САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ

«ЭКОНОМЕТРИКА»

Илона Юловна Парик

К.э.н. Доцент

Кафедра статистики и эконометрики

Основная литература

- Эконометрика: учебник / И.И. Елисеева [и др.]; под ред. И. И.Елисеевой. - М.: Издательство Юрайт, 2012
- Эконометрика : учеб. для студентов вузов по специальности 080601 "Статистика" и др. междисциплинар. специальностям / [И.И.Елисеева и др.] ; под ред. И.И.Елисеевой. - Москва: Проспект, 2011
- Курышева С.В. Анализ временных рядов и прогнозирование: учебное пособие / С.В.Курышева, М.В. Боченина. – СПб. : Изд-во СПбГЭУ, 2014
- Практикум по эконометрике: учеб. пособие / И.И. Елисеева, С.В.Курышева, Н.М.Гордеенко и др.; под ред. И. И.Елисеевой. - 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2008

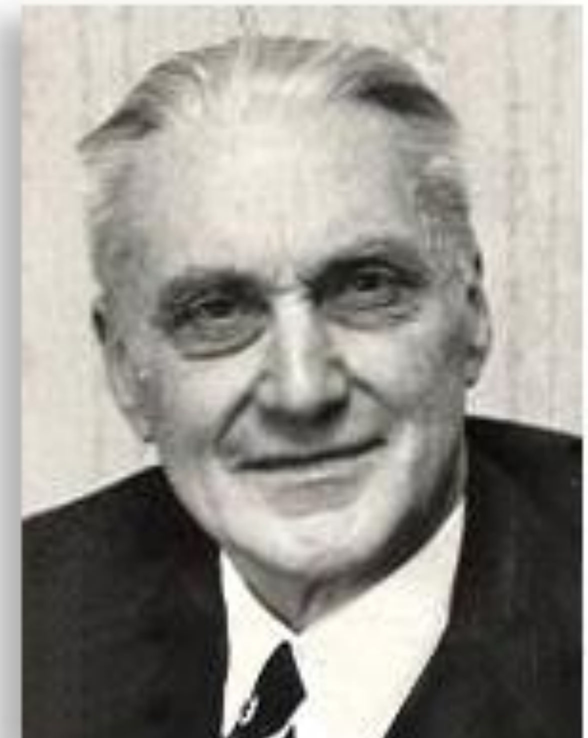
Дополнительная литература

- Айвазян С.А. Методы эконометрики. - М.: Инфра-М, 2010
- Афанасьев В.Н., Юзбашев М. М. Анализ временных рядов и прогнозирование. – М.: Финансы и статистика, 2010
- Доугерти К. Введение в эконометрику: Учебник. 2-е изд. / Пер. с англ. – М.: ИНФРА – М, 2007
- Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный дисциплина: учебник. – М.: Дело, 2009
- Чураков Е. П. Прогнозирование эконометрических временных рядов: учебник. - М.: Финансы и статистика, 2008

Рагнар Антон Киттиль Фриш

(*норв. Ragnar Anton Kittil Frisch*)

(1895-1973)



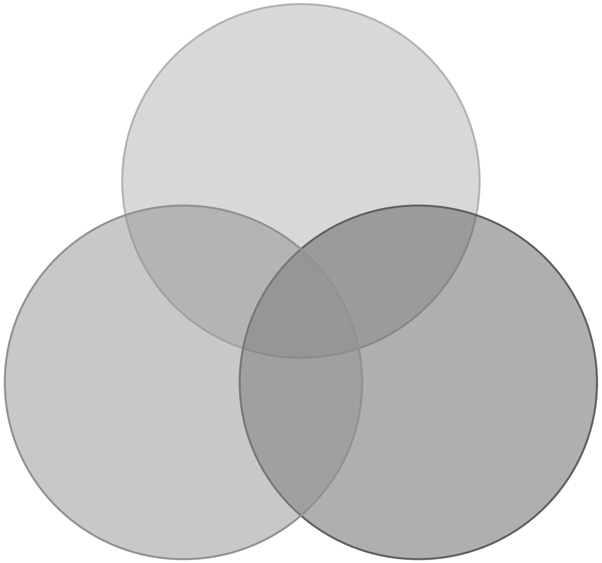
- 1926 г. норвежский экономист Рагнар Фриш (1895-1973) предложил использовать термин «эконометрика» для обозначения самостоятельной отрасли научных исследований

- Развернутое определение эконометрики было дано Рагнаром Фришем во вступительной статье первого номера журнала "Эконометрика" в 1933 г.

Эконометрика – это наука, которая дает конкретное количественное выражение общим (качественным) взаимосвязям экономических явлений и процессов, обусловленным экономической теорией

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ

БАЗОВЫЕ КОМПОНЕНТЫ ЭКОНОМЕТРИКИ



- На основе экономической теории разрабатываются концепции развития изучаемых процессов
- С помощью статистики эти процессы выражаются в статистических показателях
- Математико-статистические методы позволяют строить модели изучаемых процессов, оценивать их параметры, степень соответствия реальным данным и прогнозировать развитие изучаемого явления

Главный инструмент эконометрики – эконометрическая модель, параметры которой оцениваются с помощью методов математической статистики

Этапы построения эконометрической модели

- Теоретическое описание рассматриваемого экономического процесса с отражением существующих тенденций
- Сбор данных, анализ их качества
- Спецификация модели. Устанавливаются экзогенные (внешние) и эндогенные (внутренние) переменные, выявляются связи и соотношения, определяется вид модели исходя из соответствующей теории связи между переменными
- Оценка параметров модели
- Верификация модели, то есть проверка достоверности построенной модели
- Интерпретация результатов

ПРИМЕНЕНИЕ ПАРНОЙ РЕГРЕССИИ В ЭКОНОМЕТРИЧЕСКИХ ИССЛЕДОВАНЯХ

1. Выбор типа математической функции при построении уравнения регрессии
2. Оценка параметров уравнения парной линейной регрессии
3. Показатели силы связи в моделях парной регрессии
4. Показатели тесноты связи в моделях парной регрессии
5. Статистическая оценка достоверности регрессионной модели
6. Интервальная оценка параметров уравнения парной регрессии
7. Использование модели парной регрессии для прогнозирования

Задачи корреляционно-регрессионного анализа

- Измерение параметров уравнения, выражающего связь между признаками. Эта задача решается оценкой параметров уравнения регрессии

- Измерение тесноты связи между признаками. Данная задача решается показателей корреляции

Виды функций, наиболее часто используемые в эконометрическом моделировании

Линейная	$y = a + bx$
Гипербола	$y = a + b/x$
Парабола второго порядка	$y = a + bx + cx^2$
Логарифмическая функция	$y = a + b \cdot \ln x$
Степенная функция	$y = ax^b$
Показательная функция	$y = ab^x$
Экспонента	$y = e^{a+bx}$

Методы выбора типа математической функции

- Аналитический метод (теоретический анализ связи рассматриваемого фактора и результата)
- Графический метод
- Экспериментальный метод

Линеаризация нелинейных уравнений

Функция	***Исходное уравнение***	***Преобразованное уравнение***
Гипербола	$y = a + \frac{b}{x}$	$1/x = z$ $y = a + bz$
Степенная	$y = ax^b$	$\ln y = \ln a + b \ln x$
Показательная	$y = ab^x$	$\ln y = \ln a + x \ln b$
Экспонента	$y = e^{a+bx}$	$\ln y = a + bx$

Оценка параметров уравнения парной линейной регрессии

- Для оценки параметров функций, линейных по параметрам, используется ***метод наименьших квадратов (МНК)***

- МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака от теоретических минимальна:

$$\sum (y-\hat{y})^2 \to \min$$

Оценка параметров уравнения парной линейной регрессии

Для линейных и нелинейных уравнений, приводимых к линейным, решается следующая система относительно a и b:

$$\begin{cases} \sum y = na + b\sum x \\ \sum yx = a\sum x + b\sum x^2 \end{cases} \qquad \begin{cases} \overline{y} = a + b\overline{x} \\ \overline{yx} = a\overline{x} + b\overline{x^2} \end{cases}$$

Формулы расчета параметров уравнения парной регрессии

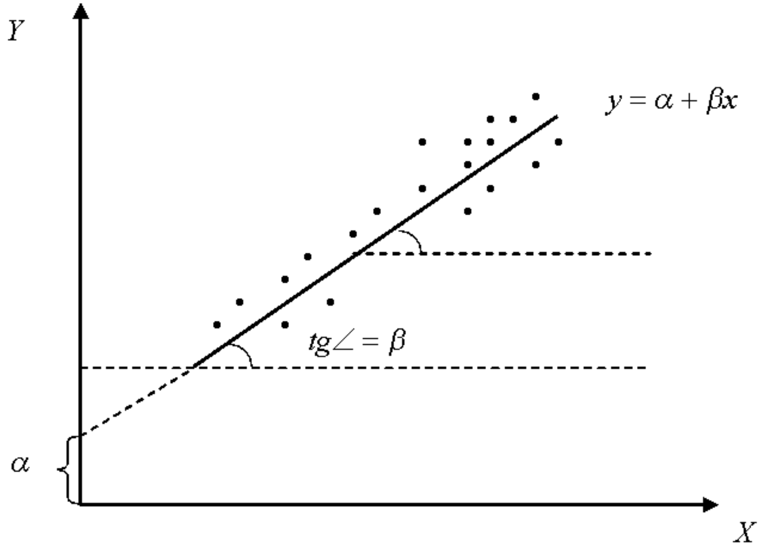
$$b = \frac{\overline{yx} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} \qquad a = \overline{y} - b\overline{x}$$

a – *свободный член уравнения регрессии (пересечение, intercept).* Экономически не интерпретируется.

b – *наклон линии регрессии (slope) или коэффициент регрессии*. Он является мерой зависимости переменной y от переменной x.

В линейном уравнении регрессии параметр b является ***абсолютным показателем силы связи***

Линия регрессии



Условия применения МНК

- Модель регрессии должна быть линейной по параметрам
- Значения ошибки (остатка)- случайные. Их изменение не образует определенной модели
- Число наблюдений должно быть больше числа оцениваемых параметров (в 5-6 раз)
- Значения переменной *x* не должны быть одинаковыми
- Изучаемая совокупность должна быть однородной
- Отсутствие взаимосвязи между фактором *x* и остатком
- Модель регрессии должна быть корректно специфицирована
- В модели не должно наблюдаться тесной взаимосвязи между факторами (условие для множественной регрессии)

Пример

Федеральный округ	Инвестиции в основной капитал на душу населения, тыс. руб. 2009 г.	Валовой региональный продукт на душу населения, тыс. руб. 2009 г.
Центральный	51,9	308,3
Северо-Западный	69,4	253,2
Южный	51,7	145
Северо-Кавказский	20	86,3
Приволжский	42,4	163,3
Уральский	109,1	358,4
Сибирский	42,7	173,4
Дальневосточный	106,4	268,3

Продолжение примера

№	y	x	$y*x$	x^2	y^2
1	308,3	51,9	16000,77	2693,61	95048,89
2	253,2	69,4	17572,08	4816,36	64110,24
3	145,0	51,7	7496,50	2672,89	21025
4	86,3	20,0	1726,00	400,00	7447,69
5	163,3	42,4	6923,92	1797,76	26666,89
6	358,4	109,1	39101,44	11902,81	128450,6
7	173,4	42,7	7404,18	1823,29	30067,56
8	268,3	106,4	28547,12	11320,96	71984,89
Итого	1756,2	493,6	124772	37427,68	444801,7
Среднее значение	219,525	61,7	15596,5	4678,46	55600,22

Линейная зависимость

$$b = \frac{15596{,}5 - 219{,}525 \cdot 61{,}7}{4678{,}46 - (61{,}7)^2} = 2{,}354$$

$$a = 219{,}525 - 2{,}354 \cdot 61{,}7 = 74{,}28$$

$$\hat{y} = 74{,}28 + 2{,}354x$$

Продолжение примера

№	y	x	lgy	lgx	lgy*lgx	$(\lg x)^2$	y^2
1	308,3	51,9	2,488974	1,715167	4,269006	2,941799	95048,89
2	253,2	69,4	2,403464	1,841359	4,425641	3,390605	64110,24
3	145,0	51,7	2,161368	1,713491	3,703484	2,93605	21025
4	86,3	20,0	1,936011	1,30103	2,518808	1,692679	7447,69
5	163,3	42,4	2,212986	1,627366	3,601338	2,64832	26666,89
6	358,4	109,1	2,554368	2,037825	5,205354	4,15273	128450,6
7	173,4	42,7	2,239049	1,630428	3,650608	2,658295	30067,56
8	268,3	106,4	2,428621	2,026942	4,922672	4,108492	71984,89
Итого	1756,2	493,6	18,42484	13,89361	32,29691	24,52897	444801,7
Среднее значение	219,525	61,7	2,303105	1,736701	4,037114	3,066121	55600,22

Степенная зависимость

$$y = ax^b$$

$$\lg y = \lg a + b \lg x$$

$$b = \frac{\overline{\lg y \lg x} - \overline{\lg y} \cdot \overline{\lg x}}{\overline{(\lg x)^2} - (\overline{\lg x})^2} = 0{,}746$$

$$\lg a = \overline{\lg y} - b\overline{\lg x} = 1{,}007$$

Степенная зависимость

$$\lg y = 1{,}007 + 0{,}746 \lg x$$

$$\hat{y} = 10^{1,007} x^{0,746} = 10{,}16 x^{0,746}$$

Показатели силы связи в моделях парной регрессии

- **Абсолютные.** Показывают, на сколько единиц в среднем изменяется результативный признак при изменении факторного признака на одну единицу. В линейном уравнении параметр b - абсолютный показатель силы связи

- **Относительные (*коэффициенты эластичности*).** Показывают, на сколько процентов в среднем изменяется результативный признак при изменении факторного признака на один процент

$$Э = f'(x)\frac{x}{y}$$

Абсолютные и относительные показатели силы связи для основных видов функций

Функция	Исходное уравнение	Показатели силы связи	
		Абсолютный	Относительный
Линейная	$y = a + bx$	b	$\frac{bx}{a+bx} = b\frac{x}{y}$
Степенная	$y = ax^b$	abx^{b-1}	b
Показательная	$y = ab^x$	$ab^x \ln b$	$x \ln b$
Гипербола	$y = a + \frac{b}{x}$	$-\frac{b}{x^2}$	$-\frac{b}{ax+b}$
Парабола второго порядка	$y = a + bx + cx^2$	$b + 2cx$	$\frac{(b+2cx)x}{a+bx+cx^2}$

Продолжение примера

$$\hat{y} = 74{,}28 + 2{,}354x$$

С увеличением инвестиций в основной капитал на 1 тыс. руб. ВРП на душу населения возрастает в среднем на 2,354 тыс. руб.

Продолжение примера

Линейная функция:

$$Э = b\frac{\bar{x}}{\bar{y}} = 2{,}354\frac{61{,}7}{219{,}525} = 0{,}662\%$$

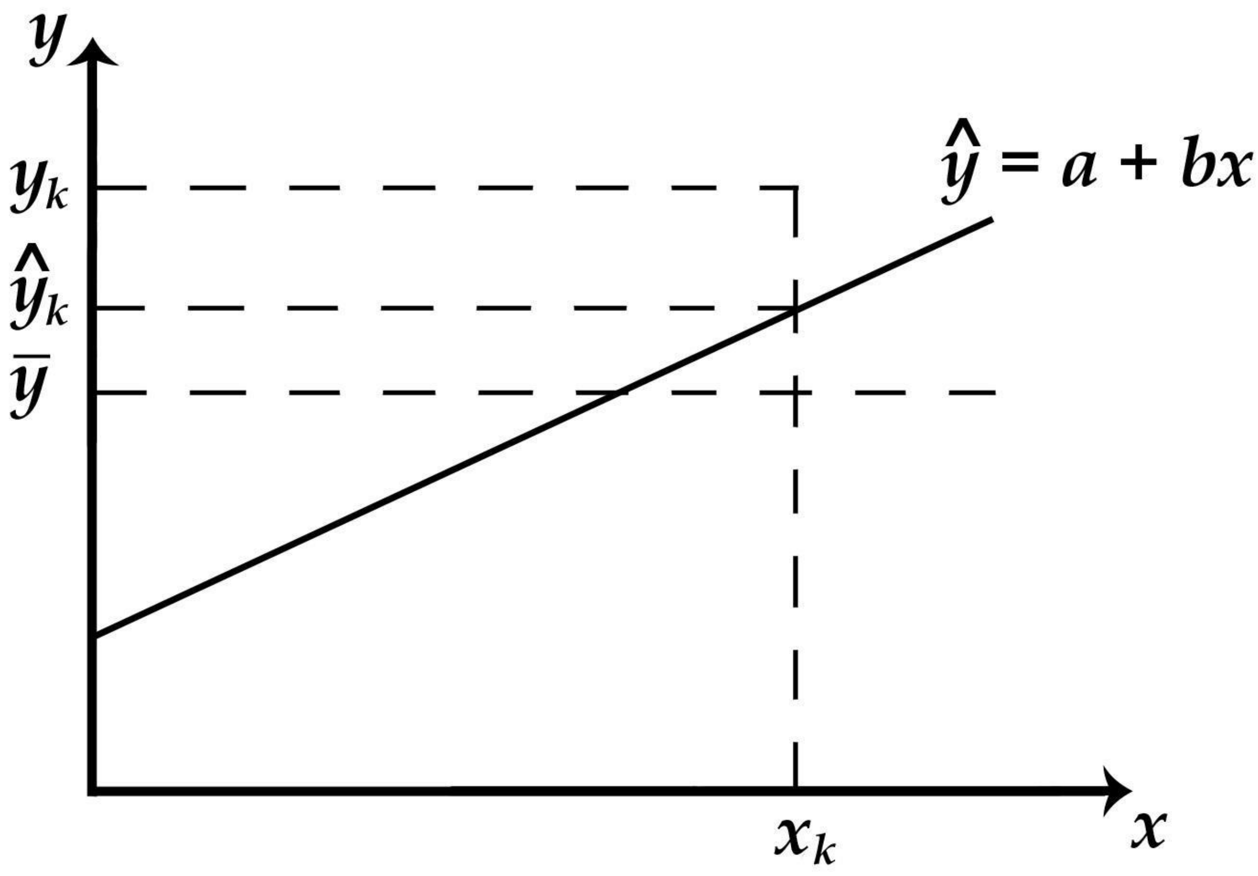
Степенная функция:

$$Э = b = 0{,}746\%$$

Показатели тесноты связи в моделях парной регрессии

Коэффициент детерминации

показывает долю вариации (дисперсии) результативного признака, объясняемую регрессией, в общей вариации результата


y
$\hat{y} = a + bx$
y_k
$\hat{y}_k$
$\bar{y}$
x_k
x

Правило сложения дисперсий

$\Sigma\left(y-\overline{y}\right)^2 = SS_T$ - общая сумма квадратов отклонений (*total sum of squares*)

$\Sigma\left(\hat{y}-\overline{y}\right)^2 = SS_R$ - факторная сумма квадратов отклонений (*sum of squares due to regression*)

$\Sigma\left(y-\hat{y}\right)^2 = SS_E$ - остаточная сумма квадратов отклонений (*sum of squares due to error*)

$$SS_T = SS_R + SS_E$$

Коэффициент детерминации

$$r^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

$$0 \le r^2 \le 1$$

Коэффициент корреляции

$$r_{yx} = b\frac{\sigma_x}{\sigma_y} = \frac{\overline{yx} - \overline{y}\cdot\overline{x}}{\sigma_y\sigma_x}$$

$$-1 \le r \le 1$$

Шкала значений коэффициента (индекса) корреляции

- До 0,3 связь слабая
- 0,3-0,5 связь умеренная
- 0,5-0,7 связь заметная
- 0,7-0,9 связь высокая
- 0,9-1,0 связь весьма высокая, близкая к функциональной

Свойства линейного коэффициента корреляции

- Это стандартизованный коэффициент регрессии
- Сравним для признаков, имеющих различные единицы измерения
- Если связь между *y* и *x* отсутствует, то $r_{yx} = 0$; если $r_{yx} = 0$, это не всегда означает отсутствия связи (связь может быть нелинейной)
- $r_{yx} = r_{xy}$

Продолжение примера

Линейная функция

$$\hat{y} = 74{,}28 + 2{,}354x$$

$$\sigma_y{}^2 = \overline{y^2} - \left(\overline{y}\right)^2 = 55600{,}22 - 219{,}525^2 = 7408{,}99$$

$$SS_T = \sigma_y^2 \cdot n = 7408{,}99 \cdot 8 = 59271{,}92$$

Продолжение примера

Расчет остаточной суммы квадратов отклонений по линейной функции

№	y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$
1	308,3	196,4526	111,8474	12509,84
2	253,2	237,6476	15,5524	241,8771
3	145,0	195,9818	-50,9818	2599,144
4	86,3	121,36	-35,06	1229,204
5	163,3	174,0896	-10,7896	116,4155
6	358,4	331,1014	27,2986	745,2136
7	173,4	174,7958	-1,3958	1,948258
8	268,3	324,7456	-56,4456	3186,106
Итого	Х	Х	Х	20629,75

Продолжение примера

Расчет теоретических значений результативного признака линейной функции

$$\hat{y} = 74{,}28 + 2{,}354x$$

$$\hat{y}_1 = 74{,}28 + 2{,}354 \cdot 51{,}9 = 196{,}4526$$

$$\hat{y}_2 = 74{,}28 + 2{,}354 \cdot 69{,}4 = 237{,}6476$$

…

Продолжение примера

Расчет коэффициента детерминации для линейной функции

$$r^2 = 1 - \frac{20629,75}{59271,92} = 0,652$$

Продолжение примера

Расчет теоретических значений результативного признака степенной функции

$$\hat{y} = 10{,}16x^{0,746}$$

$$\hat{y}_1 = 10{,}16 \cdot 51{,}9^{0,746} = 13206{,}91$$

$$\hat{y}_2 = 10{,}16 \cdot 69{,}4^{0,746} = 169{,}3616$$

...

Продолжение примера

Расчет остаточной суммы квадратов отклонений по степенной функции

№	y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$
1	308,3	193,3787	114,9213	13206,91
2	253,2	240,1861	13,0139	169,3616
3	145	192,8225	-47,8225	2286,989
4	86,3	94,94281	-8,64281	74,69818
5	163,3	166,3063	-3,00628	9,037691
6	358,4	336,5976	21,80243	475,3461
7	173,4	167,1833	6,216696	38,64731
8	268,3	330,3636	-62,0636	3851,888
Итого	X	X	X	20112,88

Продолжение примера

Расчет показателей корреляции

$$r = \sqrt{r^2} = \sqrt{0,652} = 0,807$$

$$r_{yx} = b\frac{\sigma_x}{\sigma_y} = 2,354\frac{29,522}{86,075} = 0,807$$

$$\sigma_y = \sqrt{7408,99} = 86,075$$

$$\sigma_x = \sqrt{4678,46 - 61,7^2} = 29,522$$

Статистическая проверка гипотез

Статистической гипотезой называется предположение о свойстве генеральной совокупности, которое можно проверить, опираясь на данные выборки. Обозначается буквой *H* (лат. *hypothesis*)

Статистическая оценка достоверности регрессионной модели

- Выдвигается $H_0 : r^2$ в генеральной совокупности $= 0$
- Выдвигается $H_1 : r^2$ в генеральной совокупности $\neq 0$
- Определяется уровень значимости α (1 минус доверительная вероятность)
- Рассчитывается критерий Фишера F
- Определяется табличное значение критерия Фишера $F_{tab.}$
- Фактическое значение сравнивается с табличным

- ***Критическая область*** – это область, попадание значения статистического критерия в которую приводит к отклонению H_0 . Вероятность попадания значения критерия в эту область равна приятому *уровню значимости* (1 минус доверительная вероятность)

- ***Область допустимых значений*** - область, попадание значения статистического критерия в которую приводит к принятию нулевой гипотезы

Оценка значимости уравнения регрессии

- Если $F>F_{tab.}$, то гипотеза о случайной природе оцениваемых характеристик отклоняется и признается статистическая значимость и надежность уравнения

- Если $F<F_{tab.}$, то гипотеза о случайной природе оцениваемых характеристик не отклоняется и признается статистическая незначимость, ненадежность уравнения регрессии

- *Число степеней свободы (degrees of freedom - df)* - число свободно варьируемых переменных
- $$df_T = df_R + df_E$$
  $$(n-1) = m + (n-m-1)$$

n - число единиц совокупности

m - число параметров при переменных(число факторов)

Факторная дисперсия на одну степень свободы	$MS_R = \dfrac{SS_R}{df_R} = \dfrac{\sum(\hat{y} - \bar{y})^2}{m}$
Остаточная дисперсия на одну степень свободы	$MS_E = \dfrac{SS_E}{df_E} = \dfrac{\sum(y - \hat{y})^2}{n - m - 1}$
F-критерий Фишера	$F = \dfrac{MS_R}{MS_E}$ $F = \dfrac{r_{yx}^{\ 2}}{1 - r_{yx}^{\ 2}} \cdot \dfrac{n - m - 1}{m}$

Продолжение примера

Расчет критерия Фишера

Для линейной функции:

$$F = \frac{r_{yx}^2}{1 - r_{yx}^2} \cdot \frac{n - m - 1}{m} = \frac{0,652}{1 - 0,652} \cdot \frac{8 - 1 - 1}{1} = 11,2$$

Для степенной функции:

$$F = \frac{0,661}{1 - 0,661} \cdot \frac{8 - 1 - 1}{1} = 11,7$$

Таблица дисперсионного анализа

Источник вариации	*df*	*SS*	*MS*	*F*
Регрессия	1	38642,17	38642,17	11,24
Остаток	6	20629,75	3438,29	X
Итого	7	59271,92	X	X

Оценка качества модели на основе ошибки аппроксимации

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

Продолжение примера

Расчет ошибки аппроксимации для линейной функции

№	x	y	$\hat{y}$	$y-\hat{y}$	$\left\lvert\frac{y-\hat{y}}{y}\right\rvert\cdot 100$
1	51,9	308,3	196,4526	111,8474	36,2788
2	69,4	253,2	237,6476	15,5524	6,1423
3	51,7	145,0	195,9818	-50,9818	35,1599
4	20,0	86,3	121,36	-35,06	40,6257
5	42,4	163,3	174,0896	-10,7896	6,6072
6	109,1	358,4	331,1014	27,2986	7,6168
7	42,7	173,4	174,7958	-1,3958	0,8050
8	106,4	268,3	324,7456	-56,4456	21,0382
Итого	Х	Х	Х	Х	154,2739

$$\overline{A}=\frac{154,2739}{8}=19,28\%$$

Оценка значимости коэффициентов регрессии

- Выдвигается H_0 : коэффициент регрессии в генеральной совокупности равен 0
- Выдвигается H_1 : коэффициент регрессии в генеральной совокупности не равен 0
- Определяется уровень значимости α
- Рассчитывается критерий Стьюдента t_b
- Определяется табличное (критическое) значение критерия Стьюдента $t_{tab.}$
- Фактическое значение сравнивается с табличным

- Если $t>t_{tab.}$, то H_0 отклоняется, то есть параметр b не случайно отличается от нуля, и сформировался под влиянием систематически действующего фактора

- Если $t<t_{tab.}$, то H_0 не отклоняется, и признается случайная природа формирования параметра b

Расчет критерия Стьюдента

$$t_b = \frac{b}{Se_b}$$

Se_b - случайная ошибка коэффициента регрессии

$$Se_b = \sqrt{\frac{MS_E}{\sum(x-\bar{x})^2}}$$

Продолжение примера

$$Se_b = \sqrt{\frac{MS_E}{\sum (x-\bar{x})^2}} = \sqrt{\frac{3438,29}{6972,38}} = 0,7022$$

$$\sigma_x{}^2 = \overline{x^2} - \left(\bar{x}\right)^2 = 4678,46 - (61,7)^2 = 871,548$$

$$\sum \left(x-\bar{x}\right)^2 = \sigma_x{}^2 \cdot n = 871,548 \cdot 8 = 6972,38$$

Продолжение примера

$$t_b = \frac{b}{Se_b} = \frac{2,354}{0,7022} = 3,35$$

$$t_b{}^2 = F = 3,35^2 = 11,22$$

- $t_{tab.}$ =2,447
- $t > t_{tab.}$

Построение доверительных интервалов для коэффициентов регрессии

$$\Delta_b = \pm t_{tab.} \cdot Se_{(b)}$$

$$b - \underbrace{t_{tab.} \cdot Se_{(b)}}_{\Delta_b} \le b \le b + \underbrace{t_{tab.} \cdot Se_{(b)}}_{\Delta_b}$$

$$\Delta_b = \pm 2{,}447 \cdot 0{,}7022 = 1{,}7183$$

$$2{,}354 - 1{,}718 \le b \le 2{,}354 + 1{,}718$$

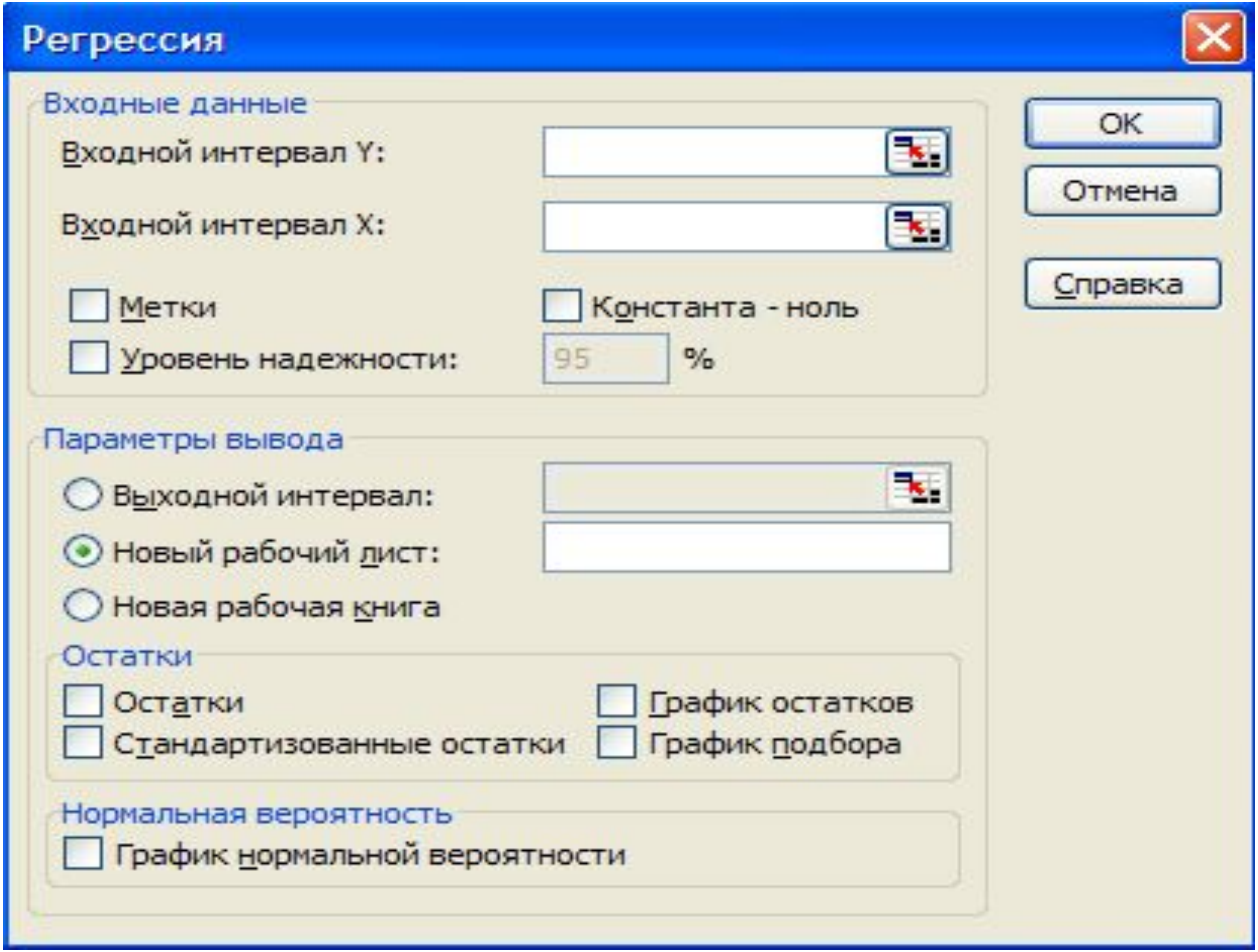
$$0{,}636 \le b \le 4{,}072$$

Расчет показателей регрессии и корреляции с помощью пакета анализа в Excel

- ***Установка пакета анализа:***
  - Кнопка «*Office*»
  - Параметры *Excel*
  - Надстройки
  - Надстройки *Excel*
  - Перейти
  - Пакет анализа
- ***После установки пакета анализа:***
  - Данные
  - Анализ данных
  - Регрессия

Расчет показателей регрессии и корреляции с помощью пакета анализа в Excel

- *В диалоговом окне* ***«регрессия»*** *задаются следующее параметры:*
- -***Входной интервал Y***, - водится ссылка на диапазон ячеек, содержащий данные результативного признака
- ***Входной интервал X***, - водится ссылка на диапазон ячеек, содержащий данные факторного признака
- -Если данные выделяются с названием граф, то устанавливается флажок ***метки***
- -*Параметры вывода*: ***выходной интервал*** (вводится ссылка на любую свободную ячейку на данном рабочем листе); *другой рабочий лист* или *другая рабочая книга*
- -ОК


Регрессия
Входные данные
Входной интервал Y:
Входной интервал X:
Метки
Константа - ноль
Уровень надежности:
95
%
Параметры вывода
Выходной интервал:
Новый рабочий лист:
Новая рабочая книга
Остатки
Остатки
Стандартизованные остатки
График остатков
График подбора
Нормальная вероятность
График нормальной вероятности
ОК
Отмена
Справка

Использование модели парной регрессии для прогнозирования

Точечный прогноз осуществляется путем подстановки в найденное уравнение регрессии прогнозного значения x_p:

$$\hat{y}_p = a + bx_p$$

- ***Интервальный прогноз***

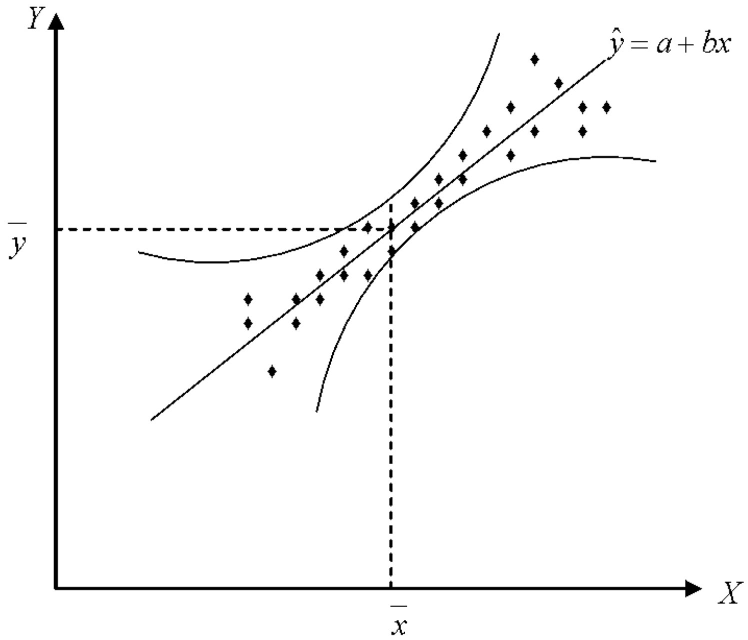
Определяется средняя ошибка прогнозного индивидуального значения *y*:

$$Se_{(\hat{y}_p)} = \sqrt{MS_E\left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}\right]}$$

Строится доверительный интервал прогноза:

$$\hat{y}_p - t_{tab.} \cdot Se_{(\hat{y}_p)} \le \hat{Y}_p \le \hat{y}_p + t_{tab.} \cdot Se_{(\hat{y}_p)}$$

95%-ый доверительный интервал



Продолжение примера

С вероятностью 0,95 построить доверительный интервал прогнозного значения валового регионального продукта на душу населения в предположении, инвестиции в основной капитал на душу населения увеличатся на 10% от своего среднего уровня.

- Найдем прогнозное значение инвестиций:

$$x_p = 493{,}6 \cdot 1{,}1 = 542{,}96$$

- Найдем по найденному уравнению регрессии прогнозное значение регионального продукта:

$$\hat{y}_p = 74{,}28 + 2{,}354 \cdot 542{,}96 = 1352{,}41$$

Продолжение примера

- Определим среднюю ошибку прогнозного индивидуального значения y:

$$Se_{(\hat{y}_p)} = \sqrt{MS_E\left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x - \bar{x})^2}\right]} =$$

$$= \sqrt{3438{,}29\left(1 + \frac{1}{8} + \frac{(542{,}96 - 493{,}6)^2}{6972{,}38}\right)} = 71{,}2$$

Продолжение примера

- Построим доверительный интервал прогноза:

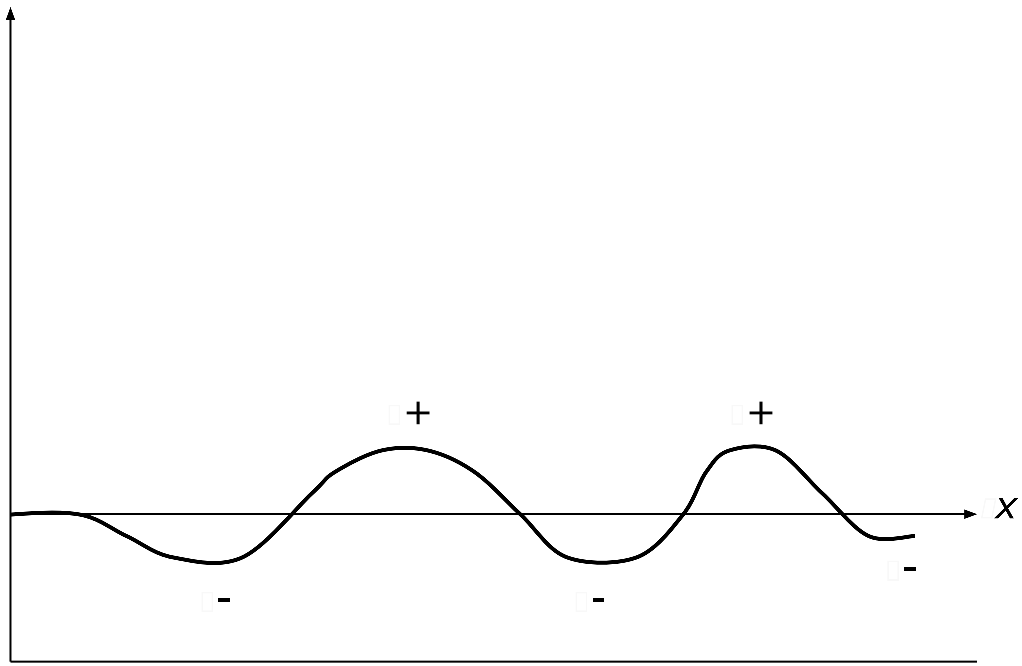
$$1352{,}41 - 2{,}447 \cdot 71{,}2 \le \hat{Y}_p \le 1352{,}41 + 2{,}447 \cdot 71{,}$$

$$1178{,}18 \le \hat{Y}_p \le 1526{,}64$$

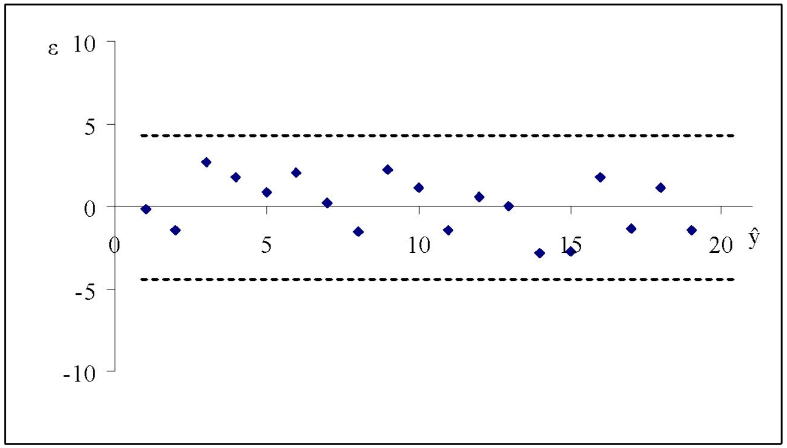
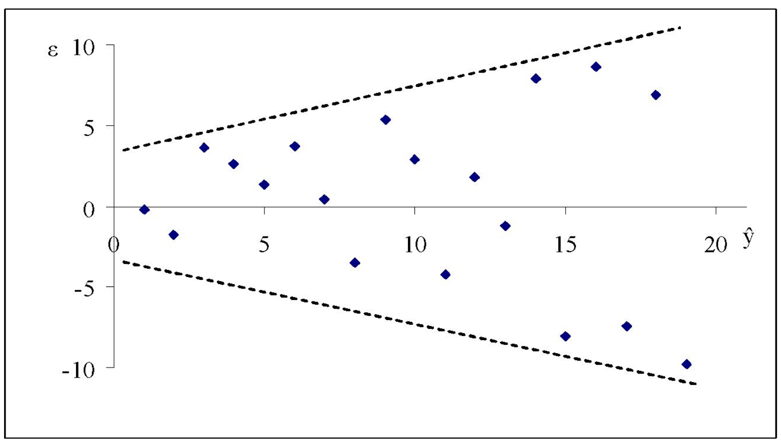
Свойства остатков

- Отсутствие связи между остатками и объясняющей переменной
- Отсутствие связи между остатками и предсказанными значениями
- Математическое ожидание остатков равно нулю
- Остатки имеют постоянную дисперсию. Дисперсия остатков равна единице. Постоянство дисперсии остатков называют гомоскедастичностью остатков. Если же дисперсия остатков непостоянна, то имеет место гетероскедастичность остатков
- Остатки не коррелированны между собой
- Остатки распределены по нормальному закону распределения

График остатков (residual plot) (случай гомоскедастичности)



Зависимость остатков от выровненного значения результата

	нет зависимости (гомоскедастичность)
	дисперсия остатков увеличивается с увеличением выровненного значения результата (один из случаев гетероскедастичности)