Эконометрика

Хасанова
Светлана Фанилевна

Лекция 1
«Парная регрессия»

Парная линейная регрессия

❖ **Основная цель – построить уравнение (модель) вида:**

$$Y = a + b*X$$

описывающее зависимость между зависимой переменной (результатом - Y) и независимой переменной (фактором – X).

a и b – называются параметрами модели.

Примеры зависимостей:

- Выручки предприятия от расходов на рекламу;
- Цены на нефть от курса доллара,
- Количества баллов по эконометрике от количества часов, потраченных на её изучение

и т.д.

Пример построения парной линейной регрессии

Изучается влияние объема ВВП на объем экспорта в стране.

Для корреляционно-регрессионного анализа использована выборка за 10 лет:

Годы	T	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
ВНП	X	1000	1090	1150	1230	1300	1360	1400	1470	1500	1580
Экспорт	Y	190	220	240	240	260	250	280	290	310	350

Построим график по следующему правилу:

По оси Y – зависимая переменная или изучаемая величина (экспорт)

По оси X – независимая переменная или причинный фактор (ВВП)

Каждая точка графика соответствует каждому году

График

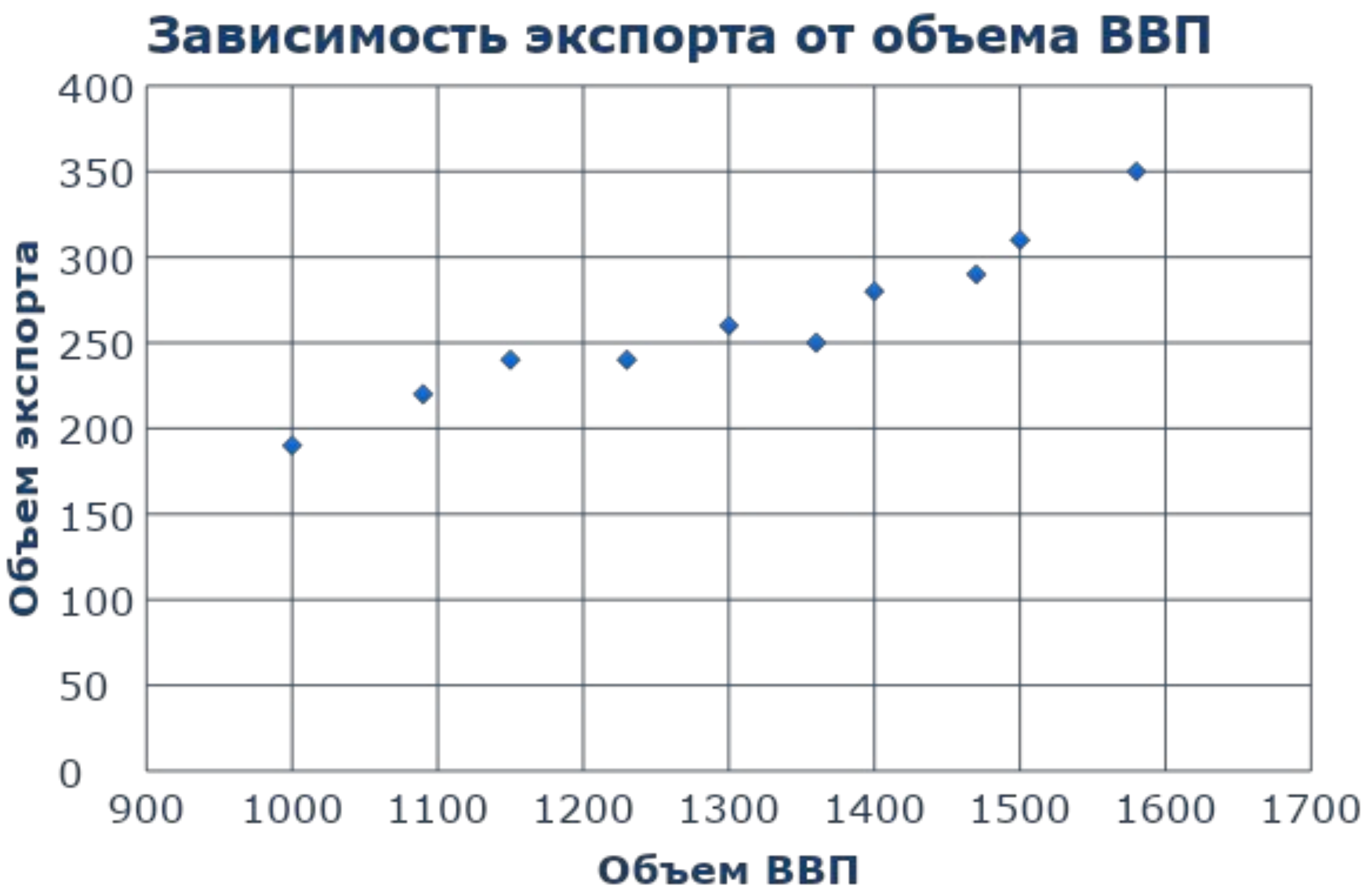
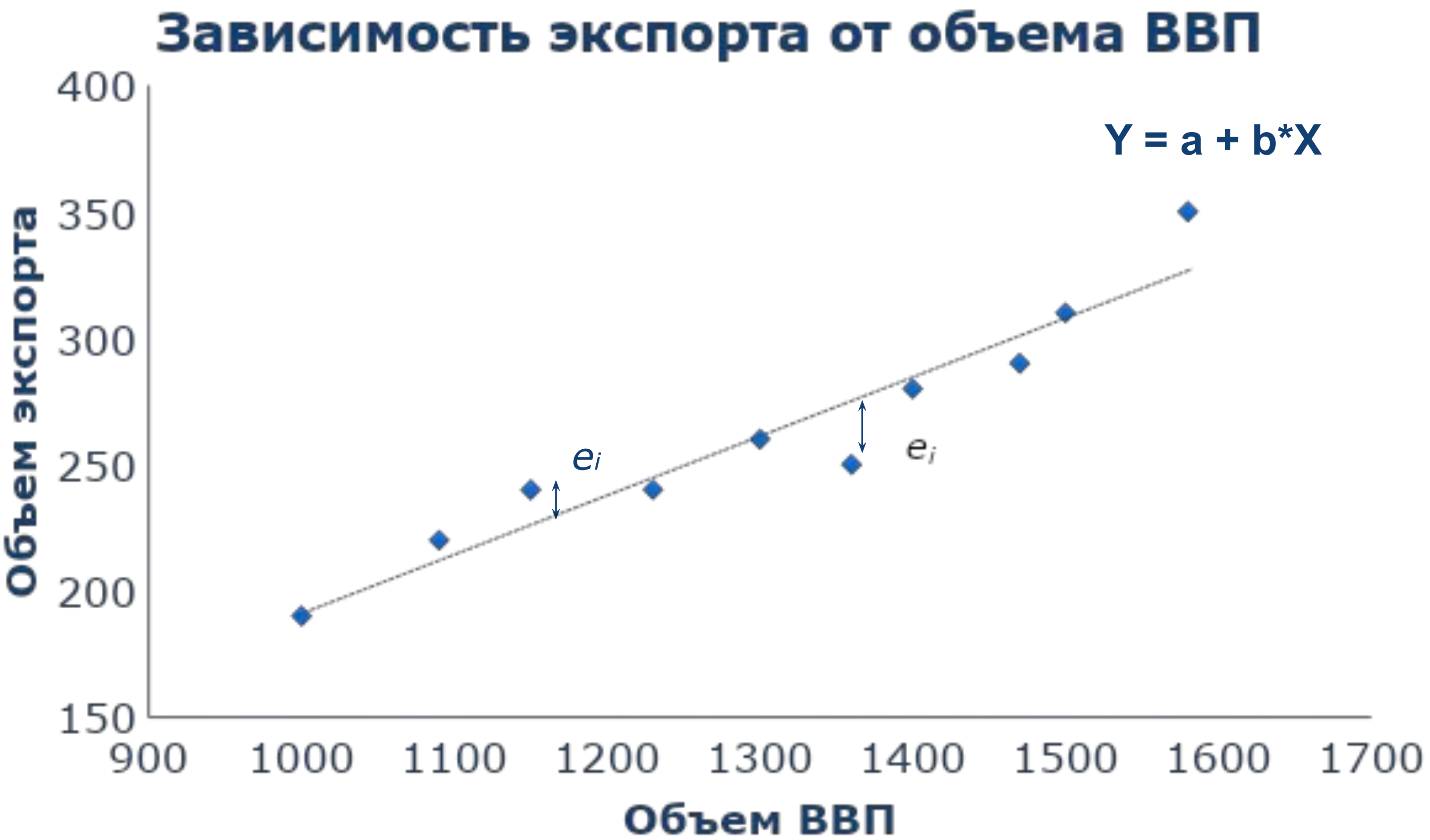

Зависимость экспорта от объема ВВП
Объем экспорта
0
50
100
150
200
250
300
350
400
900
1000
1100
1200
1300
1400
1500
1600
1700
Объем ВВП

График линейной зависимости



$$\sum e_i^2 = \sum (y_i - y_{предсказ})^2 \to \min$$

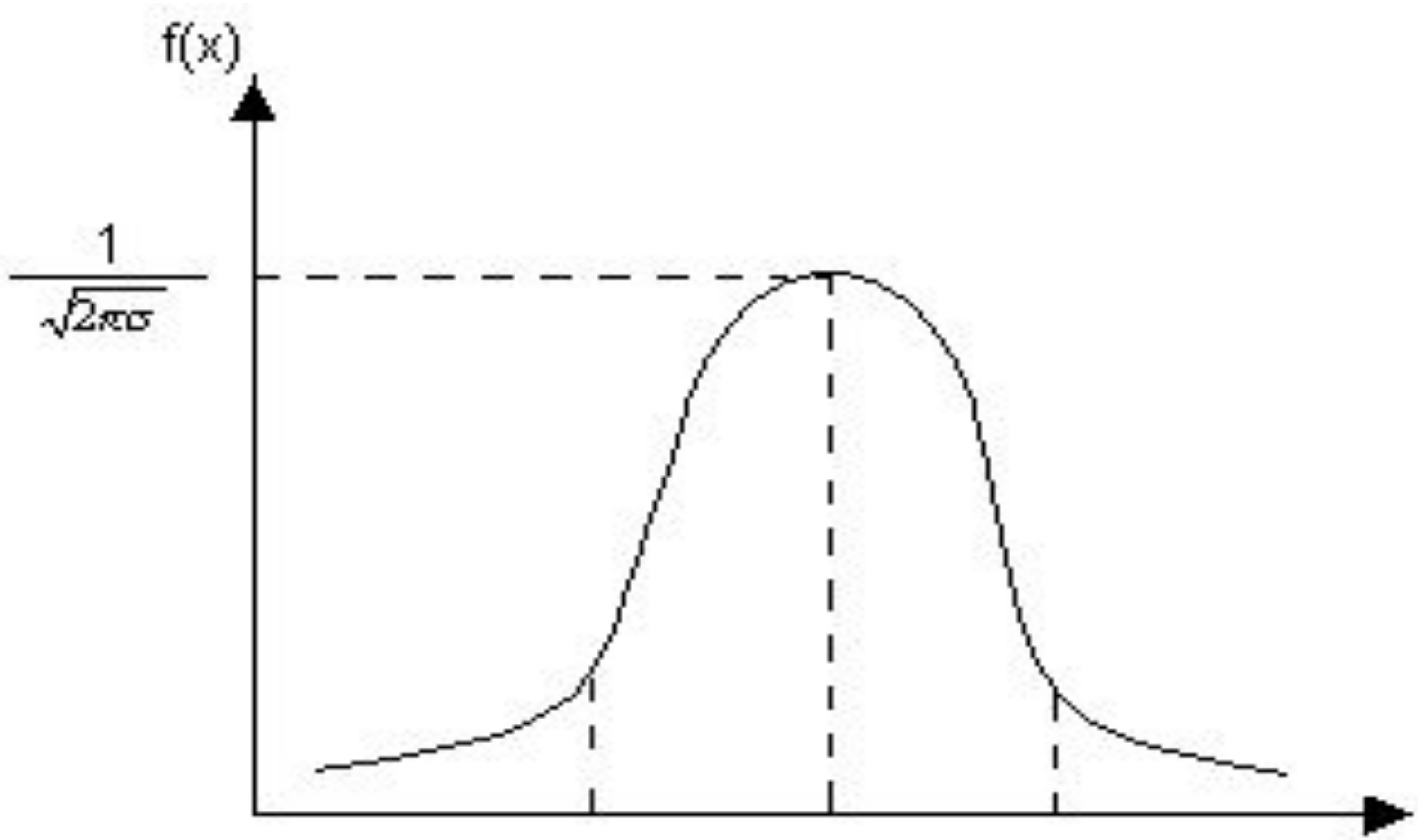
Формулы для нахождения параметров

- При помощи **метода наименьших квадратов (МНК)** выведены формулы для нахождения параметров уравнения (коэффициентов регрессии)

	Формула	Экономический смысл
b =	$b = \dfrac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}$	Показывает среднее изменение результата с изменением фактора на одну единицу
a =	$a = \bar{y} - b \cdot \bar{x}$	Значение y при x=0

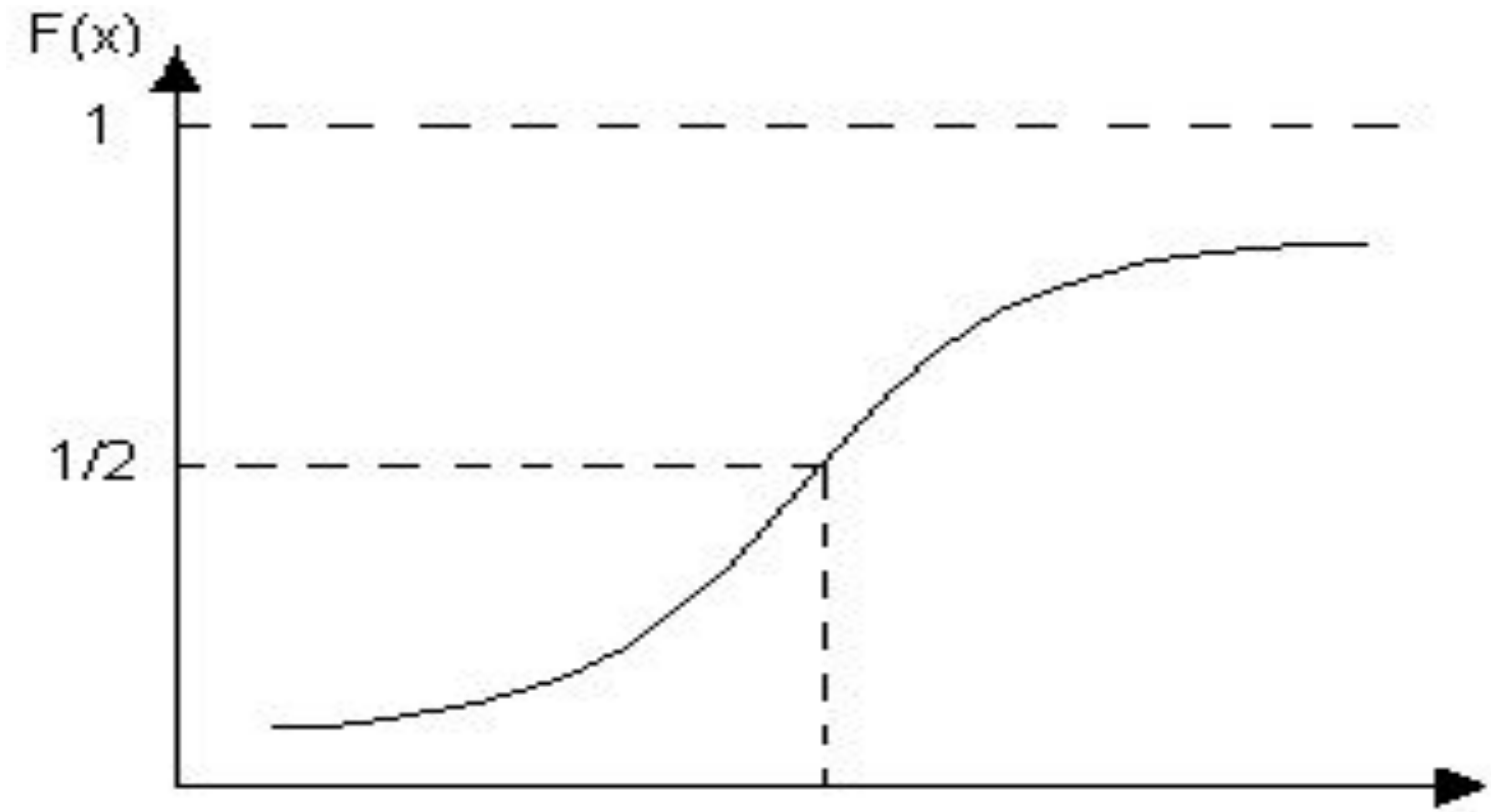
Требование нормального распределения остатков

График плотности вероятности нормального распределения



Требование нормального распределения остатков

Интегральная функция распределения нормальной СВ



Коэффициент корреляции

Его значения находятся в границах: -1 ≤ r ≤1.

***b* > 0, то r > 0, связь прямая;**

***b* < 0, то r < 0, связь обратная .**

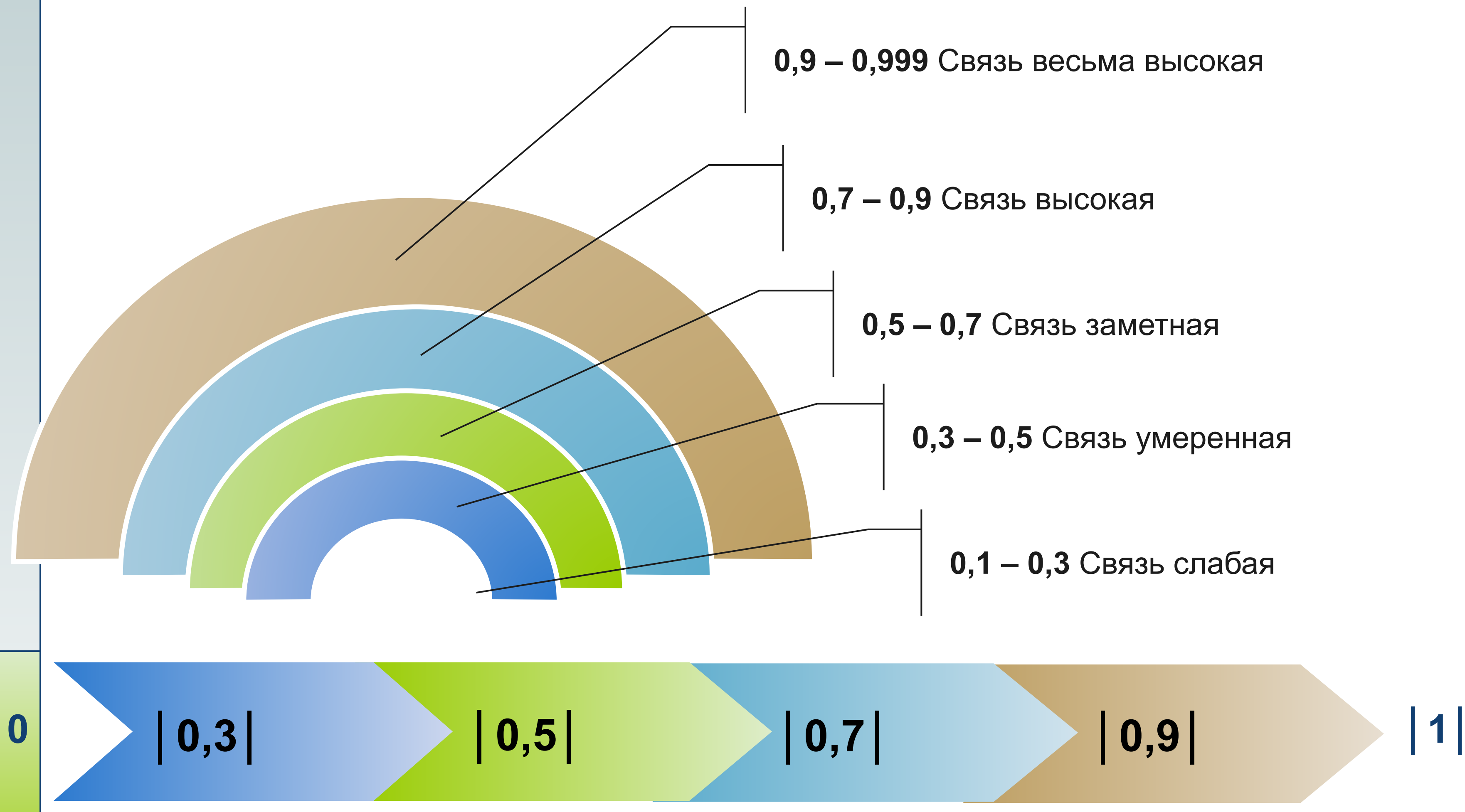
$$r = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y}$$

Показывает:

1.
Тесноту связи

2.
Направление связи

Шкала Чеддока для интерпретации коэффициента корреляции



Проверка качества подбора модели (вида уравнения)

TSS СумКО общая **=** **ESS** СумКО объясненная **+** **RSS** СумКО остаточная

$$\sum (y-\bar{y})^2 = \sum (\hat{y}-\bar{y})^2 + \sum (y-\hat{y})^2$$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$R^2 = r^2$$

$$0 \le R^2 \le 1$$

Коэффициент детерминации R^2- характеризует долю объясненной регрессией вариации *y*, в общей вариации результативного признака

Характеристики модели

Число степеней свободы - это число независимо варьируемых значений признака.
Дисперсия на одну степень свободы - получается делением каждой СКО на свое число степеней свободы

Факторная

$$D_{факт.} = \frac{\sum(\hat{y}_x - \bar{y})^2}{1}$$

k

Остаточная

$$D_{ост.} = \frac{\sum(y - \hat{y}_x)^2}{n-2}$$

n – k - 1

Общая

$$D_{общ.} = \frac{\sum(y - \bar{y})^2}{n-1}$$

n - 1

к – количество независимых переменных (для парной регрессии = 1)

Проверка статистической значимости модели

$$H_0 : D_{факт.} = D_{ост.} \quad H1 : D_{факт.} \rangle D_{ост.}$$

$$F_{набл} = \frac{D_{факт.}}{D_{ост.}}$$

Fраспобр(α;1;n-2)

если:

$$F_{набл.} < F_{табл.}$$

$$F_{набл.} \geq F_{табл.}$$

Модель статистически не значима

Модель статистическ и значима

Если справедлива H0, то дисперсии не отличаются друг от друга. Для H0 необходимо опровержение, чтобы факторная дисперсия превышала остаточную в несколько раз.

Проверка статистической значимости параметров

H0: b=0   H1: b≠0

$$t_{набл} = \frac{b}{m_b}$$

$t_{табл}$ = СТЬЮДРАСПОБР(α;n-2)

Стандартная ошибка параметра

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

$$m_b = \sqrt{\frac{D_{остат}^{\ \ 2}}{\sum (x - \bar{x})^2}}$$

если:

$$|t_{набл}| \langle t_{табл}$$

Параметр статистическ и не значим

$$|t_{набл}| \geq t_{табл}$$

Параметр статистическ и значим

Доверительный интервал для параметра

- параметр b с надежностью α лежит в интервале:

$$b_0 - m_b \cdot t_{табл}(\alpha; n-2) \le b \le b_0 + m_b \cdot t_{табл}(\alpha; n-2)$$

- значимость (α) = 1 – надежность(β)

Классические уровни α:

- **0,01**
- **0,05**
- **0,1**

Анализ данных

Регрессионная статистика	
Множественный R	0,719494
R-квадрат	0,517671
Нормированный R-квадрат	0,473823
Стандартная ошибка	53,13959
Наблюдения	13

Объясненная доля

Вероятность, с которой модель не значима

Дисперсионный анализ

	df	SS	MS	F	Значимость F
Регрессия	1	33338,02	33338,02	11,80602	0,005564
Остаток	11	31061,98	2823,816		
Итого	12	64400			

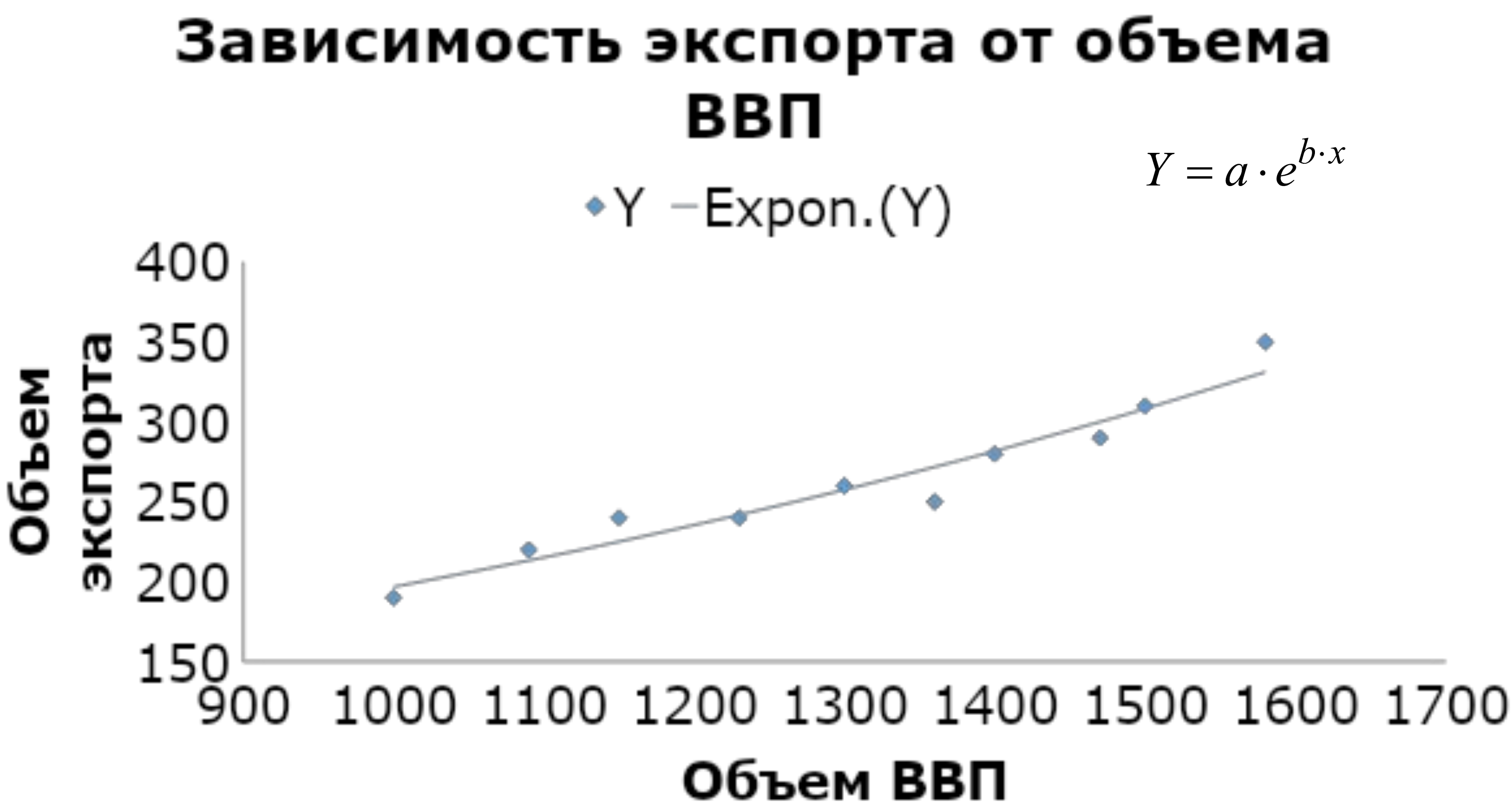
Вероятность, с которой не значим параметр

Y = -133,5 + 13,5*b

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	-133,542	78,10357	-1,7098	0,115324	-305,446	38,36313
X	13,17708	3,83502	3,435988	0,005564	4,736262	21,6179

Парная нелинейная регрессия

Экспоненциальная функция:



$$Y = a \cdot e^{b \cdot x}$$

График нелинейной зависимости

Логарифмическая функция

$$Y = a + b \cdot \ln(x)$$

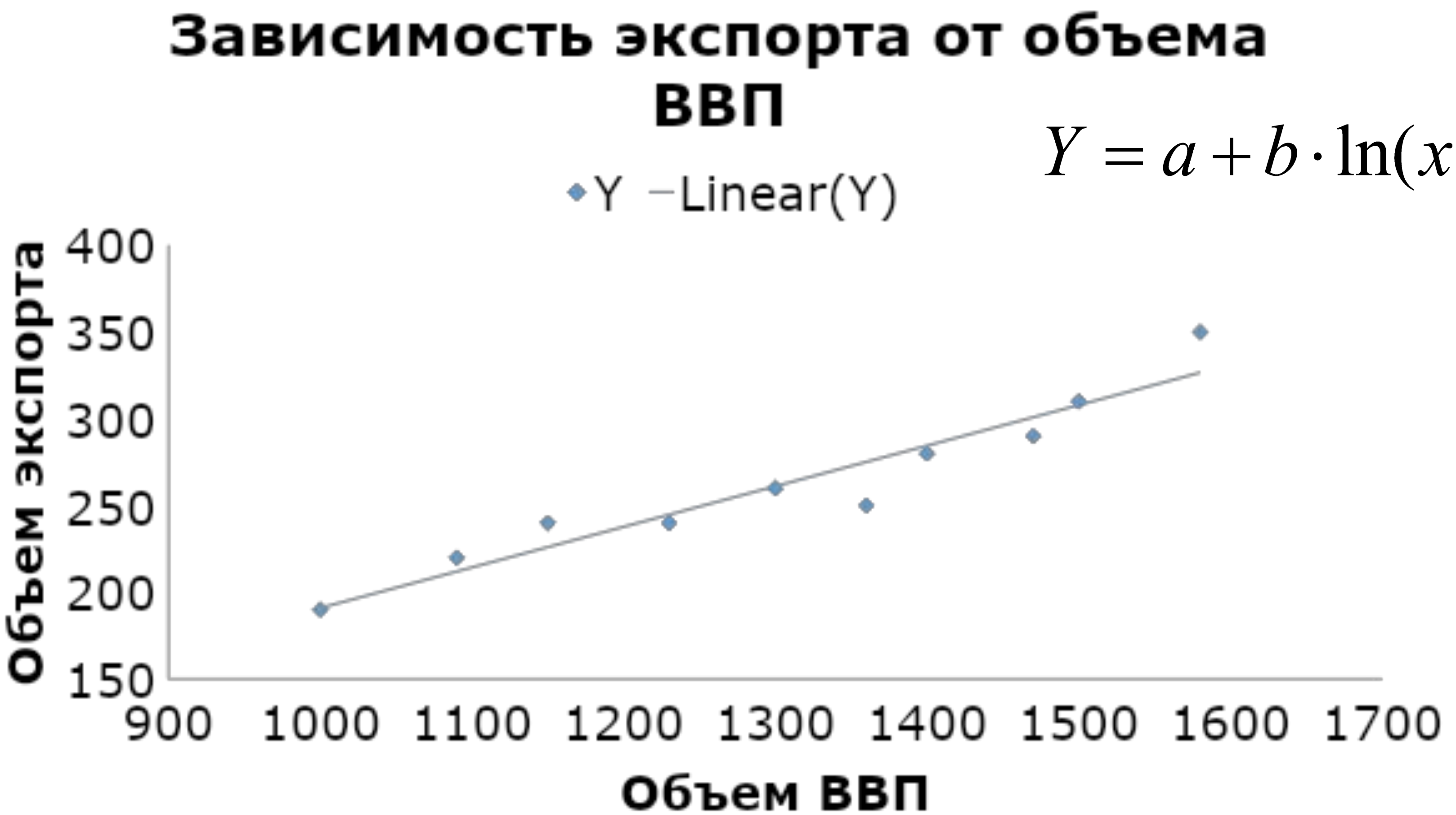
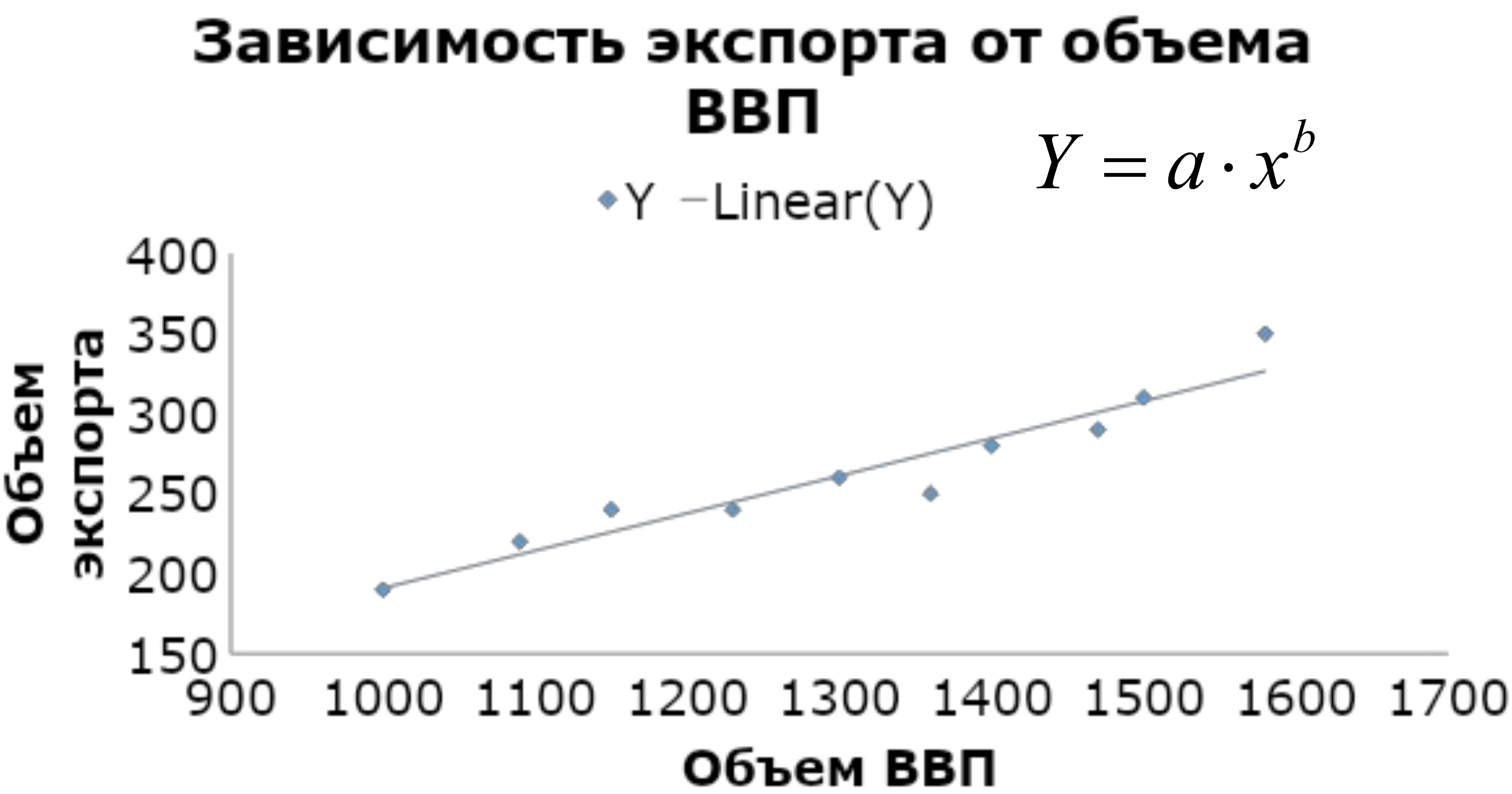


График нелинейной зависимости

Степенная функция

$$Y = a \cdot x^b$$



Эластичность

- Эластичность определяется по формуле:

$$Э_b = y' \cdot \frac{x}{y}$$

- Для парной регрессии: $$э_b = \frac{bx}{a + b \cdot x}$$

- Для степенной: $$э_b = b$$

Уравнение множественной регрессии

Линейная модель:

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p + e$$

- На любой экономический показатель чаще всего оказывает влияние не один, а несколько факторов.
- Например, спрос на некоторое благо определяется не только ценой данного блага, но и ценами на замещающие и дополняющие блага, доходом потребителей и многими другими факторами.

Уравнение множественной регрессии в стандартизированном масштабе

- Уравнение регрессии в стандартизованном масштабе имеет вид:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_p t_{x_p} + \varepsilon$$

- стандартизированные параметры:

$$t_y = \frac{y - \bar{y}}{\sigma_y}; \quad t_{x_j} = \frac{x_j - \bar{x}_j}{\sigma_{x_j}}, \quad j = \overline{1, n}$$

- Система уравнений:

$$\begin{cases} \beta_1 + \beta_2 r_{x_2 x_1} + \beta_3 r_{x_3 x_1} + \beta_p r_{x_p x_1} = r_{yx_1} \\ \beta_1 r_{x_1 x_2} + \beta_2 + \beta_3 r_{x_3 x_2} + \beta_p r_{x_p x_2} = r_{yx_2} \\ \dots \\ \beta_1 r_{x_1 x_p} + \beta_2 r_{x_2 x_p} + \beta_3 r_{x_3 x_p} + \beta_p = r_{yx_p} \end{cases}$$

Проверка качества уравнения множественной регрессии

- Коэффициент детерминации:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}$$

- Скорректированный к-т детерминации:

$$\bar{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-p-1} = R^2 - \frac{p}{n-p-1}\left(1 - R^2\right)$$

Проверка статистической значимости модели

$$H_0 : R^2 = 0 ; H1 : R^2 \rangle 0$$

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k-1}{k}$$

Fраспобр(α;k;n-k-1)

если:

$$F_{набл.} < F_{табл.}$$

$$F_{набл.} \geq F_{табл.}$$

Модель статистически не значима

Модель статистическ и значима

Если справедлива Н0, то дисперсии не отличаются друг от друга. Для Н0 необходимо опровержение, чтобы факторная дисперсия превышала остаточную в несколько раз.