений и найти методом определителей параметры регрессии

$$y = a + b\frac{1}{x} + e$$

Для удобства проведём замену 1/x на z

$$y = a + z + e$$

Согласно методу наименьших квадратов неизвестные параметры ***a*** и ***b*** выбираются таким образом, чтобы сумма квадратов отклонений эмпирических значений зависимой переменной ***y*** от значений, найденных по уравнению регрессии, была минимальной.

Метод наименьших квадратов

$$\sum(y-y_{расч})^2 \to \min.$$

Система нормальных уравнений для определения параметров ***a*** и ***b*** линейной регрессии выглядит следующим образом:

$$\begin{cases} n\cdot a+b\sum z = \sum y \\ a\sum z+b\sum z^2=\sum zy, \end{cases}$$

где n – количество наблюдений.

Для подстановки числовых параметров в систему уравнений необходимо заполнить вспомогательную таблицу:

Таблица 11. Вспомогательная таблица для нахождения коэффициентов регрессии

№	y	x	z	$z*y$	z^2	y^2
1						
…						
21						
Сумма						
Среднее						

Из системы получаем матрицу

$$\left(\begin{array}{cc|c} n & \sum z & \sum y \\ \sum z & \sum z^2 & \sum zy \end{array}\right)$$

И считаем определители

$$\Delta = n \cdot \sum z^2 - \sum z \cdot \sum z,$$

$$\Delta a = \sum y \cdot \sum z^2 - \sum z \cdot \sum zy,$$

$$\Delta b = n \cdot \sum zy - \sum y \cdot \sum z,$$

Δ – главный определитель матрицы,

Δa – определитель матрицы a,

Δb – определитель матрицы b

$$a = \frac{\Delta a}{\Delta},$$

$$b = \frac{\Delta b}{\Delta}.$$

Записать уравнение регрессии с найденными параметрами.

2. Определить параметры функции в соответствии с методом наименьших квадратов

$$b = \frac{\overline{yz} - \bar{y} \cdot \bar{z}}{\overline{z^2} - \bar{z}^2}, \qquad a = \bar{y} - b \cdot \bar{z}.$$

3. Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Качество построенной модели определяется с помощью показателей корреляции, детерминации, критериев Фишера и Стьюдента.

При использовании линейной регрессии используется линейный коэффициент корреляции r_{xy}. Линейный коэффициент корреляции находится в определенных пределах: (-1)<= r_{xy} <=(+1). При этом чем ближе r_{xy} к нулю, тем слабее корреляция, чем ближе r_{xy} к (-1) или к (+1), тем сильнее корреляция, т.е. зависимость *x* и *y* близка к линейной.

$$r_{xy} = b\frac{\sigma_x}{\sigma_y}$$

σ_x – среднее квадратическое отклонение фактора,
σ_y – среднее квадратическое отклонение результата

$$\overline{\sigma_x^2} = z^2 - \bar{z}^2,$$

$$\overline{\sigma_y^2} = y^2 - \bar{y}^2,$$

$$\sigma_x = \sqrt{\sigma_x^2},$$

$$\sigma_y = \sqrt{\sigma_y^2}.$$

О наличии существенной линейной связи между переменными y и x_i, можно говорить при значении $|r_{yxi}| > 0{,}5$ - - 0,6.

Коэффициент детерминации

$$D = r_{xy}^{\ 2}.$$

Критерий Фишера позволяет оценить качество составления всей модели. Расчётное значение критерия должно быть больше табличного. Табличное значение критерия определяется с помощью параметров *n* и *m*, где *m* – число оцениваемых параметров уравнения регрессии (для парной регрессии $m = 2$), n – число наблюдений. Согласно основной идее дисперсионного анализа для парной регрессии число степеней свободы уравнения регрессии $k_1 = m - 1$, а число степеней свободы остаточной дисперсии $k_2 = n - m$.

$$F_{факт} = \frac{r_{xy}^2}{1 - r_{xy}^2}(n-2),$$

Критерий Стьюдента позволяет оценить качество параметров уравнения.

Оценка значимости коэффициентов регрессии и корреляции с помощью t – критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_a = \frac{a}{m_a}; \quad t_b = \frac{b}{m_b}; \quad t_r = \frac{r}{m_r}.$$

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum(y - \hat{y}_x)^2 / (n-2)}{\sum(z - \bar{z})^2}};$$

$$m_a = \sqrt{\frac{\sum(y - \hat{y}_x)^2}{(n-2)} \cdot \frac{\sum z^2}{n\sum(z - \bar{z})^2}};$$

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

$$\widehat{y} = a + b\frac{1}{x} + e.$$

Параметры a и b были определены ранее.

Расчётное значения t-статистики должно быть больше критического (табличного).

Средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y - \hat{y}}{y}\right| \cdot 100\%.$$

Допустимый предел значений A – не более 8–10 %.

4. Сформулировать выводы по работе.

Кривая Энгеля

Таблица 12. Исходные данные для построения кривой Энгеля

№	*y*	*x*
1	19,46507	125
2	19,76849	117,3
3	17,99983	112,25
4	18,91755	112,25
5	21,81207	132,25
6	18,37898	117,2
7	18,26849	118,85
8	21,87516	127,1
9	20,61574	127,1
10	20,88457	137,1
11	19,30039	142,05
12	18,9174	132,05
13	20,13511	133,7
14	22,09058	147
15	20,28172	138,65
16	21,28088	133,6
17	21,77556	145,25
18	21,63858	135,25
19	19,52915	135,25
20	20,92817	136,9
21	19,47454	141,9

Кривая Энгеля характеризует взаимосвязь общей суммы доходов и доли расходов на продукты питания

$$y = a - \frac{b}{x} + e,$$

X – норма безработицы, %

Y – прирост заработной платы, %.

Порядок выполнения работы:

1. Составить систему нормальных уравнений и найти методом определителей параметры регрессии

$$y = a + (-b)\frac{1}{x} + e.$$

Для удобства проведём замену 1/x на z

$$y = a + (-b)z + e.$$

Согласно методу наименьших квадратов неизвестные параметры ***a*** и ***b*** выбираются таким образом, чтобы сумма квадратов отклонений эмпирических значений зависимой переменной ***y*** от значений, найденных по уравнению регрессии, была минимальной.

Метод наименьших квадратов

$$\sum(y-y_{расч})^2 \rightarrow \min$$

Система нормальных уравнений для определения параметров ***a*** и ***b*** линейной регрессии выглядит следующим образом:

$$\begin{cases} n\cdot a+(-b)\cdot\sum z = \sum y \\ a\cdot\sum z+(\_b)\cdot\sum z^2=\sum(z\cdot y) \end{cases}$$

где n – количество наблюдений.

Количество наблюдений должно по крайней мере в 7 раз превышать количество переменных в регрессионной модели.

Для подстановки числовых параметров в систему уравнений необходимо заполнить вспомогательную таблицу:

Таблица 13. Вспомогательная таблица для нахождения коэффициентов регрессии

№	*y*	*x*	*z*	*z*y*	z^2	y^2
1						
…						
21						
Сумма						
Среднее						

Из системы получаем матрицу

$$\left[\begin{array}{cc|c} n & \sum z & \sum y \\ \sum z & \sum z^2 & \sum zy \end{array}\right]$$

И считаем определители

$$\Delta = n \cdot\sum z^2 - \sum z\cdot\sum z,$$

$$\Delta a = \sum y\cdot\sum z^2 - \sum z\cdot\sum zy,$$

$$\Delta\text{-}b = n\cdot\sum zy - \sum y\cdot\sum z$$

Δ – главный определитель матрицы,

Δa – определитель матрицы a,

Δb – определитель матрицы b

$$a = \frac{\Delta a}{\Delta},$$

$$-b = \frac{\Delta - b}{\Delta},$$

$$b = -b.$$

Записываем уравнение регрессии с найденными параметрами.

2. Определить параметры функции в соответствии с методом наименьших квадратов

$$\overline{-b} = \frac{yz - \bar{y}\cdot\bar{z}}{z^2 - \bar{z}^2}$$

$$a = \bar{y} - b\cdot\bar{z},$$

$$b = -b$$

3. Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Качество построенной модели определяется с помощью показателей корреляции, детерминации, критериев Фишера и Стьюдента.

При использовании линейной регрессии используется линейный коэффициент корреляции r_{xy}. Линейный коэффициент корреляции находится в определенных пределах: (-1) < = r_{xy} < = = (+1). При этом чем ближе r_{xy} к нулю, тем слабее корреляция, чем ближе r_{xy} к (-1) или к (+1), тем сильнее корреляция, т.е. зависимость *x* и *y* близка к линейной.

$$r_{xy} = -b\frac{\sigma_x}{\sigma_y}$$

σ_x – среднее квадратическое отклонение фактора,
σ_y – среднее квадратическое отклонение результата

$$\overline{\sigma}_x^2 = z^2 \ - \bar{z}^2$$

$$\overline{\sigma}_y^2 = y^2 \ - \bar{y}^2$$

$$\sigma_x = \sqrt{\sigma_x^2}$$

$$\sigma_y = \sqrt{\sigma_y^2}$$

О наличии существенной линейной связи между переменными y и x_i, можно говорить при значении $|r_{yxi}| > 0{,}5 - \ - 0{,}6$.

Коэффициент детерминации.

$$D = r_{xy}{}^2 .$$

Критерий Фишера позволяет оценить качество составления всей модели. Расчётное значение критерия должно быть больше табличного. Табличное значение критерия определяется с помощью параметров n и m, где m – число оцениваемых параметров уравнения регрессии (для парной регрессии $m = 2$), n – число наблюдений. Согласно основной идее дисперсионного анализа для парной регрессии число степеней свободы уравнения регрессии $k_1 = m - 1$, а число степеней свободы остаточной дисперсии $k_2 = n - m$.

$$F_{факт} = \frac{r_{xy}^2}{1 - r_{xy}^2}(n-2).$$

Критерий Стьюдента позволяет оценить качество параметров уравнения.

Оценка значимости коэффициентов регрессии и корреляции с помощью t-критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_a = \frac{a}{m_a}; \quad t_b = \frac{-b}{m_b}; \quad t_r = \frac{r}{m_r}.$$

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum\left(y-\hat{y}_x\right)^2 /(n-2)}{\sum\left(z-\bar{z}\right)^2}};$$

$$m_a = \sqrt{\frac{\sum\left(y-\hat{y}_x\right)^2}{(n-2)} \cdot \frac{\sum z^2}{n\sum\left(z-\bar{z}\right)^2}};$$

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

$$\hat{y} = a + (-b)\frac{1}{x} + e \cdot$$

Параметры a и b были определены ранее.

Расчётное значения *t*-статистики должно быть больше критического (табличного).

Средняя ошибка аппроксимации - среднее отклонение расчетных значений от фактических:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y-\hat{y}}{y}\right| \cdot 100\%.$$

Допустимый предел значений А – не более 8–10%.

4. Сформулировать выводы по работе.

СПИСОК ЛИТЕРАТУРЫ

1. Айвазян С.А. Эконометрика. Краткий курс. М: Маркет Дс, 2010, 104 с.

2. Айвазян С.А. Эконометрика-2. Продвинутый курс с приложениями в финансах: учебник. Магистр, 201, 944 с.

3. Бородич С.А. Эконометрика: практикум. М: ИНФРА-М, 2014. 329 с.

4. Буравлёв А. Эконометрика: учеб. пособие. М: Бином, 2012, 166 с.

5. Герасимов А.Н., Гладилин А.В. Эконометрика. Теория и практика. Кнорус, 2011, CD.

6. Гладилин А.В., Герасимов А.Н., Громов Е.И. Эконометрика. М: Феникс, 2011. 304 с.

7. Горлач Б.А. Теория вероятностей и математическая статистика: учеб. пособие. СПб.: Издательство «Лань», 2013. – 320 с.

8. Кремер Н.Ш., Путко Б.А.. Эконометрика / под ред. Н.Ш. Кремера. М.:ЮНИТИ, 2010. 328 с.

9. Клентак Л.С. Элементы теории вероятностей и математической статистики: учеб. пособие. Самара: Изд-во Самар. гос. аэрокосм. ун-та, 2013. 156 с.

10. Костромин А.В. Эконометрика. Изд-во: КноРус, 2015. 232 с.

11. Новиков А.И. Эконометрика. М: ИНФРА-М, 2014. 272 с.

12. Озерная С. А. Эконометрика: метод. указания к лабораторному практикуму по специальностям «Бизнес-информатика», «Менеджмент», «Финансы и кредит». Самара, 2013. 76 с.

13. Соколов Г.А. Эконометрика. Теоретические основы: учеб. пособие. М: ИНФРА-М, 2012. 216 с.

14. Эконометрика / под ред. член-корреспондента РАН И.И. Елисеевой. М: Изд-во Юрайт, 2012. 453 с.

15. Тимофеев В.С., Фаддеенков А.В., Щеколдин В.Ю. Эконометрика: учебник для бакалавров. М: Изд-во Юрайт, 2013. 328 с.

16. Елисеева И.И. Эконометрика: учебник для магистров. М: Изд-во Юрайт, 2014. 449 с.

ПРИЛОЖЕНИЕ 1

Таблица значений *F* – критерия ФИШЕРА

k2	α	k1(число степеней свободы)											
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	**0,1**	39,9	49,5	53,6	55,8	57,2	58,2	58,9	59,4	59,9	60,2	60,5	60,7
	0,05	161	200	216	225	230	234	237	239	241	242	243	244
2	**0,10**	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,40	9,41
	0,05	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4
	0,01	98,5	99,2	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4
3	**0,10**	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,22
	0,05	10,01	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74
	0,01	34,1	30,3	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	27,1
4	**0,10**	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,91	3,90
	0,05	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91
	0,01	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,4
5	**0,10**	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,28	3,24
	0,05	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,71	4,68
	0,01	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,96	9,89
6	**0,10**	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,92	2,90
	0,05	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
	0,01	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72

Продолжение табл.

7	**0,10**	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,68	2,67
	0,05	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57
	0,01	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47
8	**0,10**	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,52	2,50
	0,05	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28
	0,01	11,3	8,65	7,59	7,01	6,63	6,37	3,18	6,03	5,91	5,81	5,73	5,67
9	**0,10**	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,40	2,38
	0,05	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07
	0,01	10,5	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11
10	**0,10**	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,30	2,28
	0,05	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91
	0,01	10,0	7,56	6,55	5,99	5,54	5,39	5,20	5,06	4,94	4,85	4,77	4,71
11	**0,10**	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,23	2,21
	0,05	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79
	0,01	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40

Окончание табл.

k2	**α**	**k1(число степеней свободы)**									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
12	**0,05**	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	**0,05**	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	**0,05**	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	**0,05**	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	**0,05**	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	**0,05**	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	**0,05**	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	**0,05**	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	**0,05**	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35

ПРИЛОЖЕНИЕ 2.

Таблица критических значений t-критерия Стьюдента

k – число степеней свободы, α – уровень значимости

α \ k	0,4	0,25	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
1	0,325	1,000	3,078	6,341	12,706	31,821	63,657	31,831	6,366
2	0,289	0,816	1,886	2,920	4,303	6,965	9,925	22,327	31,6
3	0,277	0,765	1,638	2,353	3,182	4,541	5,841	10,214	12,94
4	0,271	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,859
6	0,265	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	0,262	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587

Окончание табл.

\ α / k \	0,4	0,25	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
11	0,259	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,258	0,694	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,258	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,257	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,257	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850

Учебное издание

ЭКОНОМЕТРИКА. ПАРНАЯ РЕГРЕССИЯ

Методические указания к лабораторным работам

Составители: ***Котенко Андрей Петрович,***
Кузнецова Ольга Александровна

Редактор Ю.Н. Литвинова
Доверстка Т.С. Зинкина

Подписано в печать 20.08.2016. Формат 60х84 1/16.
Бумага офсетная. Печать офсетная. Печ. л. 3,0.
Тираж 100 экз. Заказ . Арт. – 63/2016.

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ имени академика С.П.КОРОЛЕВА»
443086 САМАРА, МОСКОВСКОЕ ШОССЕ, 34

ИЗДАТЕЛЬСТВО САМАРСКОГО УНИВЕРСИТЕТА
443086 САМАРА, МОСКОВСКОЕ ШОССЕ, 34