

Лабораторная работа №1
по дисциплине «Эконометрика»
Простая линейная регрессия

Тема:

Модели простой и множественной линейной регрессии

Цели:

- 1) Изучить основные положения регрессионного анализа.
- 2) Рассмотреть принципы статистического оценивания параметров модели для случая простой линейной регрессии.
- 3) Освоить метод наименьших квадратов для случая простой линейной регрессии.
- 4) Научиться пользоваться готовыми регрессионными моделями.

Ход работы:

Контрольные вопросы

- 1) Что такое эконометрическая модель (запишите пример)?
- 2) Для чего служит эконометрическая модель?
- 3) Какие этапы построения эконометрической модели?
- 4) Что такое эндогенные переменные?
- 5) Что такое экзогенные переменные?
- 6) Что такое шоковые переменные?
- 7) Что такое параметры модели?
- 8) Чем отличаются модели простой и множественной регрессии (запишите несколько примеров)?
- 9) Какие две основные задачи регрессионного анализа?
- 10) Какие существуют методы оценивания параметров модели?
- 11) Что значит 1-ое классическое предположение модели МНК: $M\{\varepsilon_t\} = 0$, $t = \overline{1, T}$? Приведите примеры, когда оно не выполняется.
- 12) Что значит 2-ое классическое предположение модели МНК: $M\{\varepsilon_t \cdot \varepsilon_\tau\} = 0$, $t \neq \tau$, $t, \tau = \overline{1, T}$? Приведите примеры, когда оно не выполняется.
- 13) Что значит 3-ее классическое предположение модели МНК: $D\{\varepsilon_t\} = \sigma^2$, $t = \overline{1, T}$? Приведите примеры, когда оно не выполняется.
- 14) Что значит 4-ое классическое предположение модели МНК: экзогенные переменные измеряются без ошибок и образуют линейно-независимые векторы? Приведите примеры, когда оно не выполняется.
- 15) Что значит 5-ое классическое предположение модели МНК: $\varepsilon_t \sim N(0, \sigma^2)$, $t = \overline{1, T}$? Приведите примеры, когда оно не выполняется.
- 16) Как строятся гипотезы о проверке значимости параметров модели?
- 17) Как строятся доверительные интервалы для параметров модели?
- 18) Что такое коэффициент детерминации? Как с его помощью оценивается адекватность модели?
- 19) Что такое исправленный коэффициент детерминации? В чем его отличие от обычного коэффициента детерминации?
- 20) Как строятся доверительные интервалы для зависимой переменной?
- 21) Какую информацию можно узнать анализируя остатки?
- 22) Какое распределение должны иметь остатки?

Часть I. Построение модели простой линейной регрессии в Excel

- 1) В Excel создайте следующую таблицу:

t	x_t	y_t
1	14	222
2	16	241
3	17	243
4	19	285
5	18	253
6	15	247
7	14	246
8	18	276
9	17	261
10	15	254
11	13	229

12	16	249
13	17	252
14	19	279
15	18	262
16	20	275
17	12	211
18	13	220
19	12	218
20	14	232

2) Постройте точечный график данной зависимости. Масштаб по оси Y установите от 200 до 300, по оси X – от 12 до 20.

3) По полученному графику дайте предварительную оценку следующим характеристикам модели $\hat{y}_t = \hat{a}_0 + \hat{a}_1 \cdot \hat{x}_t$:

3.1) знак и значение \hat{a}_0 ;

3.2) знак и значение \hat{a}_1 .

4) Достройте исходную таблицу до следующей:

t	x_t	y_t	x_t^2	$x_t - \bar{x}$	$(x_t - \bar{x})^2$	$y_t - \bar{y}$	$(y_t - \bar{y})^2$	$(x_t - \bar{x}) \cdot (y_t - \bar{y})$
1	14	122						
2	16	141						
...
...
20	14	132						
сумма	-							
среднее	-							

5) На основе данных полученной таблицы рассчитайте коэффициенты модели $\hat{a}_1 = \frac{\sum_{t=1}^T (x_t - \bar{x}) \cdot (y_t - \bar{y})}{\sum_{t=1}^T (x_t - \bar{x})^2}$, где

$$T = 20 \text{ и } \hat{a}_0 = \bar{y} - \hat{a}_1 \cdot \bar{x}.$$

6) Достройте к таблице следующие столбцы:

t	\hat{y}_t	e_t	e_t^2	s_t^2
1				
2				
...
...
20						
сумма	-					
среднее	-					

где $\hat{y}_t = \hat{a}_0 + \hat{a}_1 \cdot \hat{x}_t$ – это прогнозное значение модели; $e_t = y_t - \hat{y}_t$ – это ошибка модели; $s_t^2 = \frac{e_t^2}{T-2}$ – это слагаемое, сумма которых дает несмещенную оценку дисперсии случайных переменных.

7) Рассчитайте несмещенную оценку дисперсии случайных переменных $s^2 = \sum_{t=1}^T s_t^2$.

8) Рассчитайте несмещенную оценку дисперсии переменной \hat{a}_1 , $s_{a_1}^2 = \frac{s^2}{T \cdot (\overline{x^2} - (\bar{x})^2)}$.

9) Рассчитайте несмещенную оценку дисперсии переменной \hat{a}_0 , $s_{a_0}^2 = s_{a_1}^2 \cdot \overline{x^2}$.

10) С помощью функции СТЬЮДРАСПОБР($1-\gamma$, $T-2$) рассчитайте квантиль распределения Стьюдента $t_{\gamma, T-2}$ для уровней $\gamma = 0.90, 0.95, 0.99$.

11) Рассчитайте наблюдаемое значение статистики Стьюдента для параметра \hat{a}_1 , $t_{\hat{a}_1} = \frac{\hat{a}_1}{\frac{s_{a_1}}{a_1}}$ и проверьте гипотезу о статистической значимости параметра для уровней $\gamma = 0.90, 0.95, 0.99$.

12) Рассчитайте наблюдаемое значение статистики Стьюдента для параметра \hat{a}_0 , $t_{\hat{a}_0} = \frac{\hat{a}_0}{\frac{s_{a_0}}{a_0}}$ и проверьте гипотезу о статистической значимости параметра для уровней $\gamma = 0.90, 0.95, 0.99$.

13) Постройте доверительные интервалы для параметра \hat{a}_1 , $(\hat{a}_1 - t_{\gamma, T-2} \cdot s_{a_1}; \hat{a}_1 + t_{\gamma, T-2} \cdot s_{a_1})$ для уровней $\gamma = 0.90, 0.95, 0.99$.

14) Постройте доверительные интервалы для параметра \hat{a}_0 , $(\hat{a}_0 - t_{\gamma, T-2} \cdot s_{a_0}; \hat{a}_0 + t_{\gamma, T-2} \cdot s_{a_0})$ для уровней $\gamma = 0.90, 0.95, 0.99$.

15) Рассчитайте коэффициент детерминации модели $R^2 = 1 - \frac{\sum_{t=1}^T e_t^2}{\sum_{t=1}^T (y_t - \bar{y})^2}$, дайте предварительное заключение об адекватности модели.

16) С помощью функции ФРАСПОБР($1-\gamma$, 2, $T-2$) рассчитайте квантиль распределения Фишера $F_{\gamma, 2, T-2}$ для уровней $\gamma = 0.90, 0.95, 0.99$.

17) Рассчитайте наблюдаемое значение статистики Фишера для коэффициента детерминации $F_{R^2} = \frac{R^2}{1-R^2} \cdot (T-2)$ и проверьте гипотезу о статистической значимости коэффициента детерминации для уровней $\gamma = 0.90, 0.95, 0.99$.

18) С помощью полученной модели постройте прогнозы \hat{y}_{np} для значения x_{np} от 12 до 20. Занесите данные в таблицу.

19) Для каждого из прогнозов постройте доверительный интервал вида $(\hat{y}_{np} - t_{\gamma, T-2} \cdot s \cdot \sqrt{\frac{1}{T} + \frac{(x_{np} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2}}; \hat{y}_{np} + t_{\gamma, T-2} \cdot s \cdot \sqrt{\frac{1}{T} + \frac{(x_{np} - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2}})$ для уровней $\gamma = 0.90, 0.95, 0.99$.

20) Для каждого из уровней $\gamma = 0.90, 0.95, 0.99$ постройте график $\hat{y}_{np}(x_{np})$, а также график верхней и нижней границы доверительного интервала. Масштаб по оси Y установите от 190 до 310, по оси X – от 12 до 20.

21) Постройте точечный график ошибок e_t . Сделайте вывод о распределении остатков построенной модели.

Часть II. Построение модели простой линейной регрессии в EViews

1) Откройте в EViews файл, созданный в первой части лабораторной работы. Для этого выполните команду *File-Open-Foreign Data as Workfile...*

2) В качестве диапазона для импорта укажите адрес таблицы с исходными данными (см. рис. 1). Например, если таблица расположена в начале листа *Simple*, то адрес будет выглядеть следующим образом: *Simple!\$A\$1:\$C\$21*.

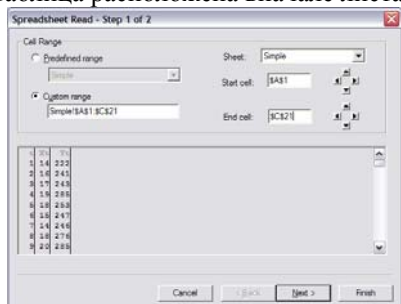


Рисунок 1. Импорт данных из Excel

3) Постройте точечный график исходных данных. Для этого выполните следующие действия:

- 3.1) выделите в рабочем окне переменные XT и YT ;
- 3.2) выполните двойной щелчок и выберите пункт меню *Open Group* (см. рис. 2);
- 3.3) в появившемся окне выполните команду *View-Graph-XY line-XY Pairs* (см. рис. 3).

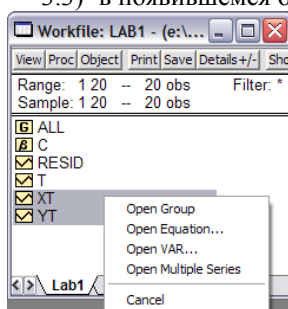


Рисунок 2. Открытие группы

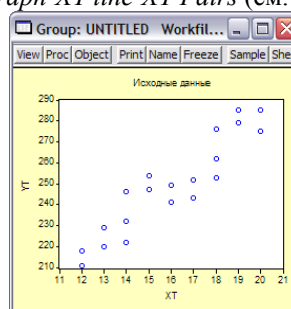


Рисунок 3. Исходные данные

- 4) Дайте оценку параметрам модели $y_t = a_0 + a_1 \cdot x_t$. Для этого выполните команду меню *Quick-Estimate Equation...* и введите следующее уравнение спецификации в явном виде $YT = C(1) + C(2) * XT$ или сокращенно $YT C XT$ (см. рис. 4). В качестве метода оценки параметров выберите МНК (*LS – Least Squares*).
- 5) В полученном отчете (см. рис. 5) найдите следующие величины:
 - 5.1) параметры модели и сопутствующие им значения t-статистик;
 - 5.2) обычный и исправленный коэффициенты детерминации;
 - 5.3) значение F-статистики.

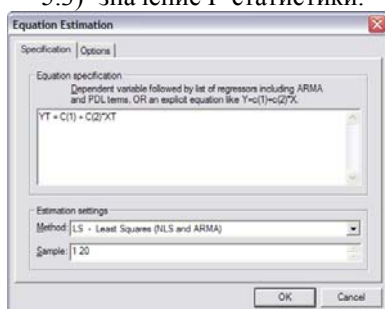


Рисунок 4. Спецификация модели

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	118.3250	12.45232	9.502246	0.0000
XT	8.164063	0.768720	10.62033	0.0000

R-squared	0.862376	Mean dependent var	248.9500
Adjusted R-squared	0.854731	S.D. dependent var	22.81845
S.E. of regression	6.697077	Akaike info criterion	7.258491
Sum squared resid	1361.505	Schwarz criterion	7.358064
Log likelihood	-70.58491	F-statistic	112.7914
Durbin-Watson stat	1.611259	Prob(F-statistic)	0.000000

Рисунок 5. Отчет спецификации

- 6) Сделайте вывод о значимости параметров модели и коэффициента детерминации.
- 7) Постройте прогнозы модели, для этого выполните команду *Forecast* и сохраните переменную как YTF (см. рис. 6). Затем выделите в рабочем окне переменные XT , YT и YTF , откройте их как группу и постройте график уравнения регрессии с экспериментальными точками (команда *View-Graph-XY line-XY Pairs*) (см. рис. 7).

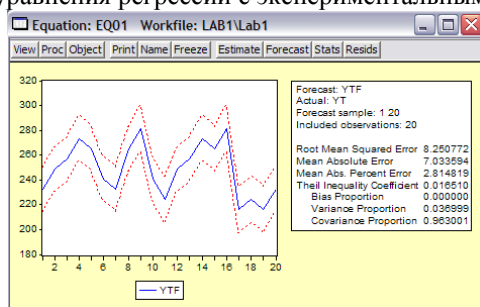


Рисунок 6. Отчет прогнозирования

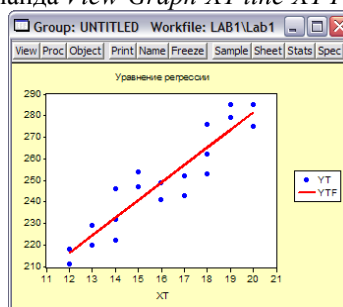


Рисунок 7. Уравнение регрессии

- 8) Постройте гистограмму распределения остатков модели (команда *View-Residual Tests-Histogram - Normality Test*), с помощью полученного отчета укажите уровень значимости, на котором может быть принята гипотеза о нормальном распределении остатков (см. рис. 8).

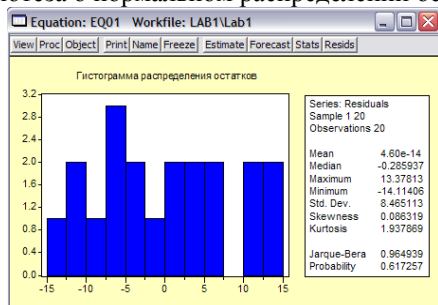


Рисунок 8. Остатки