

Тема 1. Линейная модель парной регрессии. Метод наименьших квадратов.

Кадочникова Екатерина Ивановна



- 1. Цели, предмет, задачи эконометрики. Этапы эконометрического моделирования.
- 2. Инструментарий эконометрики. Типы моделей и переменных.
- 3. Спецификация линейной модели парной регрессии. Оценки параметров линейной регрессии. Метод наименьших квадратов (МНК).
- 4. Предпосылки МНК и свойства МНК-оценок.

1. Цели, предмет, задачи эконометрики. Этапы эконометрического моделирования.

Термин «эконометрика» впервые был использован бухгалтером **П. Цьемпой, Австро-Венгрия, 1910 г.**

П. Цьемпа считал, что если к данным бухгалтерского учета применить методы алгебры и геометрии, то будет получено новое, более глубокое представление о результатах хозяйственной деятельности. Это употребление термина, как и сама концепция, не прижилось, но название «эконометрика» оказалось весьма удачным для определения нового направления в экономической науке, которое выделилось в 1930 г.

Экономика



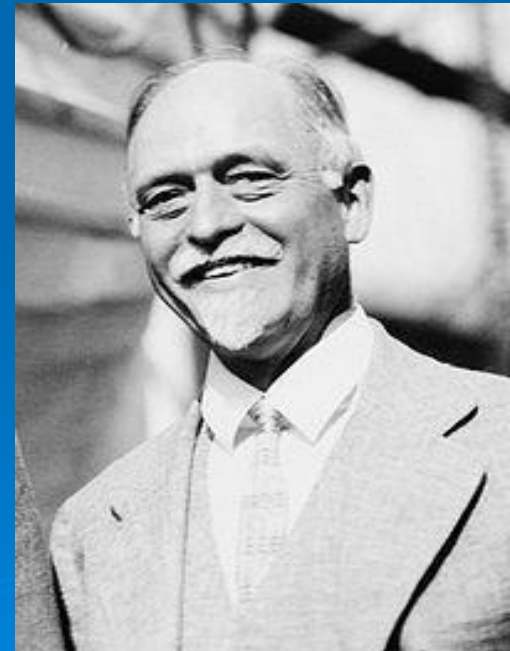
Метрика
(греч.-
метрон -
мера)

ЭКОНО
метри
ка

Становление эконометрики

1912 г. – И. Фишер (Нью-Йорк) сделал попытку создать группу ученых для стимулирования развития экономической теории путем ее связи со статистикой и математикой. Группу создать не удалось.

Ирвинг Фишер ([1867](#) (1867 - [1947](#) (1867 - 1947) — американский [экономист](#), представитель неоклассического направления в экономической науке. Окончил [Йельский университет](#) Окончил Йельский университет; доктор философии родного университета; с 1893 по 1935 г. преподавал там же. Президент [Эконометрического общества](#) Окончил Йельский университет; доктор философии родного университета; с 1893 по 1935 г. преподавал там же. Президент Эконометрического общества (1931-34). Президент [Американской экономической ассоциации](#) Окончил Йельский университет;



Становление эконометрики

1930 г., 29 декабря – на заседании Американской ассоциации развития науки по инициативе И. Фишера, Й. Шумпетера, О. Андерсона, Я. Тинбергена создано эконометрическое общество, на котором **норвежский ученый Р. Фриш** дал новой науке название «эконометрика».

Рагнар Антон Киттиль Фриш (1895 (1895 — 1973) — норвежский экономист. Лауреат Нобелевской премии Лауреат Нобелевской премии 1969 г. «за создание и применение динамических моделей к анализу экономических процессов».



Становление эконометрики

1933 г. – стал издаваться журнал «Econometrica»

1941 г. – издан первый учебник по эконометрике,
автор Я. Тинберген.

Ян Тинберген (1903
(1903—1994) — голландский
экономист.

Нобелевскую премию 1969
года Тинберген получил «за
создание и применение
динамических моделей к
анализу экономических
процессов» (на фото – третий
слева)



Становление эконометрики

1970 – е гг. – противоречия между кейнсианцами, монетаристами и марксистами привели к тому, что методы эконометрики стали применяться не только для оценки теоретических моделей, но и для доказательства причинности при выборе теоретических концепций. Появление компьютеров, создание ARIMA-моделей, VAR-моделей, развитие анализа временных рядов.

Эконометрика – это наука, которая дает **количественное выражение взаимосвязей** экономических явлений и процессов, **которые раскрыты и обоснованы экономической теорией** (И.И. Елисеева).

Эконометрика – это наука, которая на базе экономической теории, экономической статистики, экономических измерений и математико-статистического инструментария придает **количественное выражение** качественным **закономерностям, обусловленным экономической теорией** (С. А. Айвазян)

ИСТОЧНИКИ ЭКОНОМЕТРИКИ

Зарождение эконометрики является следствием *междисциплинарного подхода* к изучению экономики:


Эконометрика

Экономическая
теория

Социально-
экономическая
статистика

Основы теории
вероятностей и
математической
статистики

«**Эконометрика** – это не то же самое, что экономическая статистика. Она не идентична и тому, что мы называем экономической теорией. Эконометрика не является синонимом приложений математики к экономике. Каждая из трех отправных точек – статистика, экономическая теория и математика – необходимое, но не достаточное условие для понимания количественных соотношений в современной экономической жизни. **Это единство всех трех составляющих.** И это единство образует эконометрику» (Р. Фриш, 1933 г.)



Предметом эконометрики является определение наблюдаемых в экономике количественных закономерностей

Цель эконометрики – эмпирический (практический) вывод экономических законов

Прикладные цели – прогнозная оценка экономических показателей, априорная имитация альтернативных сценариев развития анализируемой системы

Основные задачи эконометрики

- построение эконометрической модели;
- оценка параметров построенной модели, делающих выбранную модель наиболее адекватной реальным данным;
- проверка качества найденных параметров модели и самой модели в целом;
- использование построенных моделей для объяснения поведения исследуемых экономических показателей, прогнозирования, осмысленного проведения экономической политики (С. А. Бородич)

Этапы моделирования

- 1. постановочный
- 2. априорный
- 3. спецификация модели
- 4. информационный
- 5. идентификация модели - калибровка модели
- 6. верификация модели
- 7. интерпретация результатов

1) Вопрос исследования:

*выявить фактор, влияющий на среднее
число зрителей за первые три дня проката фильма;*

2) Гипотеза исследования: *расходы на рекламу фильма
влияют на число зрителей за первые три дня проката фильма;*

3) Тестирование гипотезы и доказательства для подтверждения аргументов

4) Сбор данных

Кросс-секционные данные о расходах на рекламу и среднем числе зрителей по 20 кинофильмам

Среднее число зрителей за первые три дня проката фильма ,чел.	Расходы на рекламу фильма, руб.	Среднее число зрителей за первые три дня проката фильма ,чел.	Расходы на рекламу фильма, руб.
282	2750	305	4565
263	2430	328	5987
295	3700	335	6100
276	2860	251	2375
285	3180	292	3480
342	4270	290	3295
276	2875	387	7500
328	5295	326	5430
321	5140	347	6310
326	4870	234	2100

Типы исходных данных

- Перекрестные данные (кросс-секции)
- Временные ряды
- Панельные данные

Перекрестные данные

Множество данных, состоящих из наблюдений за **несколькими** однотипными статистическими объектами в течение **одного** периода или за один момент времени, называется **перекрестными данными** (кросс-секциями)

Показатели российских банков за июль 2015 года

Банк	Работающие активы, млн руб.	Чистая прибыль, млн. руб	Кредитный портфель, млн. руб.
Сбербанк	20759308	81583	14170300
ВТБ	7722473	21055	3466974
Промсвязьбанк	1060544	8643	717639
Югра	267518	4465	160841

http://www.banki.ru/banks/ratings/?PROPERTY_ID=40

Временные ряды

Множество данных, состоящих из наблюдений за **одним** статистическим объектом в течение **нескольких** периодов или за несколько моментов времени, называется **временным рядом**.

Год	ВВП в текущих ценах, млрд. руб.
2010	46308,5
2011	55967,2
2012	62176,5
2013	66190,1
2014	71406,4

Дата	Численность населения РФ, млн. чел.
На 01.01. 2011	142,9
На 01.01.2 012	143,0
На 01.01. 2013	143,3
На 01.01. 2014	143,7

http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/accounts/#

http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/demography/

Панельные данные

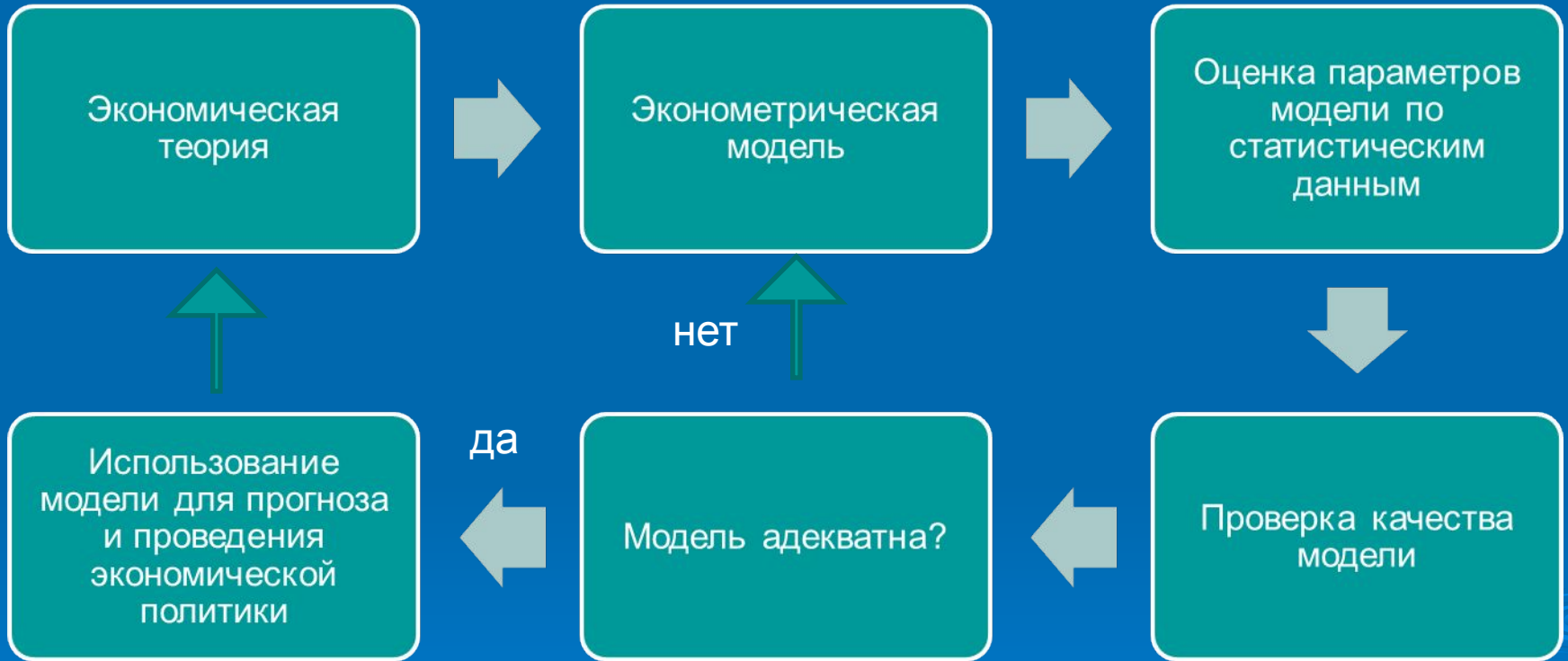
Множество данных, состоящих из наблюдений за **несколькими** однотипными статистическими объектами в течение **нескольких** временных периодов, называется **панельными данными**.

Панельные данные о товарообороте ОАО «Магнит»

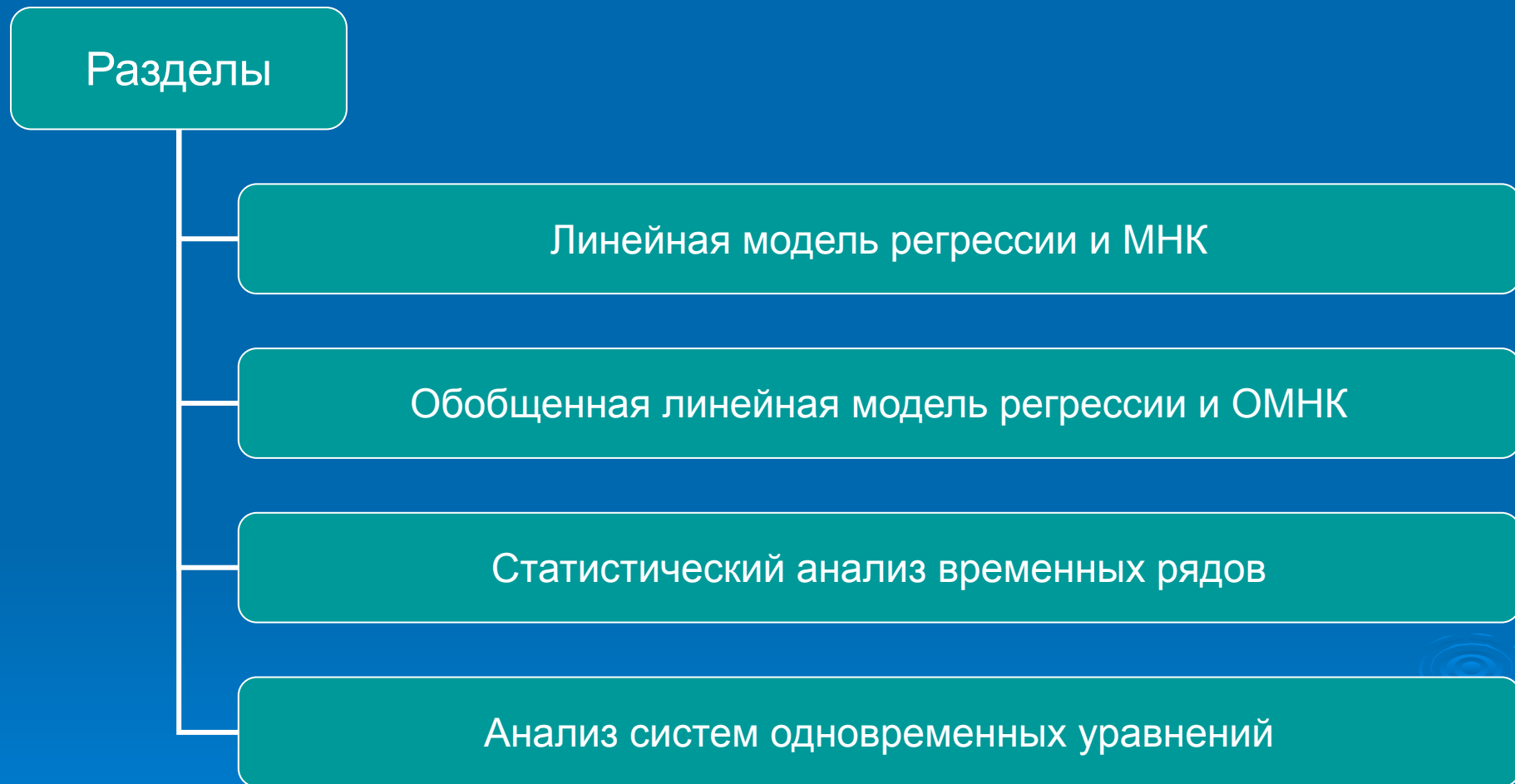
№	Y	X1	X2	№	Y	X1	X2
Казань				Москва			
1	22,836	0,2472	12,276	15	71,511	0,9682	12,432
2	30,899	0,3193	13,068	16	75,955	1,2463	13,632
3	36,245	0,5665	12,336	17	80,443	1,3287	10,668
4	41,965	0,6901	12,384	18	89,298	1,5347	9,06
5	51,458	0,8549	13,02	19	98,164	1,7201	9,372
6	60,841	1,0094	13,656	20	107,03	1,8952	9,696
7	67,012	1,1742	14,292	21	126,896	2,0806	14,208
Санкт-Петербург				Краснодар			
8	45,188	0,4635	1,74	22	100,386	1,1536	12,864
9	61,919	0,8034	2,424	23	109,824	1,3287	13,524
10	75,361	1,0094	4,524	24	119,405	1,5347	15,624
11	90,992	1,2772	6,624	25	128,887	1,7201	16,092
12	106,073	1,5347	9,012	26	138,391	1,9055	16,344
13	121,165	1,7922	10,848	27	147,906	2,1012	17,208
14	136,246	2,0497	14,412	28	157,41	2,2866	17,82

Y - годовой товарооборот (млн. руб.); X1 - торговая площадь (тыс. кв. м), X2 - среднее число посетителей в день (тыс. чел.).

<http://magnit-info.ru/>



2. Инструментарий эконометрики. Типы моделей и переменных.



Особенности эконометрического метода

- Исследование статистических зависимостей, а не функциональных.
- Отражение особенностей экономических переменных и связей между ними (оптимальность и взаимодействие переменных)
- Содержательное обоснование уравнений регрессии
- Изучение всей совокупности связей между переменными, а не изолированно взятого уравнения регрессии
- Развитие анализа временных рядов через решение проблем ложной корреляции , лага и других

Типы моделей и переменных



Экзогенные переменные
обозначаются обычно
как x .

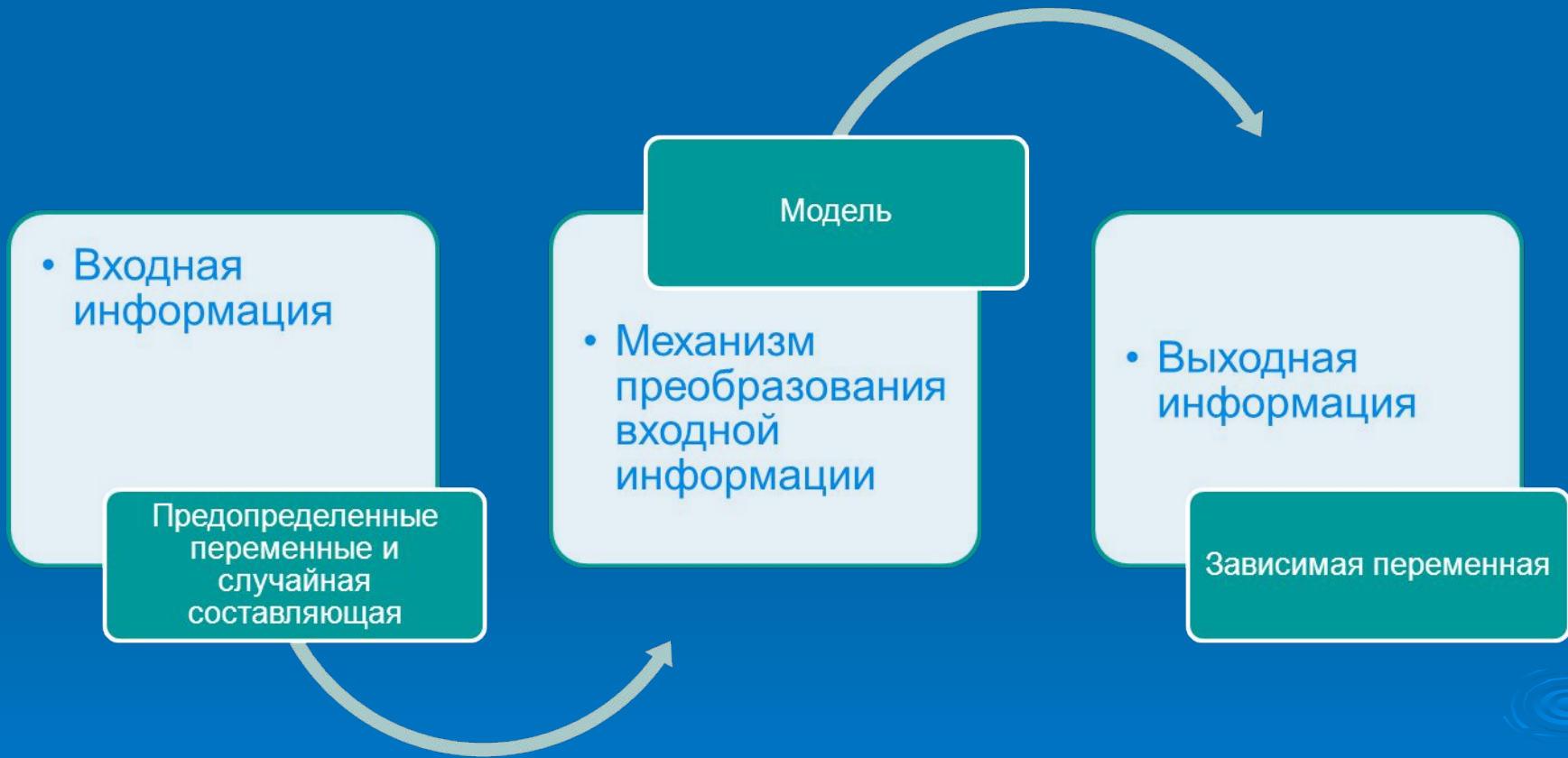
Это внешние для модели
переменные,
управляемые из
вне, влияющие на
эндогенные
переменные, но не
зависящие от них.

Эндогенные переменные
обозначаются обычно
как y .

Это внутренние,
формируемые в модели
переменные,
зависимые от
предопределенных
переменных.

Предопределенными
переменными называют
экзогенные

переменные x и
лаговые эндогенные
переменные y_{t-l} .



3. Спецификация линейной модели парной регрессии. Оценки параметров линейной регрессии. Метод наименьших квадратов (МНК).

Корреляционный анализ

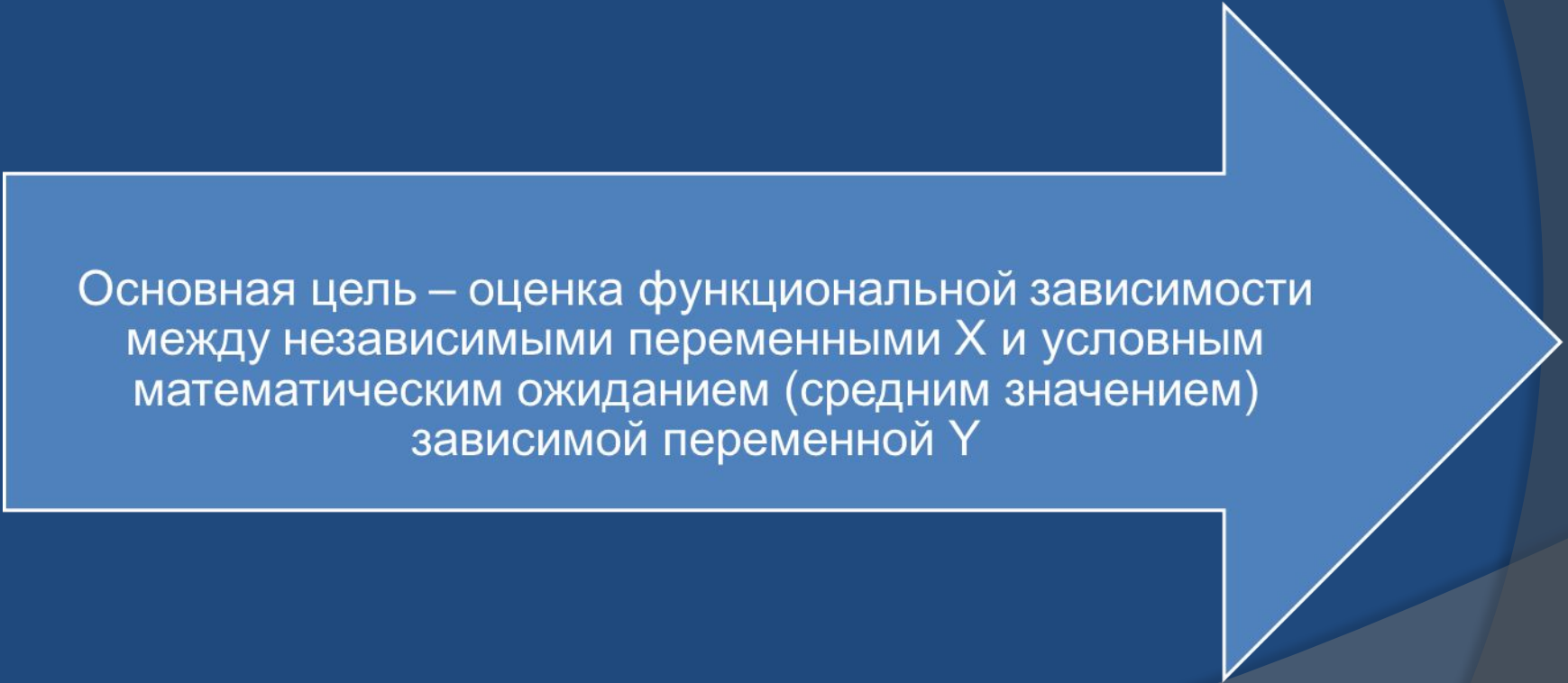
- X и Y равноценны, не делятся на независимую и зависимую
- Измеряют наличие и силу взаимосвязи между X и Y , основной мерой является коэффициент корреляции

Регрессионный анализ

- X и Y не равноценны, изменение независимой X служит причиной для изменения зависимой Y
- Анализируют как X влияет на Y «в среднем» и определяют функцию регрессии Y на X .

Цель регрессионного анализа

Термин «регрессия» был введен Фрэнсисом Гальтоном в конце 19 века.



Основная цель – оценка функциональной зависимости между независимыми переменными X и условным математическим ожиданием (средним значением) зависимой переменной Y

Виды регрессии

Модели регрессии

```
graph TD; A[Модели регрессии] --> B[По размерности:]; A --> C[По форме зависимости:]; A --> D[По направлению связи:]; B --> B1[- Простая (Парная)]; B --> B2[- Множественная]; C --> C1[- Линейная]; C --> C2[- Нелинейная]; D --> D1[- Прямая]; D --> D2[- Обратная];
```

По размерности:

- Простая (Парная)
- Множественная

По форме зависимости:

- Линейная
- Нелинейная

По направлению связи:

- Прямая
- Обратная

Простая (парная) регрессия представляет собой модель, где среднее значение зависимой переменной Y рассматривается как функция одной независимой переменной X :

$$Y_x = f(x)$$

Множественная регрессия представляет собой модель, где среднее значение зависимой переменной Y рассматривается как функция нескольких независимых переменных X_1, X_2, \dots :

$$Y_x = f(x_1, x_2, \dots, x_m)$$

Спецификация модели - формулирование вида модели, исходя из соответствующей теории связи между переменными.

Исследование начинается с теории, устанавливающей связь между явлениями.
(И. И. Елисеева)

*Определяется **состав переменных** и **математическая функция** для отражения связи между ними.*

Спецификация линейной модели парной регрессии

$$Y_i = \hat{Y}_{x_i} + \varepsilon_i$$

Y_i - фактическое значение зависимой переменной Y

\hat{Y}_{x_i} - теоретическое (среднее) значение зависимой переменной Y , найденное из уравнения регрессии

ε_i - случайная величина (остаток регрессии)

Теоретическая линейная модель парной регрессии

$$Y_i = \alpha + \beta \cdot x_i + \varepsilon_i$$

α – свободный коэффициент

β - коэффициент регрессии

ε_i – случайное отклонение (возмущение)

Случайное отклонение включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Источники его присутствия в модели: спецификация модели, выборочный характер исходных данных, особенности измерения переменных.

Эмпирическое уравнение линейной парной регрессии

$$Y_{x_i} = a + b \cdot x_i$$

Y_{x_i} - теоретическое (среднее) значение зависимой переменной Y , найденное из уравнения регрессии

b - эмпирический коэффициент регрессии

a - эмпирический свободный коэффициент

В конкретном случае – линейная модель парной регрессии:

$$Y_i = a + b \cdot x_i + e_i$$

e_i – оценка теоретического случайного отклонения ε

Типы ошибок в регрессии

Ошибки спецификации

- Неправильный выбор математической функции
- Недоучет существенного фактора

Ошибки выборки

- Неоднородные статистические данные
- Неправильный выбор временного интервала информации

Ошибки измерения

- Преднамеренные ошибки в отчетности
- Непреднамеренные ошибки из-за сокрытия информации

Методы выбора типа уравнения регрессии

Графический
метод

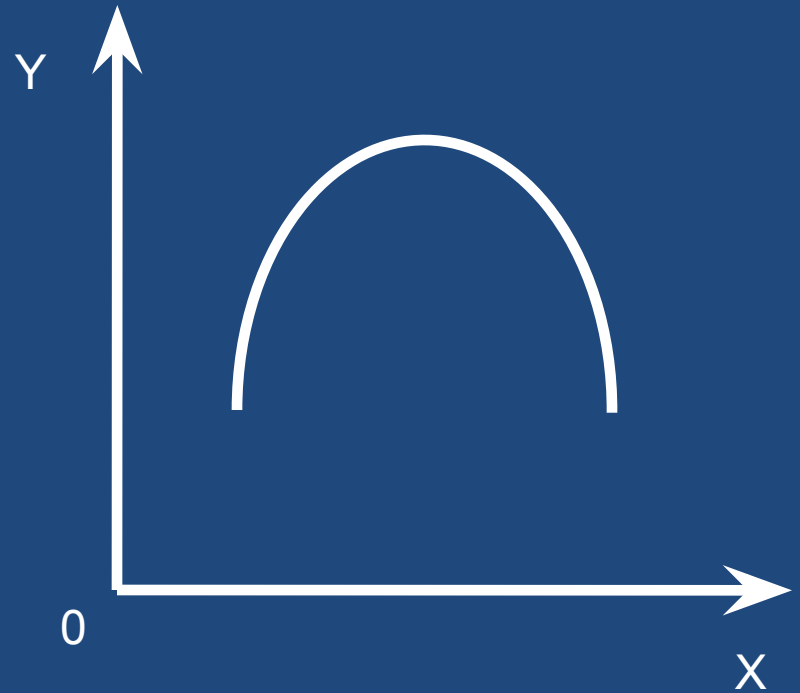
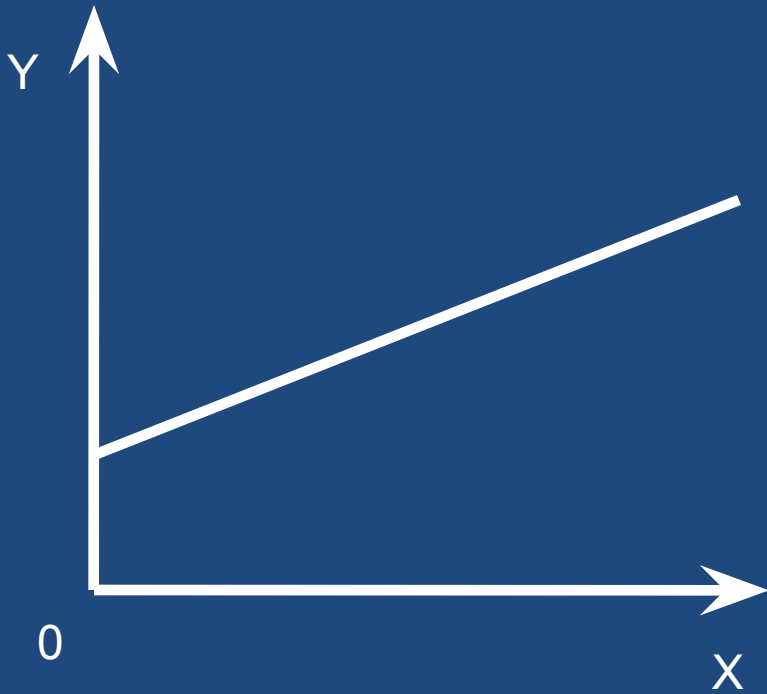
- Основан на визуальном анализе поля корреляции

Аналитический
метод

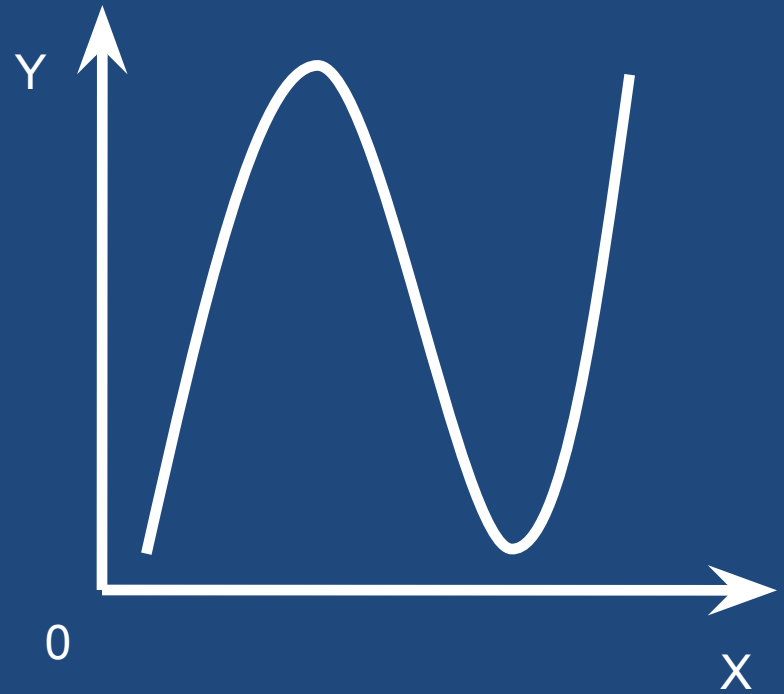
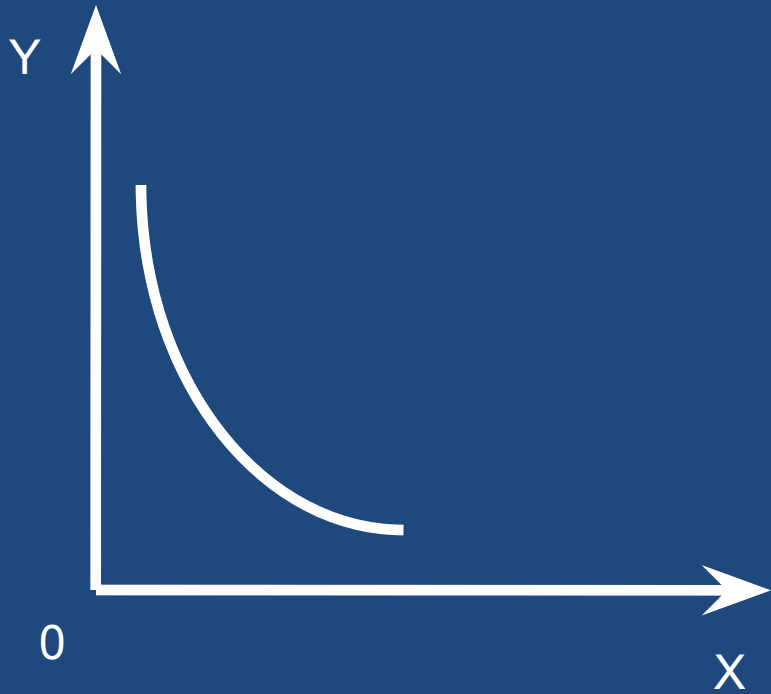
- Основан на изучении материальной природы взаимосвязи

Эксперименталь
ный метод

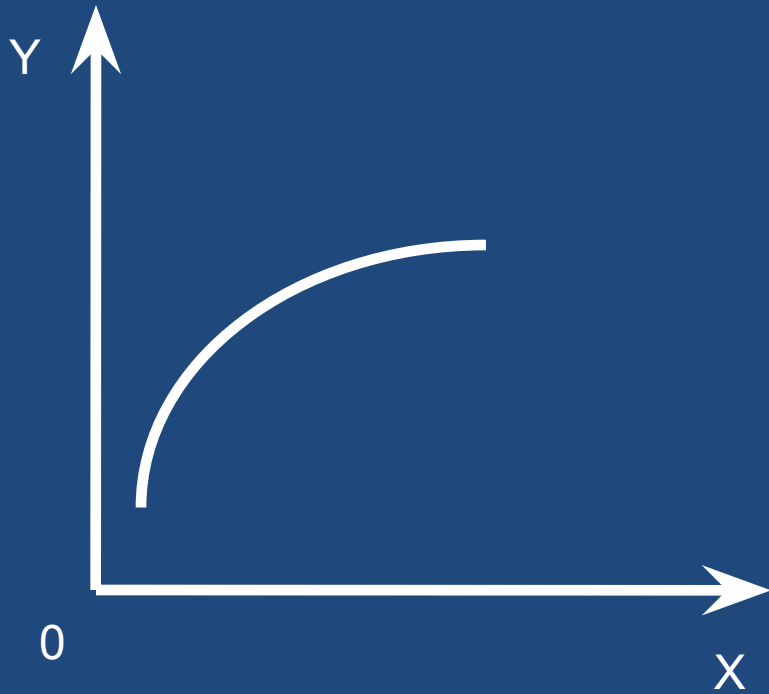
- Основан на сравнении величины остаточной дисперсии, рассчитанной при разных моделях



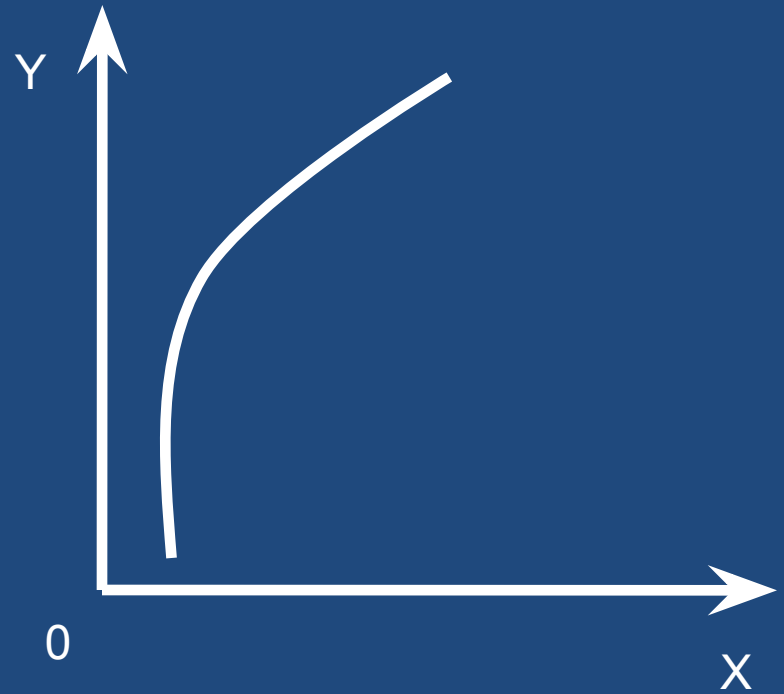
$$Y_x = a + b \cdot x \quad Y_x = a + b \cdot x + c \cdot x^2$$



$$Y_x = a + b/x \quad Y_x = a + b \cdot x + c \cdot x^2 + d \cdot x^3$$



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



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

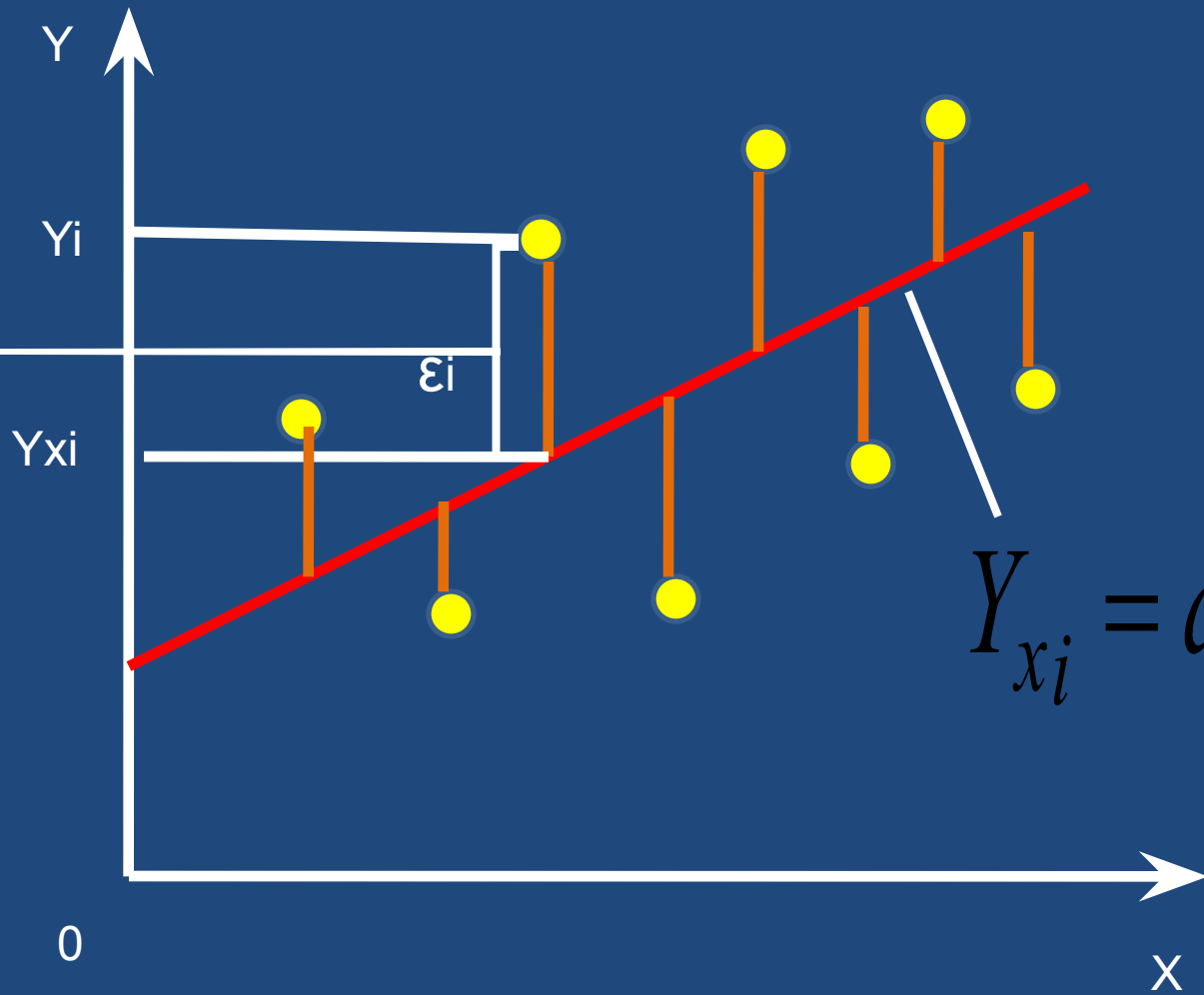
Построение линейной регрессии сводится к оценке ее параметров – a и b



Классический подход к оцениванию параметров основан на *методе наименьших квадратов*



Из множества линий на графике выбирается та, для которой *минимальна сумма квадратов* расстояний по вертикали между точками наблюдений и этой линией



$$Y_{xi} = a + b \cdot x_i$$

Суть метода наименьших квадратов (МНК) - оценки параметров таковы, что сумма квадратов отклонений фактических значений зависимой переменной Y от расчетных (теоретических) Y_x минимальна:

$$\sum_{i=1}^n (y_i - y_{xi})^2 \rightarrow \min$$

Оценка параметров регрессии

$$S = \sum (y_i - y_{x_i})^2 = \sum (y - a - b \cdot x)^2;$$

$$\frac{dS}{da} = -2 \sum y + 2 \cdot n \cdot a + 2 \cdot b \sum x = 0;$$

$$\frac{dS}{db} = -2 \sum y \cdot x + 2 \cdot a \sum x + 2 \cdot b \sum x^2 = 0.$$

Оценка параметров регрессии

$$\begin{cases} n \cdot a + b \sum x = \sum y, \\ a \sum x + b \sum x^2 = \sum y \cdot x \end{cases}$$

$$a = \bar{y} - b \cdot \bar{x},$$

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - (\bar{x})^2}$$

4. Предпосылки МНК и свойства МНК-оценок

В силу несовпадения статистической базы для генеральной совокупности и выборки оценки параметров регрессии a и b отличаются от теоретических коэффициентов α и β и не позволяют сделать вывод, насколько точно эмпирическое уравнение регрессии соответствует уравнению для всей генеральной совокупности.

Доказано, что надежность оценок параметров регрессии существенно зависит от свойств случайного отклонения ε . Для получения наилучших МНК-оценок необходимо, чтобы выполнялся ряд предпосылок относительно ε .

Предпосылки МНК

1. Математическое ожидание случайного отклонения ε_i равно нулю для всех наблюдений.

$$M(\varepsilon_i) = 0$$

2. Дисперсия случайных отклонений ε_i постоянна. Выполнение данной предпосылки называется **гомоскедатичностью**, нарушение – **гетероскедастичностью**.

$$D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2$$

Предпосылки МНК

3. Случайные отклонения ε_i и ε_j являются независимыми друг от друга для $i \neq j$. Выполнение данной предпосылки говорит об отсутствии **автокорреляции**, нарушение – о присутствии автокорреляции.

$$\sigma_{\varepsilon_i \varepsilon_j} = \text{COV}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$$

$$\sigma_{\varepsilon_i \varepsilon_j} = \text{COV}(\varepsilon_i, \varepsilon_j) = \sigma^2, i = j$$

4. Случайное отклонение должно быть независимо от объясняющих переменных

$$\sigma_{\varepsilon_i x_i} = \text{COV}(\varepsilon_i, x_i) = 0$$

5. Модель линейна относительно параметров

Свойства МНК-оценок

Теорема Гаусса- Маркова. Если предпосылки МНК выполнены, то МНК-оценки обладают следующими свойствами:

1. Оценки являются несмещенными:

$$M(a) = \alpha, M(b) = \beta$$

2. Оценки состоятельны, так как их дисперсия при увеличении объема выборки стремится к нулю:

$$D(a) \rightarrow 0, D(b) \rightarrow 0, n \rightarrow \infty$$

3. Оценки эффективны, имеют наименьшую дисперсию по сравнению с другими оценками, линейными относительно зависимой переменной

$$D(a) \rightarrow D_{\min}, D(b) \rightarrow D_{\min}$$

Предпосылки
2 и 3
нарушены

Предпосылки
1-5
выполнены

1. Оценки
являются
несмещенными

2. Оценки
состоятельны

3. Оценки
эффективны

Оценки
неэффективны

