Тема лекции и практики:

«ПОСТРОЕНИЕ ПАРНЫХ ЛИНЕЙНЫХ МОДЕЛЕЙ (парных линейных уравнений регрессии)»

ВОПРОСЫ ТЕМЫ:

1.Понятие математической модели и аппроксимации. Виды математических зависимостей и моделей. Парные и множественные математические модели данных: сходства и отличия.

2.Этапы построения парной математической модели данных.

3.Графическая интерпретация и свойства различных видов парных моделей.

4.Построение парной линейной модели: метод наименьших квадратов, теоретические аспекты, возможности MExcel и SPSS , пример.

1 Понятие математической модели, состав и целевое назначение

Системное отражение оригинала

Модель – это?

Способ существования знаний

Упрощенное и осмысленное представление об объекте

МОДЕЛЬ

Абстрактная

это образные модели: чертежи, схемы, геометрические фигуры, карты, знаковые (формулы, уравнения, зависимости, функции)

Материальная

это предметы, которые воспроизводят существенные (геометрические, физические) свойства предметов: макеты, муляжи

Математическая модель – это абстрактная, знаковая модель, отражающая количественную **зависимость** одной переменной от другой (одного явления от другого)

СОСТАВ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Переменная величина, которая влияет (фактор)
X

В мат. модели устанавливается зависимость Y от X в виде арифметических действий, в которых участвует X

Переменная величина, которая испытывает влияние (результат)
Y

$$Y = 15X + 25X^2 + 8$$

Математическая модель = зависимость
(изменение X приводит к изменению Y)

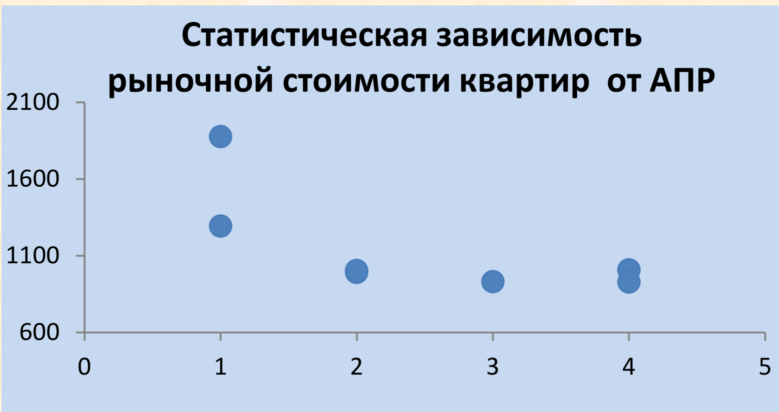
Виды зависимостей

Виды зависимостей:

- не строгая, вероятностная, неполная (корреляционная)
- функциональная (строгая)

Для построения мат. модели необходимы статистические данные переменных, включаемых в модель, т.е. выборки факторной величины и результативной. Выборки могут представлять собой изменение переменной величины в пространстве и во времени.

Пример не строгой зависимости



Пример статической выборки

Квартира	Стоимость (Y)	АПР (X)
1	Y_1	X_1
2	Y_2	X_2
....	....	....

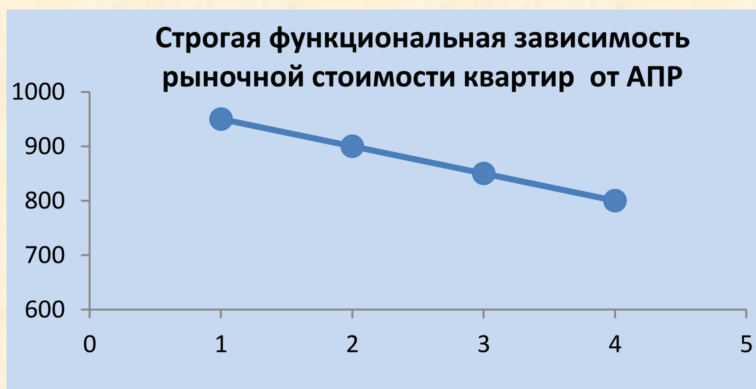
Пример динамической выборки

Год	Стоимость (Y)	Износ (X)
2000	Y_1	X_1
2001	Y_2	X_2
....	....	.....

Точка данных

Свойства не строгой зависимости (корреляционной):

1) точки данных не соединяются в правильную линию, а расположены около некоторой правильной линии;
2) одному и тому же фактическому значению фактора (или значениям нескольких факторов) могут соответствовать разные фактические значения Y;
3) если фактор изменяется, то изменение Y происходит, но без строгой закономерности;
4) Может быть найдена только наиболее близкая функция

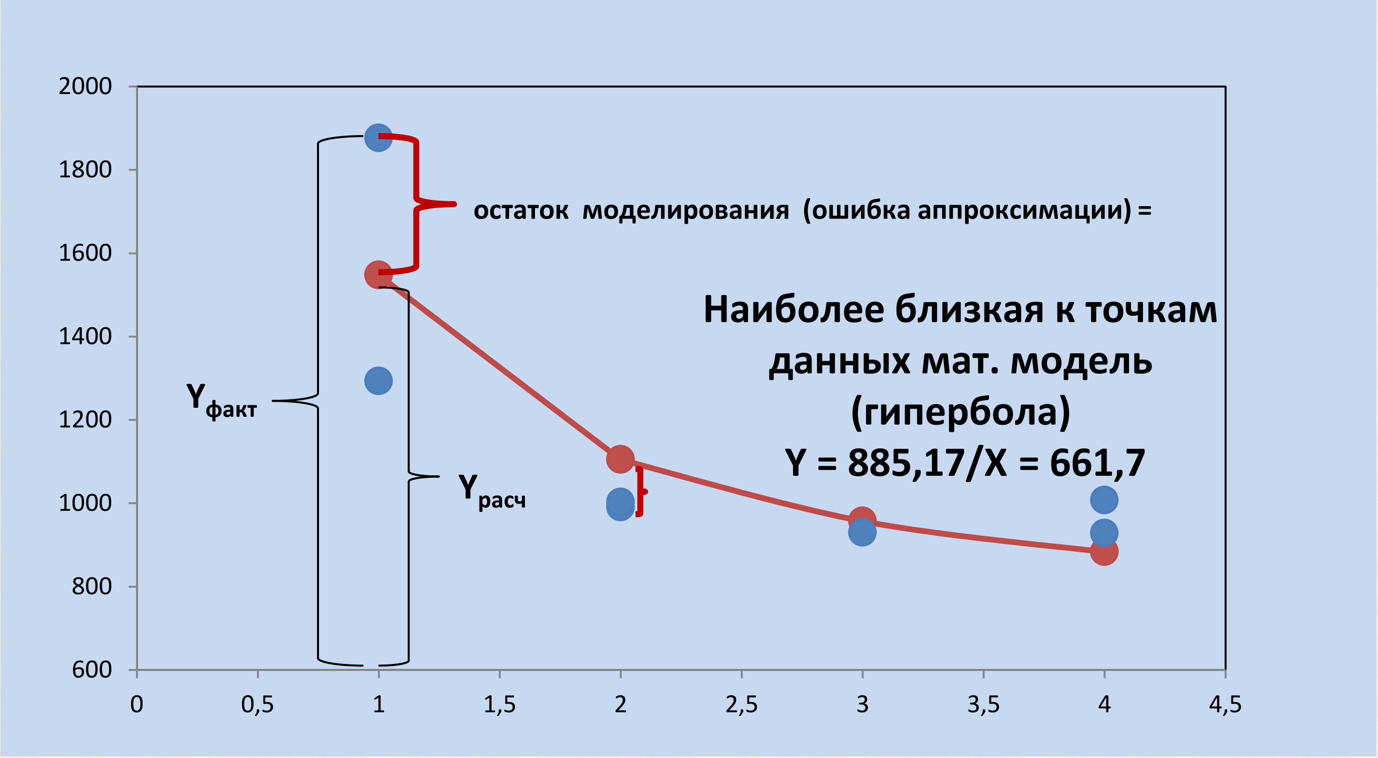


Пример строгой функциональной зависимости

Свойства строгой функциональной зависимости:

1) точки данных соединяются в правильную линию (прямую или кривую);
2) если фактор изменяется, то изменяется и Y, и в этом изменении наблюдается закономерность;
3) функция может быть определена, т.е. зависимость может быть выражена количественно

Понятие аппроксимации и остатка в точке данных



Математическая модель – это функциональная зависимость **! ! ! !**

Цель моделирования: если зависимость не строгая, найти наиболее близкую мат. модель (функциональную) к точкам данных.

Аппроксимация – замена не строгой зависимости <u>наиболее близкой мат. моделью.</u> В результате аппроксимации в точке формируется <u>остаток</u> (разность между фактическим и расчетным значением Y)

Виды математических моделей (стандартных функций)

Математическая модель (стандартная функция)

ПАРНАЯ (один фактор) Y = f(X)	МНОЖЕСТВЕННАЯ (много факторов) Y=f(X1, X2, X3, ……, Xn)
Функция одной переменной	**Функция нескольких переменных**

Вид модели	Парная модель	Множественная модель
линейная	$y = ax+b$	$y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b$
гиперболическая	$y = a/x+b$	$y = a_1/x_1 + a_2/x_2 + \dots + a_n/x_n + b$
полином 2 степени	$y = ax^2+bx+c$	$y = a_1x_1^2 + b_1x_1 + a_2x_2^2 + b_2x_2 + c$
полином 3 степени	$y = ax^3+bx^2+cx+d$	$y = a_1x_1^3 + b_1x_1^2 + c_1x_1 + a_2x_2^3+b_2x_2^2+c_2x_2+d$
степенная	$y = ax^b$	$y = ax_1^{b1}x_2^{b2}\dots x_n^{bn}$
показательная	$y = ab^x$	$y = ab_1^{x1}b_2^{x2}\dots b_n^{xn}$
обратная	$y = 1/(ax+b)$	$y = 1/(a_1x_1+a_2x_2+\dots+a_nx_n+b)$
экспоненциальная 1	$y = ae^{bx}$	$y = ae^{b1x1}e^{b2x2}\dots..$
экспоненциальная 2	$y = e^{ax+b}$	$E^{a1x1\ +\ a2x2\ \ +\dots..\ +\ anxn\ +\ b}$
логарифмическая	$y = a\mathrm{Ln}(x)+b$	$y = a_1\mathrm{Ln}(x_1) + a_2\mathrm{Ln}(x_2) + \dots.. + b$

Примеры математических моделей (не стандартных функций)

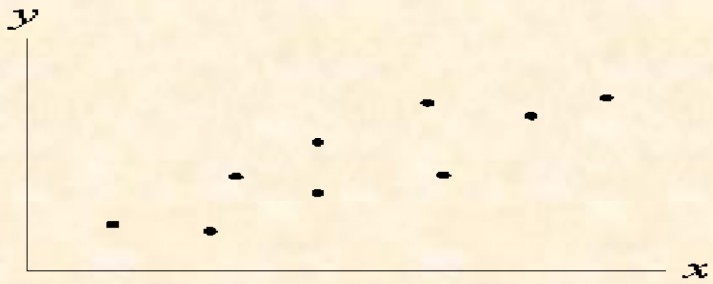
Не строгая функциональная модель чаще всего может представлять собой <u>смешение различных строгих функциональных моделей</u>

Параметры модели, которые должны быть определены в процессе моделирования

Элементы модели взяты из	Пример
Гиперболической и экспоненциальной	$y = a/e^{bx} + c$
Гиперболической и степенной	$y = a/x^{b} + c$
Гиперболической и показательной	$y = a/b^{x} + c$
Гиперболической и логарифмической	$y = a/\ln(x) + b$
Линейной и степенной	$y = ax^{b} + c$
Линейной и показательной	$y = ab^{x} + c$
……………….. Вариантов сочетаний может быть множество !!!!!!!!!!!	

Этапы построения (поиска) стандартных парных функций

1 Этап: Определяются переменные величины Y и X
2 Этап: Формируется статистическая выборка для переменных X и Y.
3 Этап: Строится поле корреляции (зависимости).

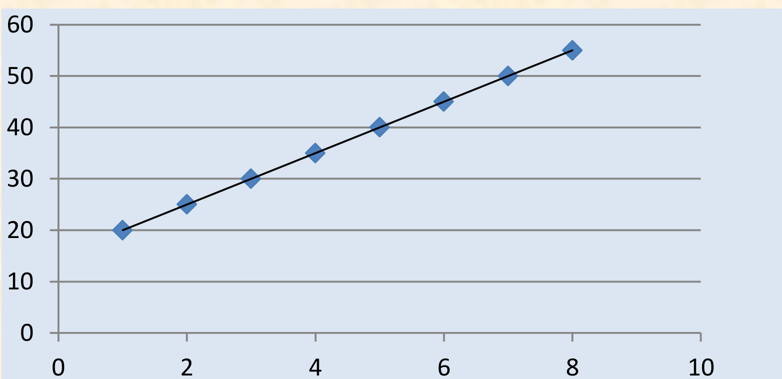


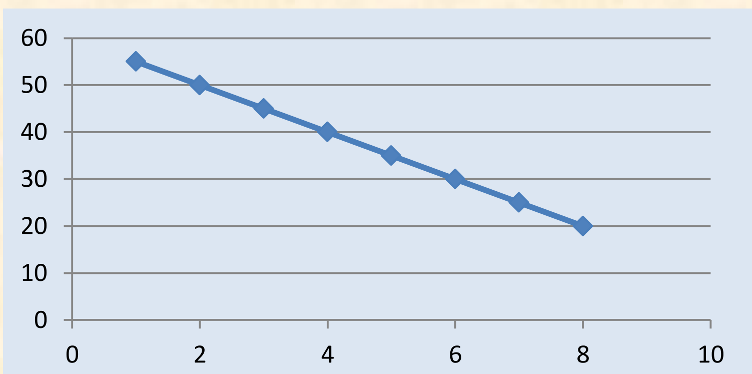
4 Этап: Визуально по характеру распределения точек данных выдвигается гипотеза с помощью какой стандартной парной функции целесообразно аппроксимировать не строгую зависимость (гипотез может быть несколько)!!!!!!!!!
5 Этап: Записывается выбранная функция в общем виде и определяются ее параметры методом наименьших квадратов или методом центральных точек. Если выбранная функция – не стандартная, то метод нахождения ее параметров – это метод итераций.
6 Этап: Рассчитанные параметры записываются в общий вид модели.
7 Этап: Рассчитываются показатели качества модели и делается вывод о возможности ее применения для прогноза Y.

Графическая интерпретация и свойства стандартной линейной модели

ЛИНЕЙНАЯ МОДЕЛЬ: y = ax + b

Знак а	Производная	Знак производной в точке	Изменение производной в точке	График модели	Направление зависимости	Степень зависимости Y от X
a > o	y' = a	+	const	положительный одинаковый наклон	прямая	не изменяется
a < o	y' = - a	-	const	отрицательный одинаковый наклон наклон	обратная	не изменяется





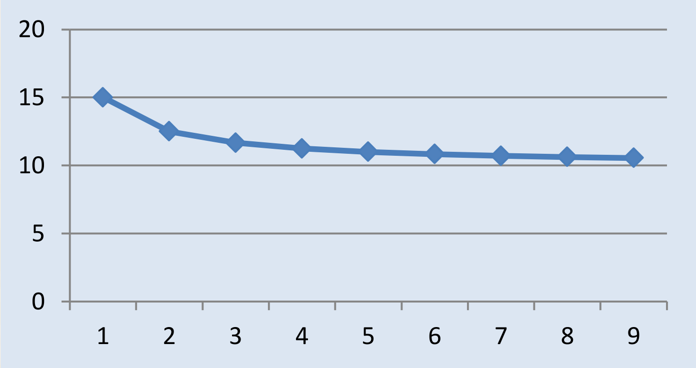
Схематично график модели: y = ax + b

Схематично график модели: y = - ax + b

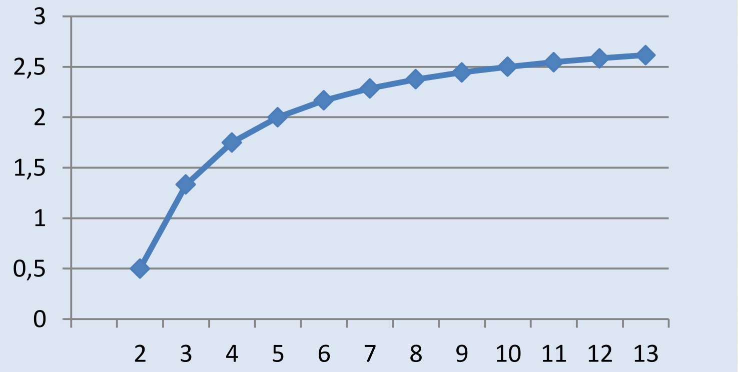
Графическая интерпретация и свойства стандартной гиперболической модели

ГИПЕРБОЛИЧЕСКАЯ МОДЕЛЬ: y = a/x + b

Знак a	Производная	Знак производной в точке	Изменение производной в точке	График модели	Направление зависимости	Степень зависимости Y от X
a > 0	$y' = -ax^{-2}$	-	снижается при возрастании X	отрицательный наклон с убывающим углом наклона	обратная	при возрастании X его влияние на Y снижается
a < 0	$y' = ax^{-2}$	+	снижается при возрастании X	положительный наклон с убывающим углом наклона	прямая	при возрастании X его влияние на Y снижается



Схематично график модели: y = a/x + b



Схематично график модели: y = - a/x + b

Графическая интерпретация и свойства стандартной степенной модели

СТЕПЕННАЯ МОДЕЛЬ: $y = ax^b$

Знак a	Знак b	Производная	Знак производной в точке	Изменение производной в точке	График модели	Направление зависимости	Степень зависимости Y от X
a > 0	b > 1	$y' = abx^{b-1}$	+	увеличивается при росте X	положительный наклон с ростом угла наклона	прямая	при росте X снижается
a > 0	b > 0	$y' = abx^{b-1}$	+	снижается при росте X	положительный с убыванием угла наклона	прямая	при росте X снижается
a > 0	b < 0	$y' = -bax^{-b-1}$	-	снижается при росте X	отрицательный с убыванием угла наклона	обратная	при росте X снижается

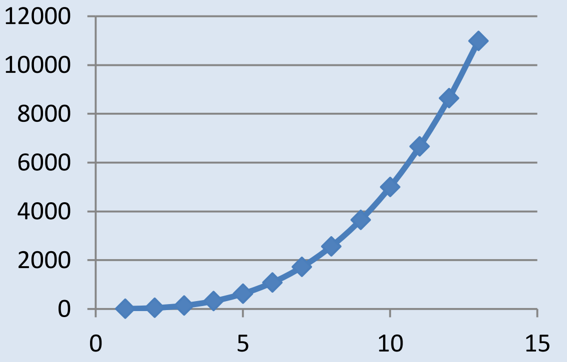


график модели: $y = ax^b$ (b>1)

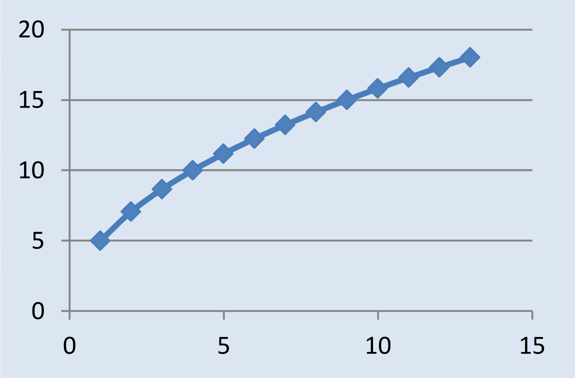


график модели: $y = ax^b$ (b>0)

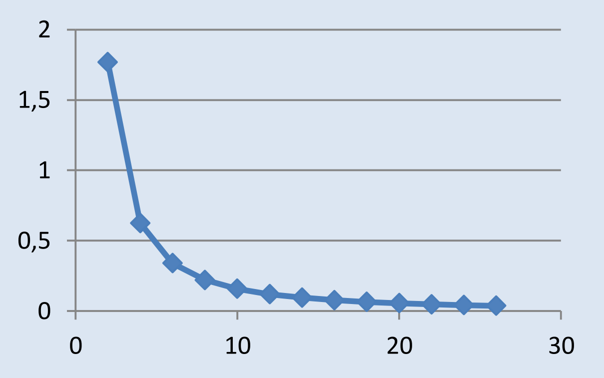


график модели: $y = ax^{-b}$ (b<0)

Графическая интерпретация и свойства стандартной показательной модели

ПОКАЗАТЕЛЬНАЯ МОДЕЛЬ: $y = ab^x$

Знак a	Знак b	Производная	Знак производной в точке	Изменение производной в точке	График модели	Направление зависимости	Степень зависимости Y от X
a > 0	b > 1	$y' = ab^x \ln(b)$	+	увеличивается при росте X в равное количество раз. Коэффициент роста = b	положительный наклон с ростом угла наклона в b раз	прямая	при росте X увеличивается
a > 0	b > 0	$y' = ab^x \ln(b)$	-	снижается при росте X в равное количество раз. Коэффициент снижения = b	отрицательный наклон с убыванием угла наклона в b раз	обратная	при росте X снижается

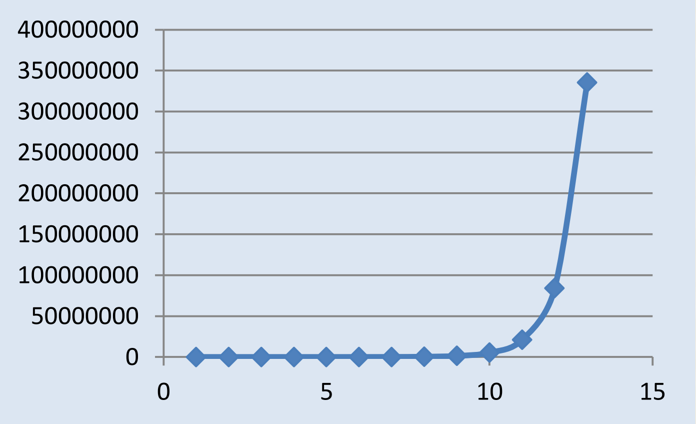


график модели: $y = ab^x$ (b>1)

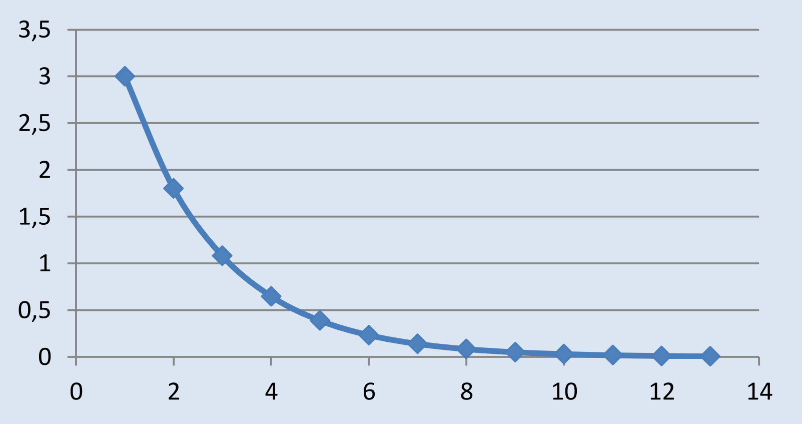
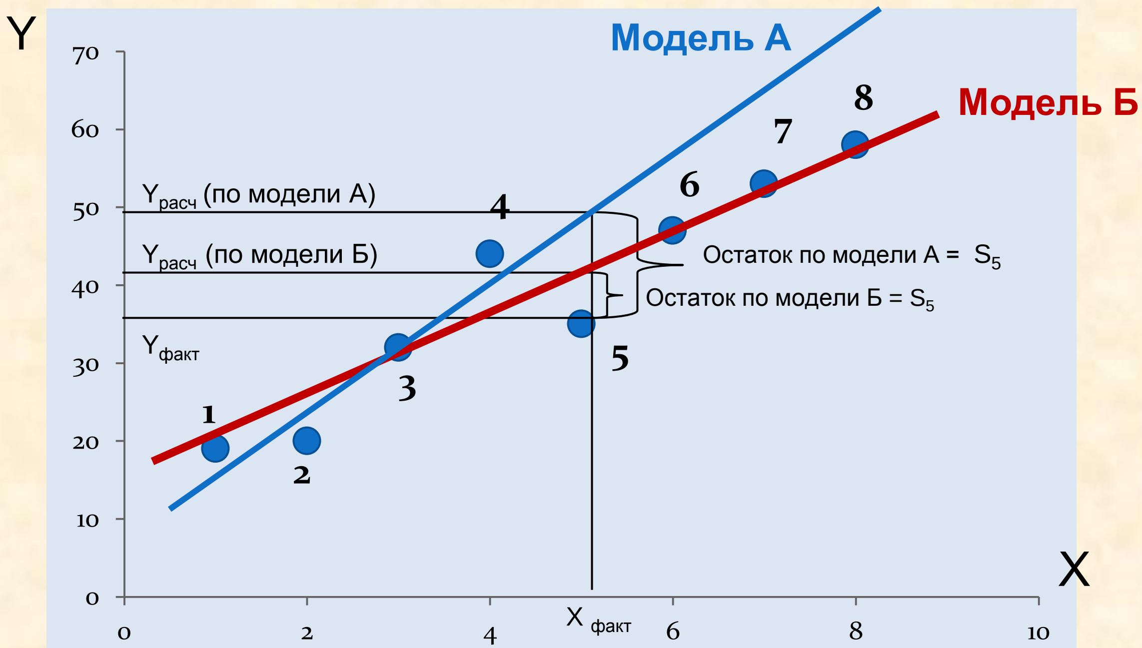


график модели: $y = ab^x$ (b>0)

Построение стандартной парной линейной модели. Метод наименьших квадратов

1 Этап: Определяются переменные величины Y и X.
2 Этап: Формируется статистическая выборка для переменных X и Y (точки данных).
3 Этап: Строится **поле корреляции** (зависимости)



4 Этап: Визуально определяется характер распределения точек. Если точки стремятся к графику линейной модели, то зависимость может быть аппроксимирована парной линейной моделью.
5 Этап: Записывается линейная модель в общем виде: **y = ax + b.** Далее необходимо определить параметры a и b. Сочетаний параметров – бесконечно. **Необходимо определить сочетание, при котором линейная модель была бы максимально приближена к точкам данных, т.е. необходимо найти модель с** минимально возможными остатками.

Условие максимальной близости модели: **S = $\sum S_i^2$ - минимальна !!!!!!!**

Метод наименьших квадратов (базируется на мат. анализе)

Построение стандартной парной линейной модели. Метод наименьших квадратов

S = $\sum \boldsymbol{S_i}^2 = \sum(\boldsymbol{y}_{\text{расчт}\, \boldsymbol{i}} - \boldsymbol{y}_{\text{фактич}\, \boldsymbol{i}})^2 = \sum(\boldsymbol{a}\boldsymbol{x}_{\boldsymbol{i}} + \boldsymbol{b} - \boldsymbol{y}_{\text{фактич}\, \boldsymbol{i}})^2$

$\boldsymbol{S} =$ **f (a, b) – сумма квадратов остатков есть функция от параметров модели**

Цель: Определить параметры A и B при которых S минимальна (задача из мат. анализа на определение минимума функции)

Свойство критической точки:

«Производная в критической точке равна нулю» !!!!!!

Последовательность нахождения параметров A и B (координаты критической точки):

1) определяется S': S'_a; S'_b
2) S' = 0 (свойство критической точки): S'_a (b = const) = 0; S'_b (a = const) = 0
3) Из полученных уравнений формируется система уравнений с двумя неизвестными параметрами A и B (координатами критической точки)
4) Определяется: является ли критическая точка минимумом или максимумом!!!!!
5) Если критическая точка – минимум, то полученные значения и есть параметры линейной модели, которая ближе остальных расположена к точкам данных.

Построение стандартной парной линейной модели. Метод наименьших квадратов

Упростим условие: **«Предположим, что у нас всего две точки данных»**

$$\mathbf{S} = \sum \boldsymbol{S_i}^2 = \sum(\boldsymbol{y}_{\text{расчт } i} - \boldsymbol{y}_{\text{фактич } i})^2 = \sum(\boldsymbol{a}\boldsymbol{x}_i + \boldsymbol{b} - \boldsymbol{y}_{\text{фактич } i})^2 = (\boldsymbol{a}\boldsymbol{x}_1 + \boldsymbol{b} - \boldsymbol{y}_1)^2 + (\boldsymbol{a}\boldsymbol{x}_2 + \boldsymbol{b} - \boldsymbol{y}_2)^2$$

$S'_a(b = const) = ((ax_1 + b - y_1)^2)' + ((ax_2 + b - y_2)^2)' = 2(ax_1 + b - y_1)(ax_1 + b - y_1)' + 2(ax_2 + b - y_2)(ax_2 + b - y_2)' = 2(ax_1 + b - y_1)x_1 + 2(ax_2 + b - y_2)x_2$

$S'_a(b = const) = 0$

$2(ax_1 + b - y_1)x_1 + 2(ax_2 + b - y_2)x_2 = 0$

$(ax_1 + b - y_1)x_1 + (ax_2 + b - y_2)x_2 = 0$

$ax_1^2 + bx_1 - y_1x_1 + ax_2^2 + bx_2 - y_2x_2 = 0$

$a(x_1^2 + x_2^2) + b(x_1 + x_2) = y_1x_1 + y_2x_2$

$a\sum x_i^2 + b\sum x_i = \sum y_i x_i$ **(условие, при котором $S'_a(b = const) = 0$)**

$S'_b(a = const) = ((ax_1 + b - y_1)^2)' + ((ax_2 + b - y_2)^2)' = 2(ax_1 + b - y_1)(ax_1 + b - y_1)' + 2(ax_2 + b - y_2)(ax_2 + b - y_2)' = 2(ax_1 + b - y_1) + 2(ax_2 + b - y_2)$

$S'_b(a = const) = 0$

$2(ax_1 + b - y_1) + 2(ax_2 + b - y_2) = 0$

$(ax_1 + b - y_1) + (ax_2 + b - y_2) = 0$

$ax_1 + b - y_1 + ax_2 + b - y_2 = 0$

$a(x_1 + x_2) + 2b = y_1 + y_2$

$a\sum x_i + nb = \sum y_i$ **(условие, при котором $S'_b(a = const) = 0$)**

Построение стандартной парной линейной модели. Метод наименьших квадратов

$$\begin{cases} \sum y_i = a\sum x_i + nb \\ \sum y_i x_i = a\sum x_i^2 + b\sum x_i \end{cases}$$

Объединение двух условий в систему. Решение системы относительно параметров **a, b** позволяет определить **критическую точку** для функции **S**

Способы решения системы

в ручную

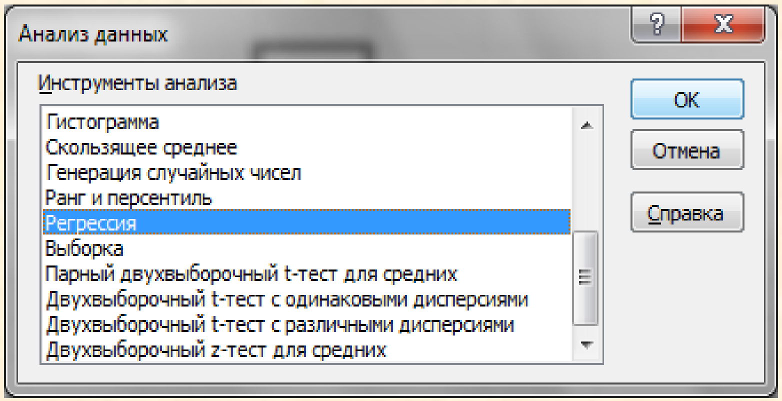
SPSS

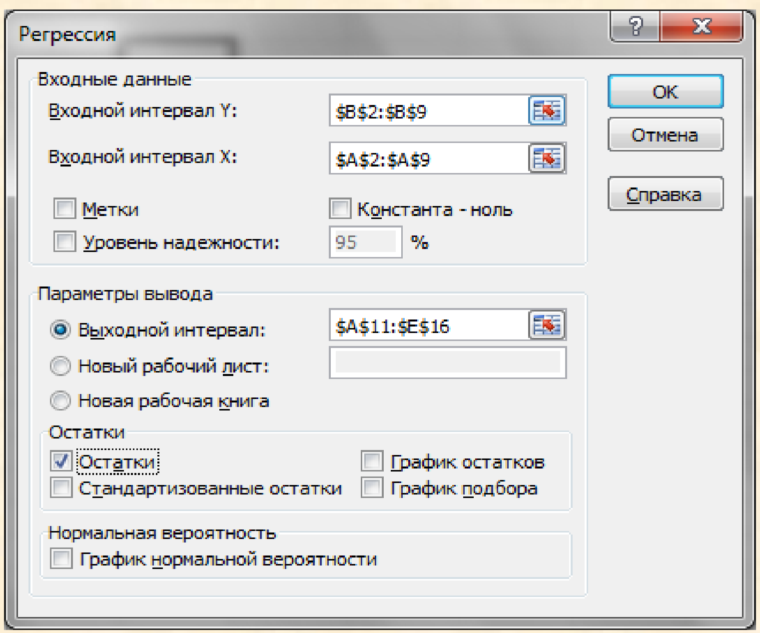
Пакет Анализа (MExcel)

? Критическая точка – это минимум или максимум функции

Построение парной линейной модели с помощью Пакета Анализа

Меню: Данные: Анализ данных





Вывод итогов

Регрессионная статистика	
Множественный R	0,98447717
R-квадрат (коэффициент детерминации)	0,96919529
Нормированный R-квадрат	0,96406117
Стандартная ошибка моделирования	1,61158997
Наблюдения	8

Дисперсионный анализ					
	df (число степеней свободы)	*SS*	*MS*	*F (критерий Фишера)*	*Значимость F*
Регрессия	1	490,291667	490,2917	188,775401	9,24E-06
Остаток	6	15,5833333	2,597222		
Итого	7	505,875			

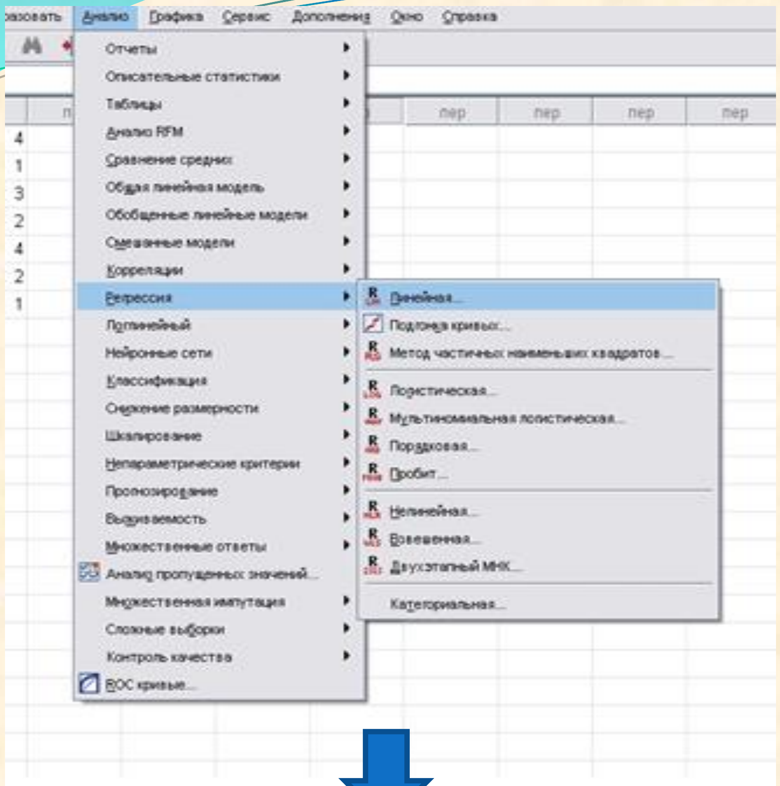
	Коэффициенты (параметры)	*Стандартная ошибка параметра*	*t-статистика*
Y-пересечение	15	1,255	11,945
Переменная X 1	3,416	0,2486	13,73

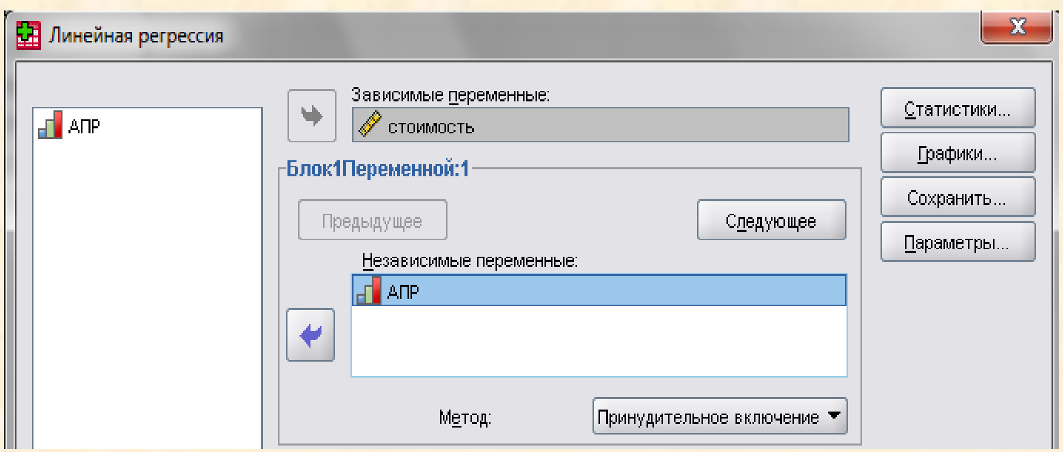
a

b

ВЫВОД ОСТАТКА		
Наблюдение	*Предсказанное Y*	*Остатки*
1	18,4166667	1,58333333
2	21,8333333	0,16666667
3	25,25	-0,25
4	28,6666667	-0,6666667
5	32,0833333	-2,0833333
6	35,5	-0,5
7	38,9166667	-0,9166667
8	42,3333333	2,66666667

Построение парной линейной модели с помощью SPSS





Вывод итогов (файл с расширением SAV)

Модель		Сумма квадратов	ст.св.	Средний квадрат	Щ	Знч.
1	Регрессия	368773,111	1	368773,111	5,349	,069[a]
	Остаток	344740,603	5	68948,121		
	Всего	713513,714	6			

a. Предикторы: (конст) АПР

b. Зависимая переменная: стоимость

Коэффициенты[a]

Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знч.	95,0%% доверительный интервал для B	
		B	Стд. Ошибка	Бета			Нижняя граница	Верхняя граница
1	(Константа)	673,250	227,401		2,961	,031	88,698	1257,802
	АПР	194,838	84,247	,719	2,313	,069	-21,726	411,403

a. Зависимая переменная: стоимость

Статистики остатков[a]

	Минимум	Максимум	Для среднего	Стд. Отклонение	N
Предсказанное значение	868,[illegible]	1452,60	1146,43	247,916	7
Остаток	-328,765	423,397	,000	239,701	7
Стд. Предсказанное значение	-1,12[illegible]	1,235	,000	1,000	7
Стд. Остаток	-1,252	1,612	,000	,913	7

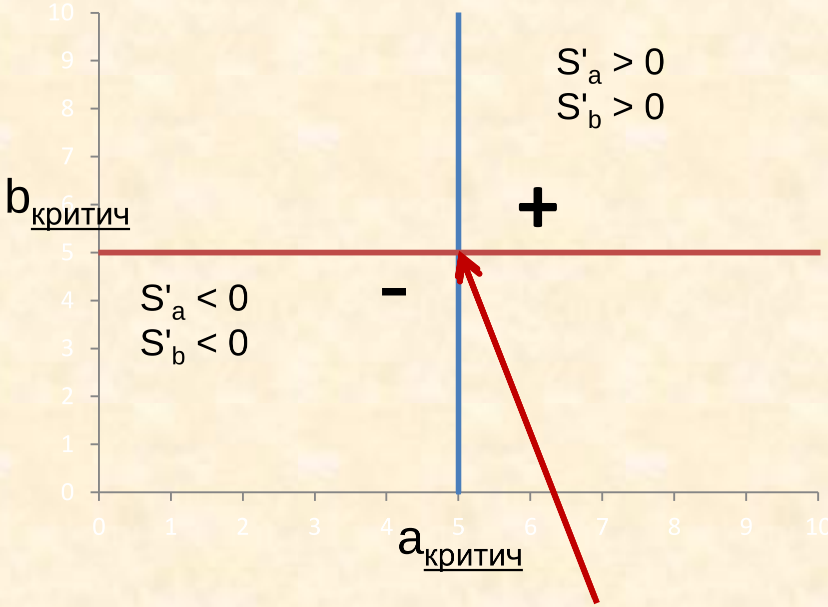
a. Зависимая переменная: стоимость

a

b

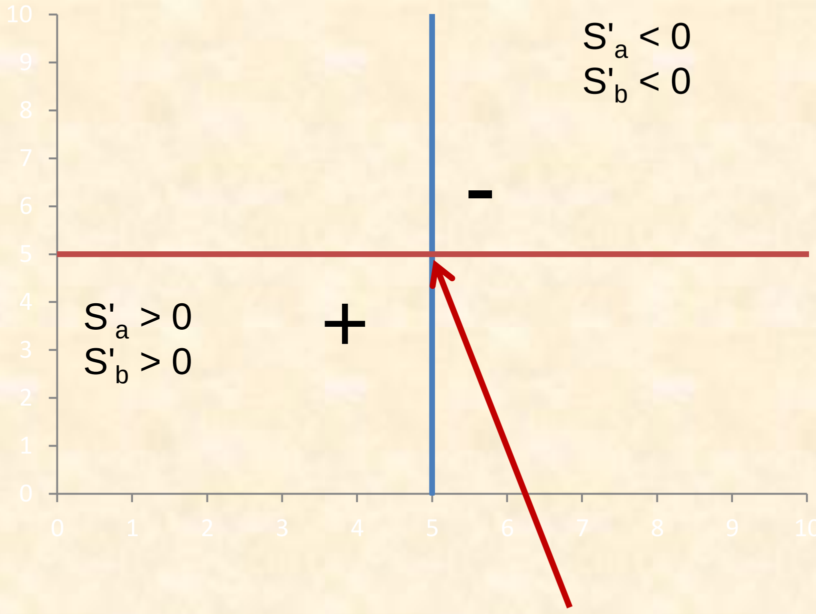
Проверка критической точки на минимум

Критическая точка - минимум



Критическая точка (a, b) - минимум

Критическая точка - максимум



Критическая точка (a, b) - максимум

Возможности метода наименьших квадратов

Линейная модель

МНК,
МЦТ

Стандартные не линейные модели

МНК,
МЦТ

Не стандартные модели

Модель приводится к линейной

МНК

Модель не приводится к линейной

Метод итераций

Последовательность действий в МНК (если система уравнений не известна)

1) выбор функции для аппроксимации

2) Математическая запись выражения для суммы квадратов остатков как функции от параметров модели

3) Нахождение с помощью мат. анализа критической точки данной функции:
   - нахождение частных производных;
   - приравнивание их к нулю;
   - запись полученных уравнений в систему;
   - решение системы относительно неизвестных параметров (найдена критическая точка для функции суммы квадратов остатков

4) проверка критической точки на минимум

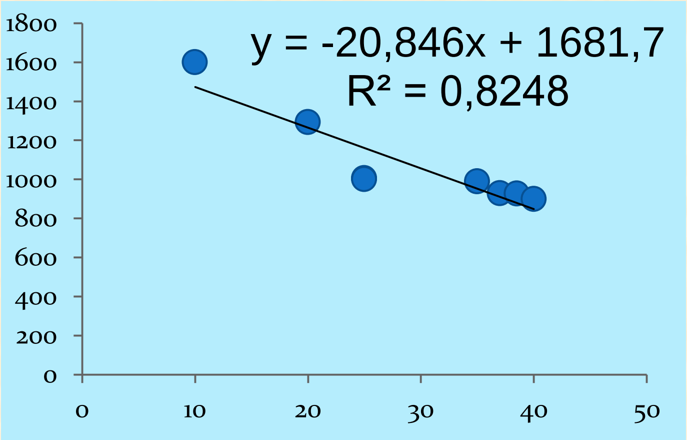
5) если критическая точка – точка минимум, то параметры выбранной модели найдены

Пример построения парной линейной модели на реальных статистических данных

Исходные данные (выборка)

Объект-квартира	Рыночная стоимость, \$/кв. метр	Износ помещения, %
1	1293	20
2	1007	25
3	929	37
4	990	35
5	1600	10
6	1002	25
7	928	38,5
8	900	40

Y (Стоимость), \$/кв. метр



X (Износ), %

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,91
R-квадрат	0,82
Нормированный R-квадрат	0,8
Стандартная ошибка	109
Наблюдения	8

	Коэффициенты	Стандартная ошибка	t-статистика
Y-пересечение	1681,74	119,5	14,07
Переменная X 1	-20,84	3,92	-5,31

Дисперсионный анализ					
	df	SS	MS	F	Значимость F
Регрессия	1	341106,5676	341106,6	28,2	0,001805092
Остаток	6	72470,30742	12078,38		
Итого	7	413576,875			

Наблюдение	Расчетный Y	Остатки
1	1264,828735	28,17126478
2	1160,599666	-153,5996656
3	910,4498985	18,55010152
4	952,1415263	37,85847367
5	1473,286874	126,7131255
6	1160,599666	-158,5996656
7	879,1811776	48,81882241
8	847,9124567	52,08754329

<u>Применение модели: </u>если износ объекта составит 50%, то стоимость должна составить 640 \$/кв. метр, если износ 51% - то стоимость 619 \$/кв. метр, т.е. разница – в 21 доллар возникает за счет разницы в износе в 1% . 21 доллар – это денежная поправка на износ.