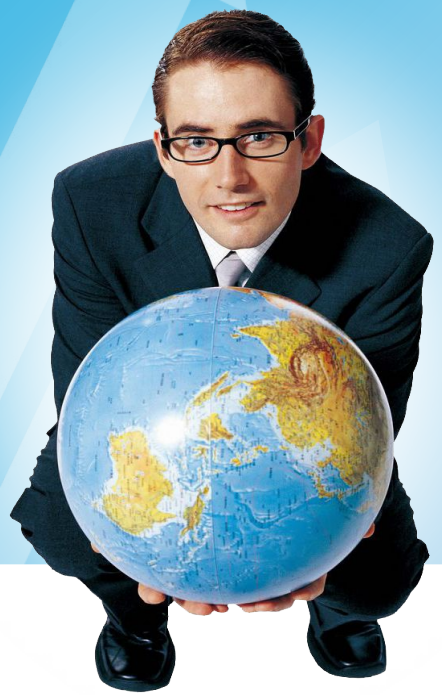
Курс Эконометрика

Хасанова
Светлана Фанилевна



www.themegallery.com

Темы занятий

- Парная регрессия
- Множественная регрессия
- Временные ряды
- Системы одновременных уравнений
- Проверка статистических гипотез

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

Для 13 клиентов спортивного отдела магазина зафиксирована сумма покупки (в у.е.) и время разговора с продавцом (мин):

Y	40	50	60	80	100	110	120	130	150	160	180	200	310
X	14	14	17	19	17	20	24	22	25	24	18	20	26

Построим график с осями Y – сумма покупки и X – время разговора

Каждая точка графика соответствует конкретному клиенту – всего 13 точек.

График

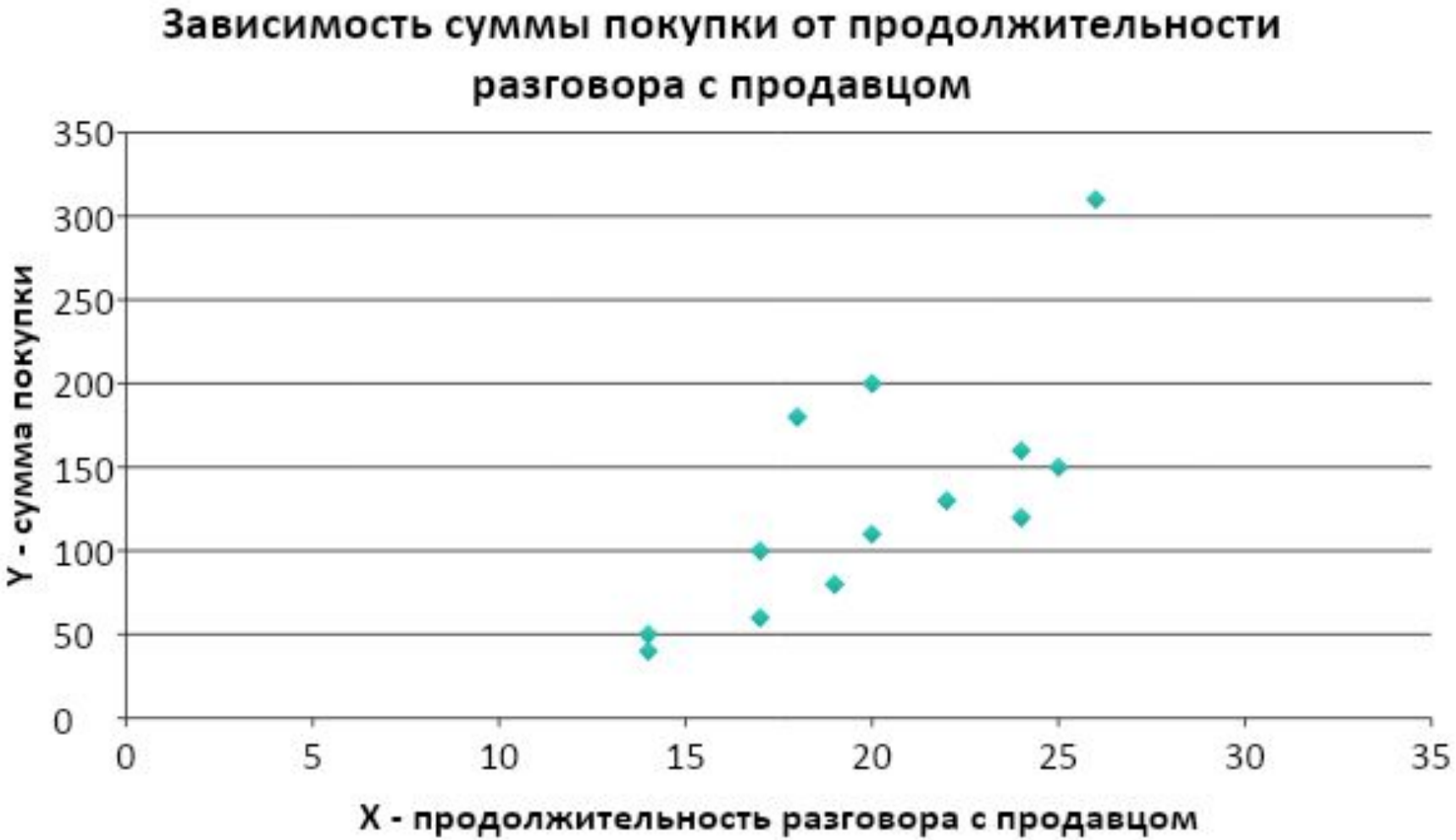
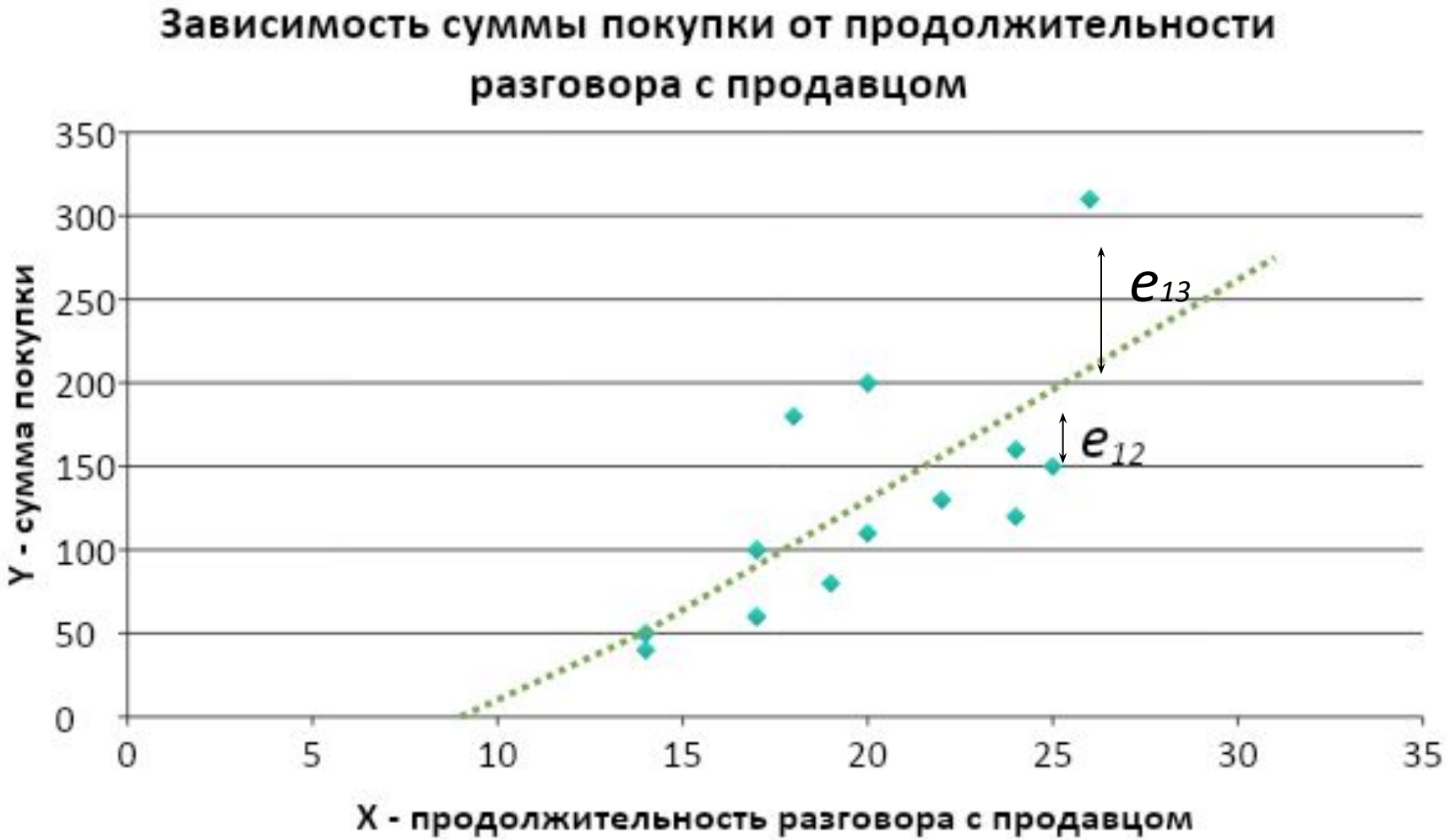

Зависимость суммы покупки от продолжительности разговора с продавцом
Y - сумма покупки
0
50
100
150
200
250
300
350
0
5
10
15
20
25
30
35
X - продолжительность разговора с продавцом

График линейной зависимости

$Y = a + b*X$



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

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

Коэффициент корреляции

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

***b* > 0, то r > 0, связь прямая;**

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

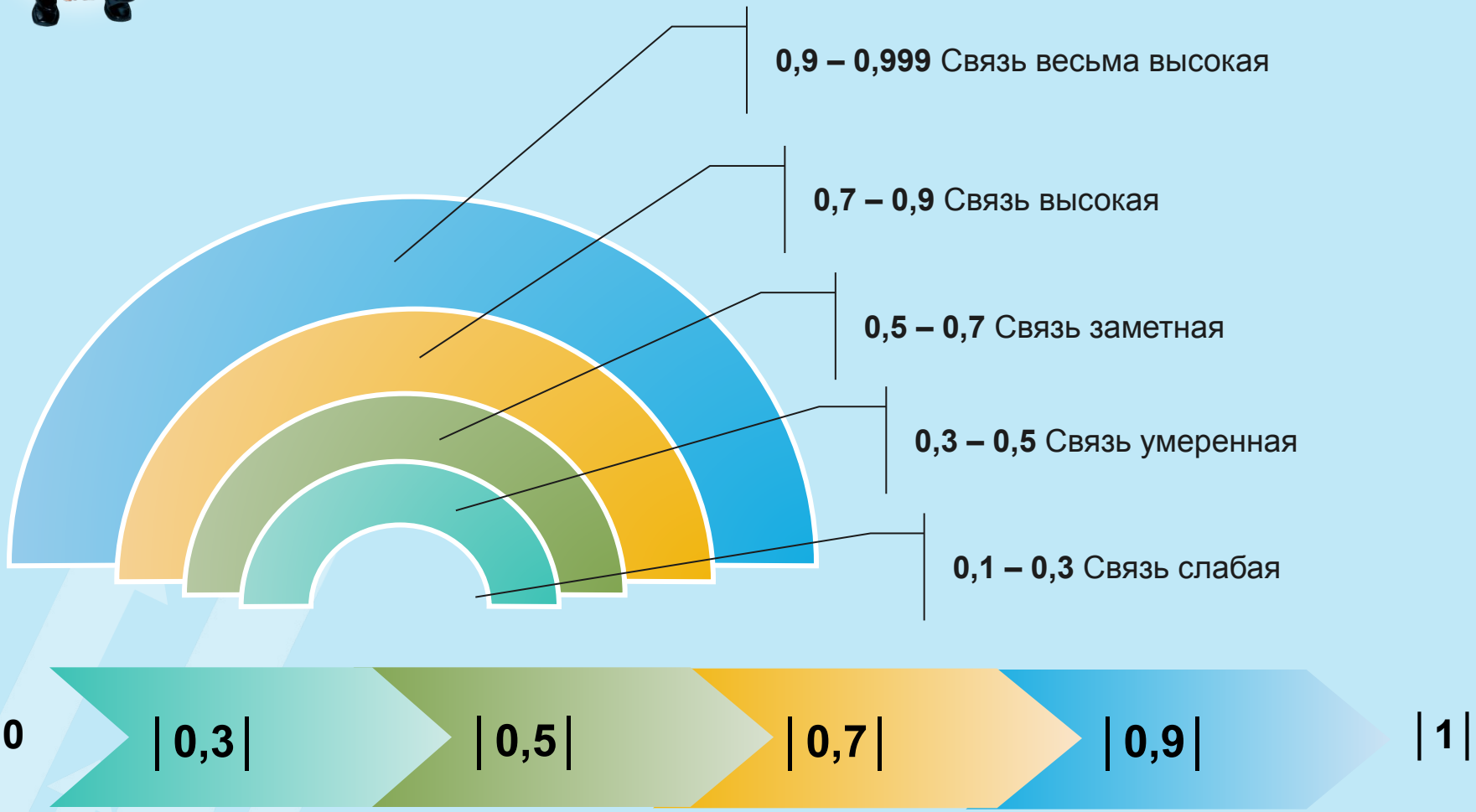
$$r = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y}$$

Показывает:

1. Тесноту связи

2. Направление связи

Шкала Чеддока для интерпретации коэффициента корреляции

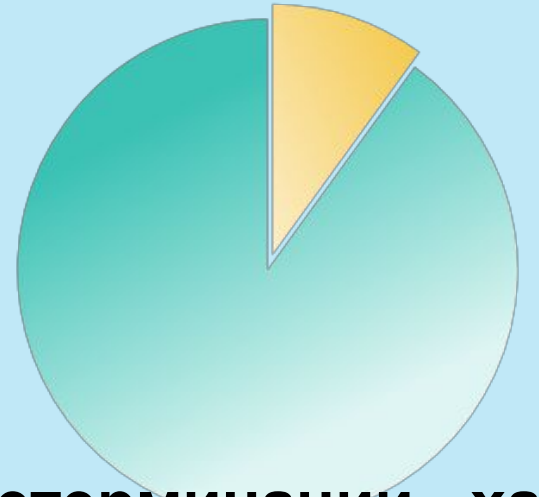


Проверка качества подбора модели (вида уравнения)

δ^2 общая = δ^2 факторная + δ^2 остаточная

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

$$R^2 = \frac{\sigma^2_{объясн.}}{\sigma^2_{y\,общ.}}$$



$$R^2 = r^2$$

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

Коэффициент детерминации - характеризует долю дисперсии результативного признака *y*, объясняемую построенной регрессией, в общей дисперсии 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	Остаточная	Общая
$D_{факт.} = \frac{\sum(\hat{y}_x - \bar{y})^2}{1}$	$D_{ост.} = \frac{\sum(y - \hat{y}_x)^2}{n-2}$	$D_{общ.} = \frac{\sum(y - \bar{y})^2}{n-1}$
k	**n – k - 1**	**n - 1**

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

Проверка статистической значимости модели

$$H_0 : D_{факт.} = D_{ост.}$$

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

Fраспобр(α;1;n-2)

если :

$F_{набл.} < F_{табл.}$

Модель статистически не значима

$F_{набл.} \geq F_{табл.}$

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

$t_{набл} < t_{табл}$:

Параметр статистически не значим

$t_{набл} \geq t_{табл}$

Параметр статистически значим

Доверительный интервал для параметра

- параметр b с надежностью α лежит в интервале:

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

- надежность (α) = 1 – вероятность (β)

Анализ данных

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

Объясненная доля

Вероятность, с которой модель не значима

Вероятность, с которой не значим параметр

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

Парная нелинейная регрессия

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

Экспоненциальная функция:

График нелинейной зависимости

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

Логарифмическая функция

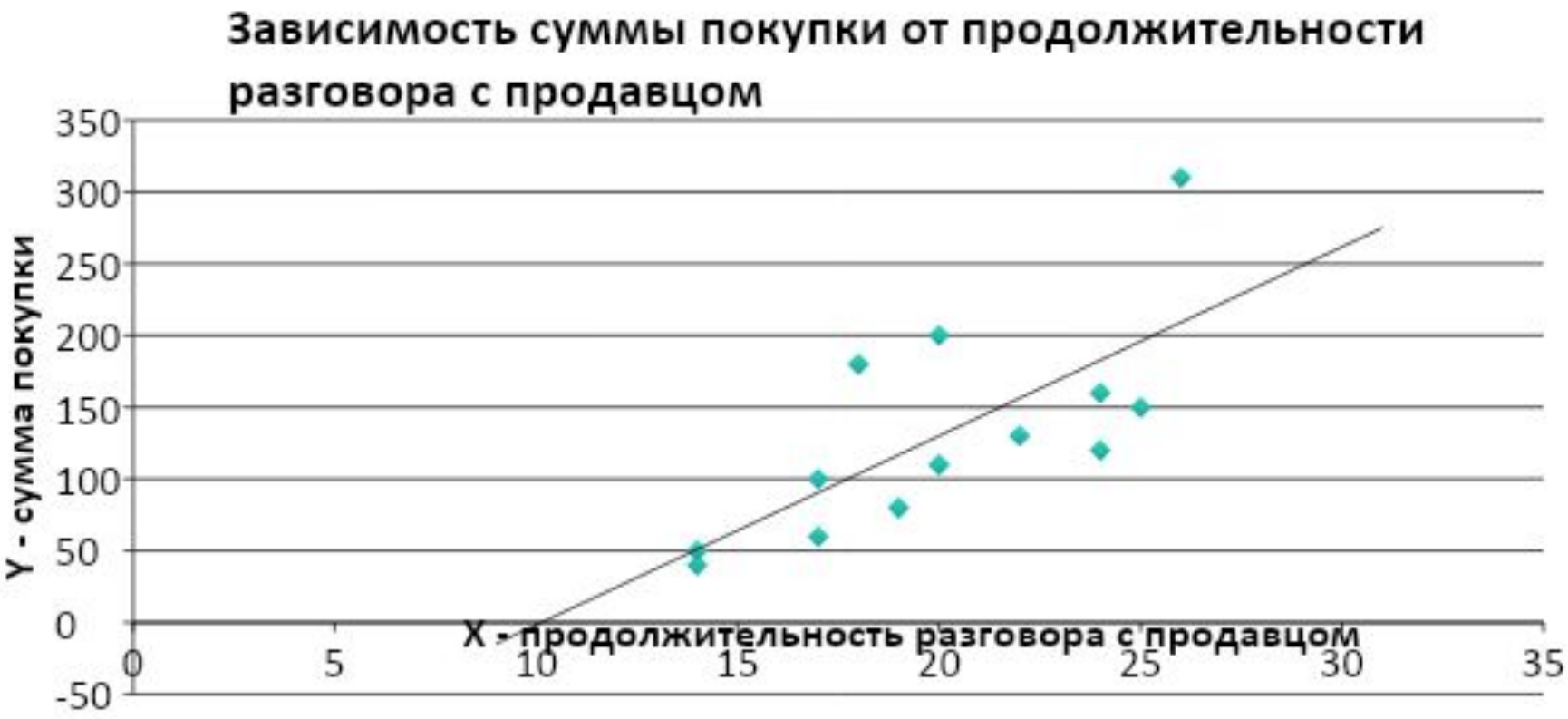
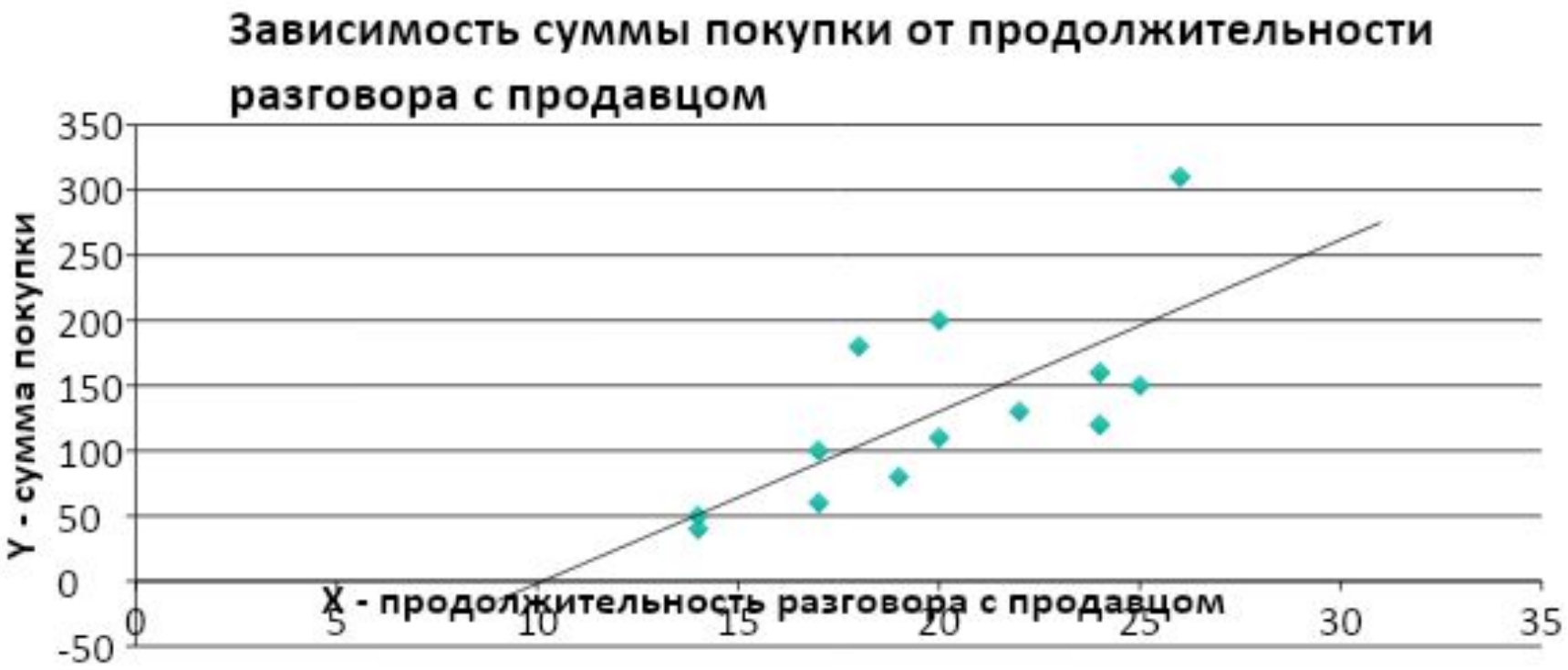


График нелинейной зависимости

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

Степенная функция



Эластичность

- Эластичность определяется по формуле:

$$Э_b = \frac{y'}{y}$$

- Для парной регрессии:

$$э_b = \frac{b}{a + b \cdot x}$$

- Для степенной:

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

$$э_b = b$$

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:

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

- 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:

$$\begin{cases} \beta_1 + \beta_2 r_{x_2 x_1} + \beta_3 r_{x_3 x_1} + \beta_p r_{x_p x_1} = r_{y x_1} \\ \beta_1 r_{x_1 x_2} + \beta_2 + \beta_3 r_{x_3 x_2} + \beta_p r_{x_p x_2} = r_{y x_2} \\ \square \quad \square \quad \square \quad \square \quad \square \\ \beta_1 r_{x_1 x_p} + \beta_2 r_{x_2 x_p} + \beta_3 r_{x_3 x_p} + \beta_p = r_{y x_p} \end{cases}$$

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:

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} = R^2 - \frac{p}{n-p-1}(1 - R^2)$$

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

Спасибо за внимание!

До свидания!

www.themegallery.com