Математические методы прогнозирования уровня спроса и закупок в компании

Доц., к.э.н. Ставицкий А.В.

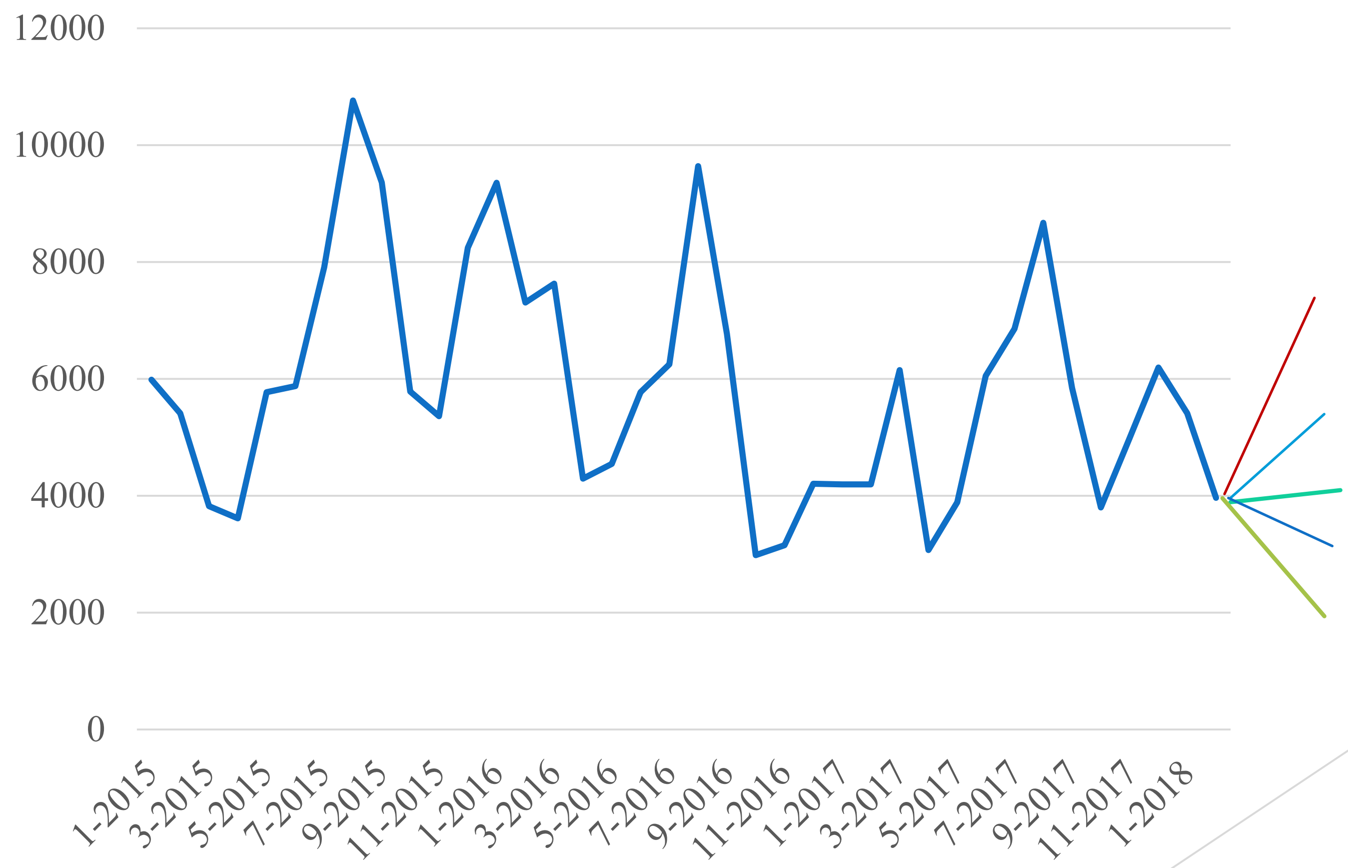
Типичная ситуация

	A	B
1	Дата	Сумма по полю Количество оборот
14	1-2016	9354
15	2-2016	7306
16	3-2016	7629
17	4-2016	4294
18	5-2016	4547
19	6-2016	5775
20	7-2016	6251
21	8-2016	9639
22	9-2016	6772
23	10-2016	2985
24	11-2016	3153
25	12-2016	4208
26	1-2017	4197
27	2-2017	4195
28	3-2017	6148
29	4-2017	3069
30	5-2017	3889
31	6-2017	6052
32	7-2017	6858
33	8-2017	8668
34	9-2017	5842
35	10-2017	3800
36	11-2017	4987
37	12-2017	6193
38	1-2018	5411
39	2-2018	3962
40		
41		

?

	A	B	C	D	E
1	Дата календаря	Количество оборот	Сумма продаж, грн.	Количество остаток	
1134	06.02.2018	125	26521	4339	
1135	07.02.2018	123	21723	3753	
1136	08.02.2018	69	18543	3739	
1137	09.02.2018	142	44979	3711	
1138	10.02.2018	215	55852	3669	
1139	11.02.2018	240	60776	3622	
1140	12.02.2018	47	11360	3613	
1141	13.02.2018	80	22286	3597	
1142	14.02.2018	140	35397	3569	
1143	15.02.2018	159	41159	3538	
1144	16.02.2018	145	39619	3509	
1145	17.02.2018	258	69705	3459	
1146	18.02.2018	216	65905	3416	
1147	19.02.2018	114	26732	4089	
1148	20.02.2018	114	34864	4066	
1149	21.02.2018	61	15827	4055	
1150	22.02.2018	145	46016	4026	
1151	23.02.2018	116	32708	4003	
1152	24.02.2018	237	73640	3957	
1153	25.02.2018	173	52281	3922	
1154	26.02.2018	64	18052	3910	
1155	27.02.2018	57	17924	3898	
1156	28.02.2018	100	29368	3875	
1157	01.03.2018				
1158	02.03.2018				
1159	03.03.2018				
1160	04.03.2018				
1161	05.03.2018				
1162	06.03.2018				
1163	07.03.2018				

Или так…



12000
10000
8000
6000
4000
2000
0
1-2015
3-2015
5-2015
7-2015
9-2015
11-2015
1-2016
3-2016
5-2016
7-2016
9-2016
11-2016
1-2017
3-2017
5-2017
7-2017
9-2017
11-2017
1-2018

Зачем?

- Обоснованное построение прогнозов
- Самостоятельный выбор моделей для прогнозирования
- Автоматизация расчета прогнозов на фирме

Временной ряд

Пусть $y_1, y_2, \ldots, y_T$ - значение наблюдений за экономическим процессом в течение T периодов.

Эта последовательность является числовыми значениями, каждое из которых имеет соответствующий индекс, который зависит от номера периода, в который он наблюдался.

Такая последовательность, записанная в порядке возрастания индекса, называется **временным рядом.**

Цель анализа

Целью прикладного анализа является построение математической модели ряда, с помощью которой можно объяснить поведение ряда и осуществить прогноз на будущие периоды.

$$\hat{y}_{T+1} - ?$$

Структура временного ряда

- Любой временной ряд можно представить как сумму детерминированного и случайного компонентов:

$$y_t = d_t + r_t,\ t = \overline{1,T}.$$

- В свою очередь детерминированный компонент состоит из трех частей: трендового, сезонного, циклического компонентов.

$$d_t = tr_t + s_t + c_t,\ t = \overline{1,T}.$$

Компонеты временного ряда - 1

- Анализ временного ряда начинается с выделения **трендового компонента.** Его присутствие нетрудно заметить, проанализировав график временного ряда. Как правило, для экономических данных очень типичным является медленный рост или падение в течение длительного периода времени.
- Наличие тренда в экономических временных рядах можно объяснить демографическими изменениями, технологическими изменениями, изменениями в структуре производства, спроса и тому подобное. Действие таких факторов является постоянным, поэтому исследователи могут описывать такие изменения с помощью кривых, которые можно задать в аналитическом виде.

Компонеты временного ряда - 2

- **Сезонный компонент** показывает колебания вокруг трендового компонента. Его наличие объясняется сезонным характером производства, потребления. Например, в четвертом квартале каждого года перед Новым годом значительно возрастает потребление товаров.
- Главная идея выделения сезонных колебаний заключается в сравнении данных за соответствующие периоды, а не за прошлые периоды, то есть, например, данные за декабрь одного года надо сравнивать с данными декабря прошлого года, а не с ноябрем.

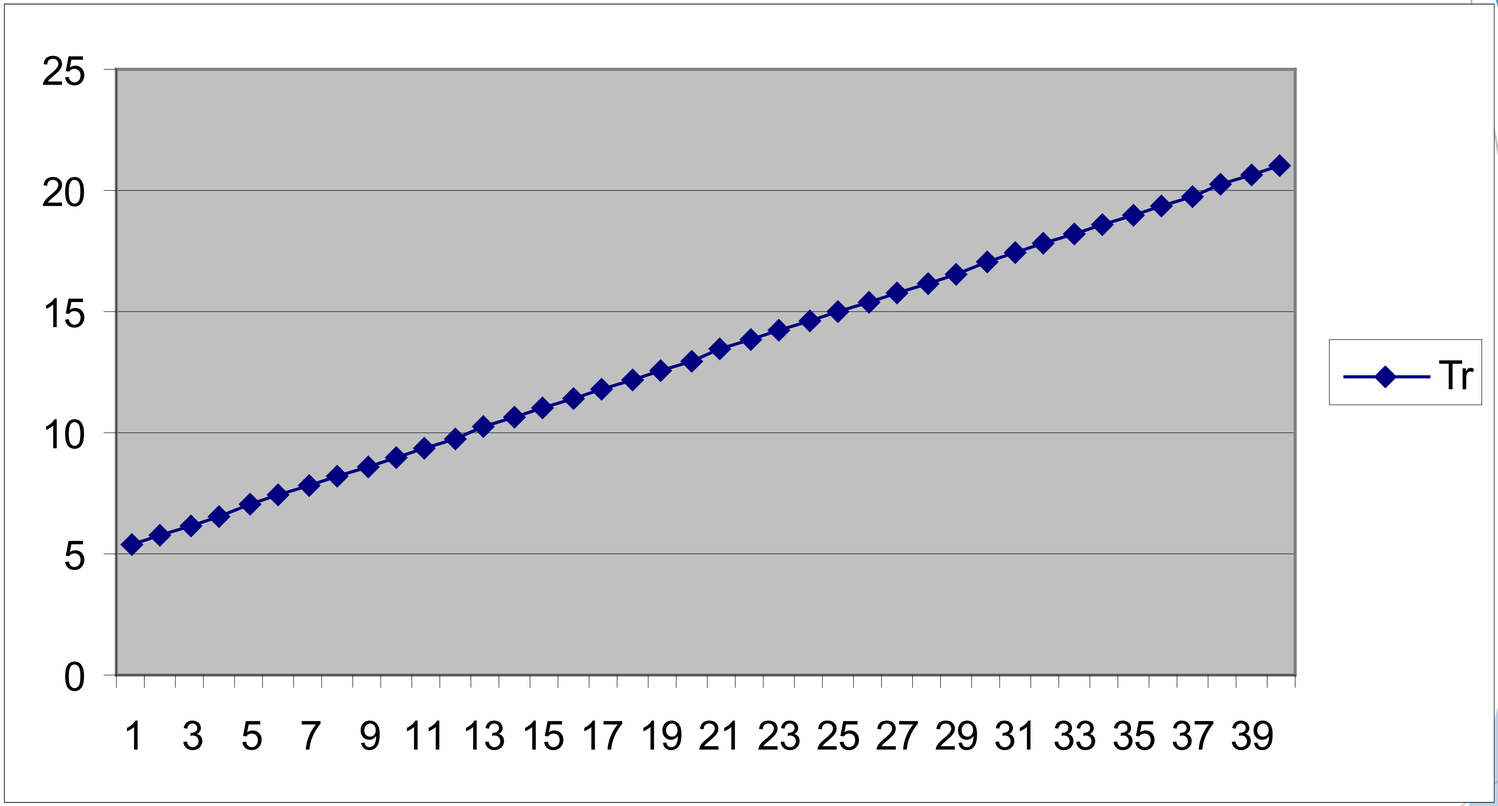
Компонеты временного ряда - 3

- **Циклический компонент** занимает промежуточное место между трендом и сезонным компонентом. Тренд - это гладкое изменение, которое проявляется на большом промежутке времени. Сезонный компонент - это периодическая функция, зависящая от времени, причем период его значительно меньше количества наблюдений.
- Циклический компонент рассматривается, в основном, как гладкое изменение, зависящая от времени, но которое не включается ни в тренд, ни в сезонный компонент.

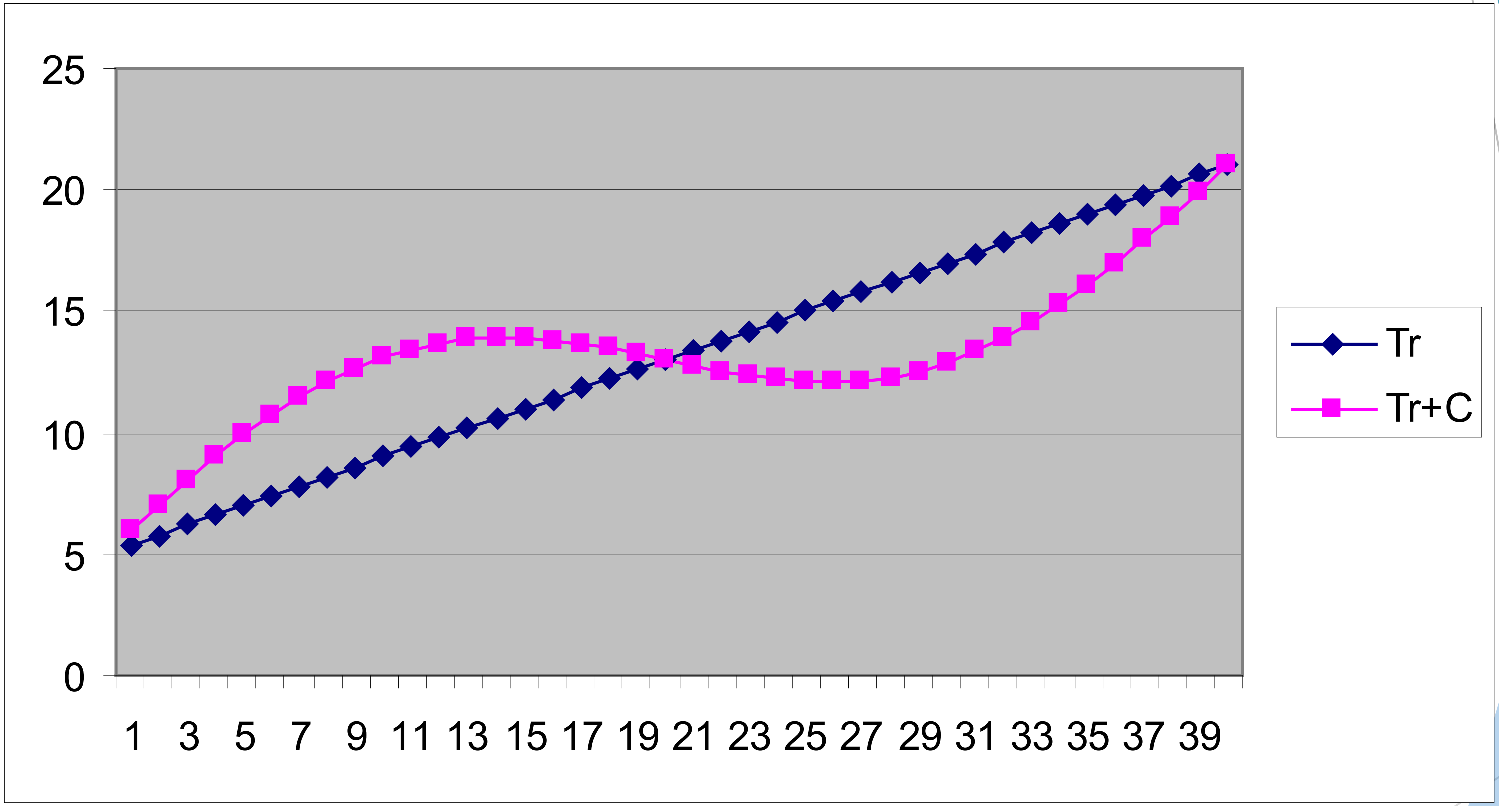
Компонеты временного ряда - 4

- **Случайный компонент** является тем, что осталось от временного ряда после исключения тренда, циклического и сезонного компонентов. Часть таких эффектов может быть отнесена к непредвиденным природным катаклизмам (землетрясения, пожары и т.д.), часть - к случайным действий людей.
- При наличии случайной компоненты невозможно прогнозировать значения временного ряда без ошибки. Но любой реальный экономический процесс включает случайный компонент.

Пример

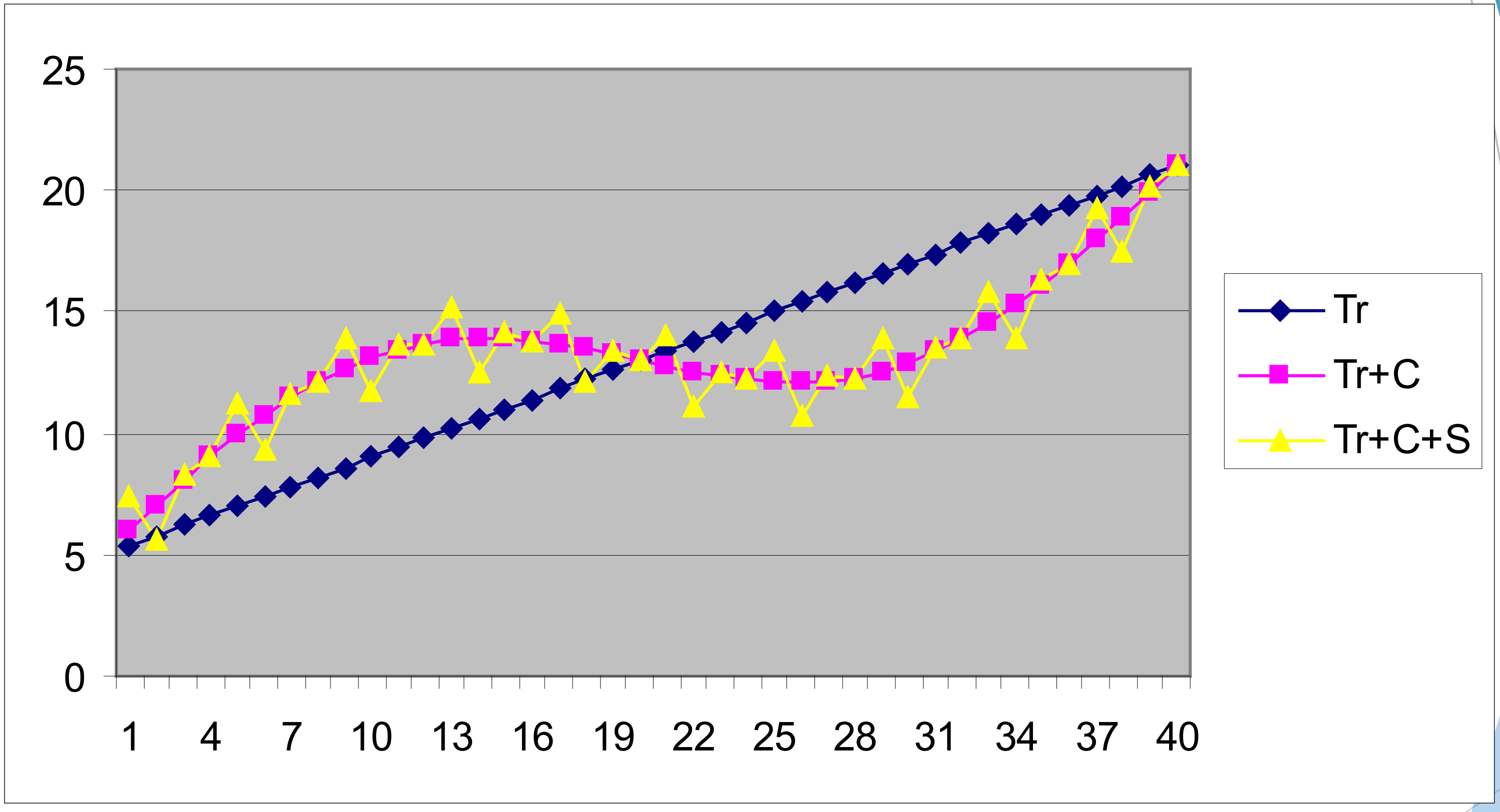


Пример

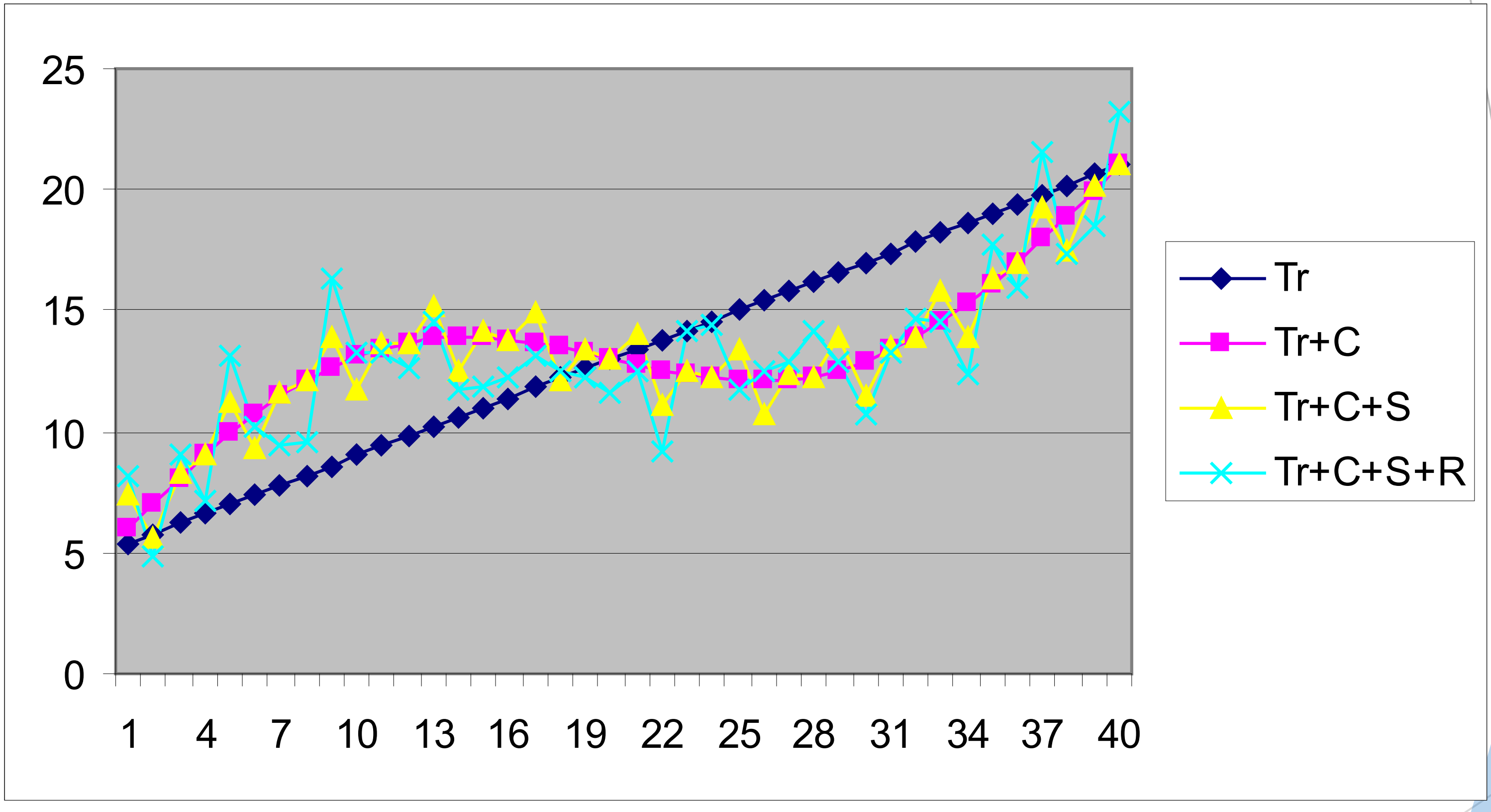


25
20
15
10
5
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39
Tr
Tr+C

Пример



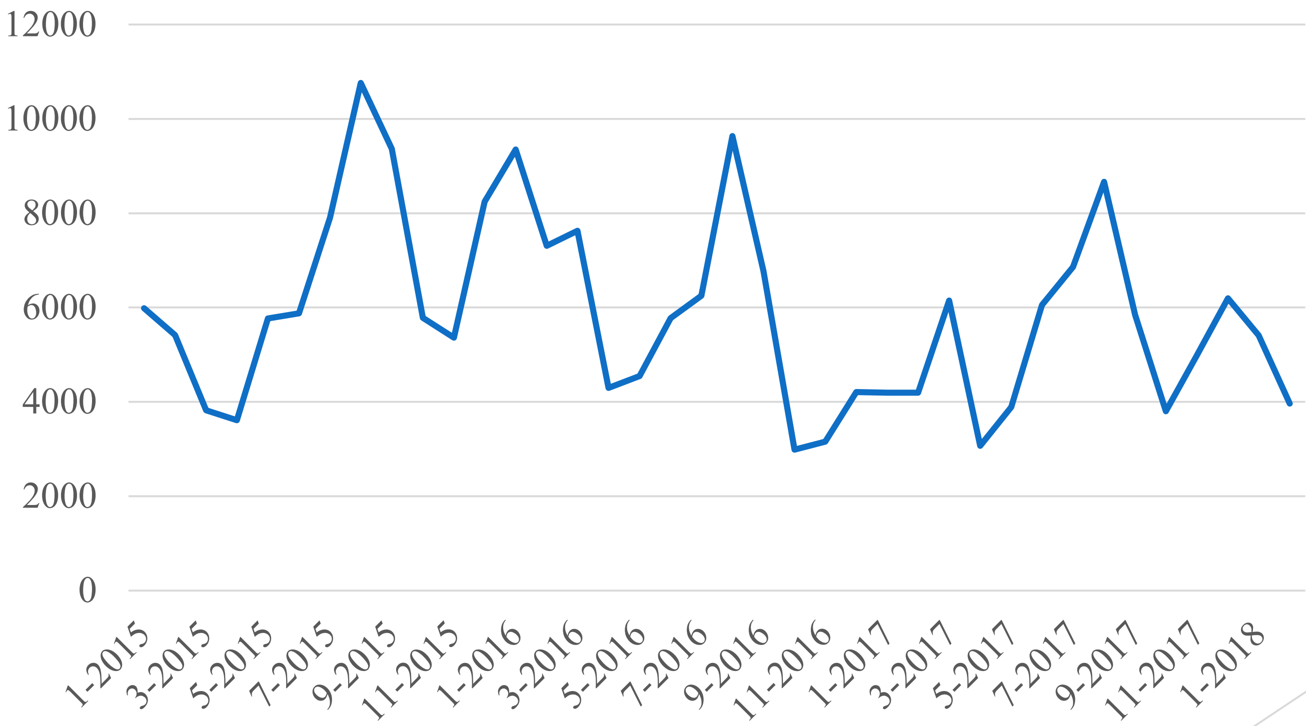
Пример



Порядок анализа

1. Построение и изучение графика ряда
2. Выбор модели
3. Прогнозирование

Построение и изучение графика



На графике определяют

- наличие тренда и его характер;
- наличие сезонных и циклических компонентов;
- степень медлительности или прерывности изменений последовательных значений ряда после устранения тренда.

Методы прогнозирования объема продаж

- методы экспертных оценок;
- методы анализа и прогнозирования временных рядов;
- каузальные (причинно-следственные) методы (попытка найти факторы, определяющие поведение прогнозируемого показателя).

Экспертные методы

Их много, но они требуют:

- Экспертов
- Денег
- Времени
- Знаний в конкретной области

Методы прогнозирования временных рядов

- Метод скользящего среднего
- Метод экспоненциального сглаживания
- Метод двойного экспоненциального сглаживания
- Метод тройного экспоненциального сглаживания
- Метод Холта-Винтерса
- Метод выделения сезонности

Причинно-следственные методы

- Трендовые модели
- Регрессионный анализ
- Модели с фиктивными переменными

Когда использовать?

Ситуация	Метод
Ничего не известно, нет данных	Экспертный метод
Есть ряд данных продаж, нет тренда	Экспоненциальное сглаживание Скользящее вреднее
Есть ряд данных продаж, есть линейный тренд	Двойное экспоненциальное сглаживание Метод Холта-Винтерса
Есть ряд данных продаж, есть квадратичный тренд	Тройное экспоненциальное сглаживание
Есть ряд данных продаж, есть сезонность	Метод выделения сезонности
Есть ряды данных продаж и других факторов	Регрессионный анализ

Точность прогнозов

- О точности прогноза принято судить по размеру ошибки прогноза - разницы между прогнозным и фактическим значением исследуемой переменной. Однако такой подход к оценке точности возможен только в двух случаях.
- Во-первых, когда период прогнозирования уже закончился, и исследователь знает фактические значения переменной. При краткосрочном прогнозировании это вполне реально.
- Во-вторых, когда прогноз разрабатывается, то есть прогнозирование осуществляется для некоторого момента времени в прошлом, уже есть фактические данные.

Важно!

- Проверка точности одного прогноза мало что может сказать исследователю. **Хороший единичный прогноз может быть получен и по неадекватной модели**, и наоборот, поэтому о качестве прогнозов применяемых методик и моделей можно судить лишь по совокупности сопоставлений прогнозов и их реализации.

Меры точности прогнозов

- Наиболее простой мерой качества прогнозов при условии, что есть данные об их реализации, может стать относительное число случаев, когда фактическая реализация попадала в доверительный интервал прогноза, к общему числу прогнозов, то есть

$$\eta = \frac{m}{m+n}$$

где

- m - количество прогнозов, подтвержденных фактическими данными;
- n - количество прогнозов, не подтвержденных фактическими данными.

Абсолютные критерии точности прогнозов

$$MSE = \frac{1}{n}\sum_t \left(y_t - \hat{y}_t\right)^2$$

Среднеквадратичная погрешность прогноза за n периодов.

$$RMSE = \sqrt{\frac{1}{n}\sum_t \left(y_t - \hat{y}_t\right)^2}$$

Корень из среднеквадратичной погрешности прогноза
за n периодов.

$$MAD = \frac{1}{n}\sum_t \left|y_t - \hat{y}_t\right|$$

Средняя абсолютная погрешность за n периодов.

Относительные критерии точности прогнозов

$$RMSPE = 100\sqrt{\frac{1}{n}\sum_{t}\left(\frac{y_t - \hat{y}_t}{y_t}\right)^2}$$

корень из среднеквадратичной погрешности в процентах от фактических значений за n периодов.

$$MAPE = \frac{100}{n}\sum_{t}\left|\frac{y_t - \hat{y}_t}{y_t}\right|$$

средняя абсолютная погрешность в процентах за n периодов.

Оценка точности микроэкономического прогноза

MAPE, RMSPE	**Точность прогноза**
меньше 5%	Высокая
5% – 10%	Хорошая
10% – 15%	Удовлетворительная
15% – 20%	Плохая
больше 20%	Неудовлетворительная

Эффективность прогнозирования

- количество усилий, затрачиваемых на построение модели и наличие готовых машинных программ;
- скорость, с которой метод улавливает существенные изменения в поведении ряда;
- существование серийной корреляции в ошибках;
- несменяемость первичных данных;
- полный объем работы в некоторых сферах деятельности - тысячи рядов ежемесячно требуют обновления, небольшие затраты и скорость имеют первостепенное значение;
- срочность прогнозирования.

Улучшение прогнозирования с помощью правильной подготовки данных

- Пики в данных
- Кризисные значения
- Инфляция

Метод скользящего среднего

- Этот метод является одним из самых простых, который позволяет выделить тренд. Для применения этого метода исследователь должен иметь достаточно длинный ряд наблюдений. Формально метод описывается выражением:

$$\tilde{y}_t = \frac{1}{k} \sum_{j=-k_1}^{k_2} y_{t+j}, \quad k = k_1 + k_2 + 1.$$

Метод скользящего среднего

- Для квартальных данных временного ряда формула приобретает вид

$$\tilde{y}_t = \frac{1}{4}\sum_{j=-3}^{0} y_{t+j} = \frac{1}{4}\left(y_{t-3} + y_{t-2} + y_{t-1} + y_t\right).$$

Обычное экспоненциальное сглаживание

- Лучше всего этот метод зарекомендовал себя, когда данные имеют очень гладкий, или даже горизонтальный тренд. Новый ряд строится по правилу:

$$S_t = \alpha y_t + (1-\alpha)S_{t-1}, \quad 0 < \alpha < 1.$$

- Выбор начального значения:

$$S_1 = y_1$$

Выбор константы

- Единственный параметр α может выбираться несколькими путями.
- Во-первых, если выбирается значение близко к 1, то будут более важными при прогнозировании последние данные временного ряда, при выборе близким к 0, более влиятельными будут прошлые значения.
- Во-вторых, можно положить $\alpha = \dfrac{2}{T+1}.$

Прогноз

- Прогноз значений временного ряда равен последнему члену последовательности

$$\hat{y}_{T+p} = S_T, \qquad p = 1,\ 2,\ \ldots .$$

Двойное экспоненциальное сглаживание Брауна

- Этот метод строится аналогично предыдущему, только процесс сглаживания делается дважды:

$$S_t^{/} = \alpha y_t + (1-\alpha)S_{t-1}^{/},$$

$$S_t^{//} = \alpha S_t^{/} + (1-\alpha)S_{t-1}^{//}, \quad 0 < \alpha < 1.$$

- Метод используется, когда данные временного ряда имеют тренд

Прогноз

- Прогноз строится как последнее значение второй последовательности:

$$\hat{y}_{T+p} = S_T^{//}, \qquad p = 1, \quad 2, \quad \ldots.$$

Тройное экспоненциальное сглаживание Брауна

- Этот метод аналогичен двум предыдущим, только сглаживание проводится трижды. Это позволяет прогнозировать нестационарные временные ряды с большими перепадами минимального и максимального значений. Новые последовательности строятся по правилу:

$$S_t^{/} = \alpha y_t + (1-\alpha) S_{t-1}^{/},$$

$$S_t^{//} = \alpha S_t^{/} + (1-\alpha) S_{t-1}^{//},$$

$$S_t^{///} = \alpha S_t^{//} + (1-\alpha) S_{t-1}^{///}, \quad 0 < \alpha < 1.$$

Прогноз

- Прогноз на следующие периоды имеет вид:

$$\hat{y}_{T+p} = S_T^{///}, \qquad p = 1, \quad 2, \quad \ldots .$$

Несезонная модель Холта-Винтерса

- Эта модель похожа на двойное экспоненциальное сглаживание, но позволяет выделять трендовый компонент с помощью второй последовательности:

$$S_2^{/} = y_2, \quad S_2^{//} = y_2 - y_1,$$

$$S_t^{/} = \alpha y_t + (1-\alpha)\left(S_{t-1}^{/} + S_{t-1}^{//}\right), \quad 0 < \alpha < 1,$$

$$S_t^{//} = \beta\left(S_t^{/} - S_{t-1}^{/}\right) + (1-\beta) S_{t-1}^{//}, \quad 0 < \beta < 1.$$

Прогноз

- Прогноз на следующие периоды:

$$\hat{y}_{T+p} = S_T^{/} + pS_T^{//}, \qquad p = 1, \quad 2, \quad \ldots .$$

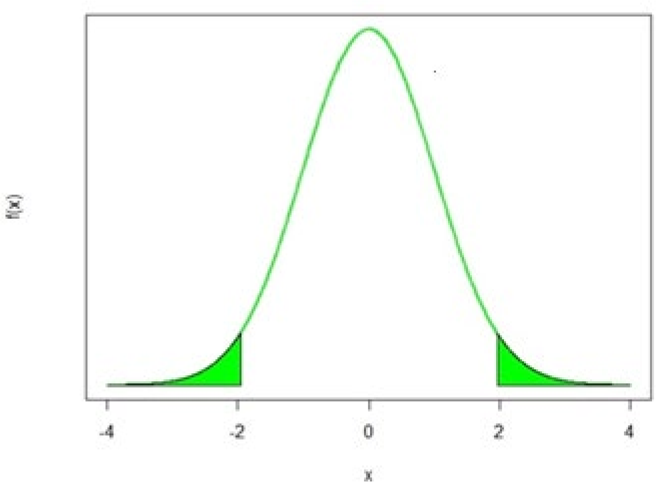
- Следует отметить, что прогнозы, сделанные по этому методу, как правило, либо сильно завышены, или занижены вследствие того, что добавляется постоянный трендовый компонент, который на практике меняется в течение года.

Доверительные интервалы

- Доверительный интервал - это допустимое отклонение наблюдаемых значений от истинных. Размер этого допущения определяется исследователем с учетом требований к точности информации.

Интервальная оценка

- Как правило, доверительный уровень обозначают следующим образом: (1-α)x100%, где величина α представляет собой площадь, ограниченную хвостом распределения, выходящим за пределы доверительного интервала.
- Величину α называют уровнем значимости доверительного интервала.
- Площади, ограниченные как левым, так и правым хвостами распределения, выходящими за пределы доверительного интервала, равны α/2.



Доверительный интервал

$$\hat{y}_{T+n} - t_\alpha \sigma_p \le y_{T+n} \le \hat{y}_{T+n} + t_\alpha \sigma_p$$

- T - длина временного ряда;
- n -период упреждения;
- $\hat{y}_{T+n}$ -точечный прогноз на момент T+n;
- t_a- значение t-статистики Стьюдента;
- σ_p - средняя квадратическая ошибка прогноза.

Сезонность во временных рядах

- Получение коэффициента сезонности: нахождение среднего отношения для каждого квартала (месяца, недели, дня) к среднегодовому (среднеквартальному, среднемесячному, средненедельному) значению.

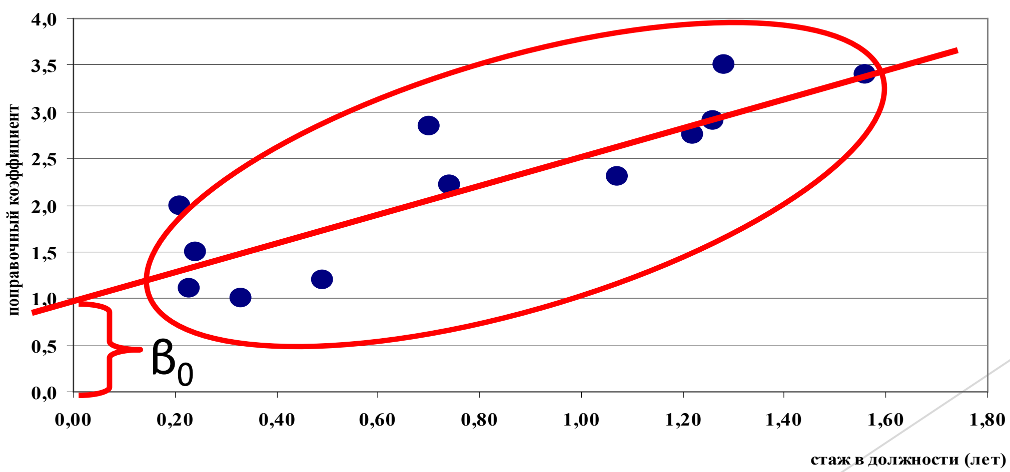
Регрессионный анализ

- Даёт нам правила, определяющие линию регрессии, которая лучше других предсказывает одну переменную на основании другой.
- По оси Y располагают переменную, которую мы хотим предсказать, а по оси X – переменную, на основе которой будем предсказывать.
- Предсказанное значение Y обычно обозначают как $\hat{Y}$

Простая линейная регрессия

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

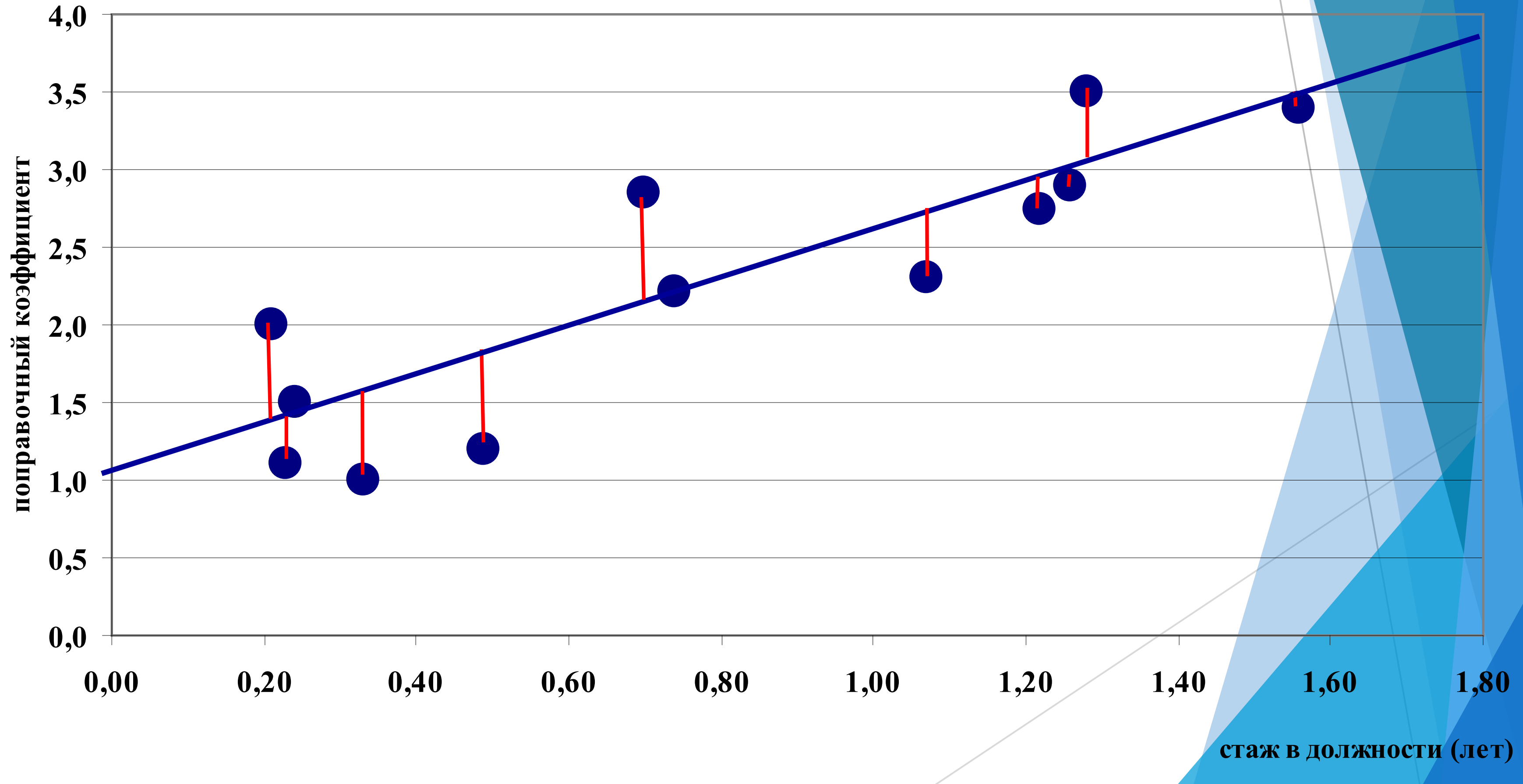
- Y – зависимая переменная
- X – независимая переменная
- β_0 и β_1 - коэффициенты регрессии
- β_1 – характеризует НАКЛОН прямой;
- β_0 – определяет точку пересечения прямой с осью OY.



«Лучшая» линия регрессии

$$e_t = Y_t - \widehat{Y}_t$$

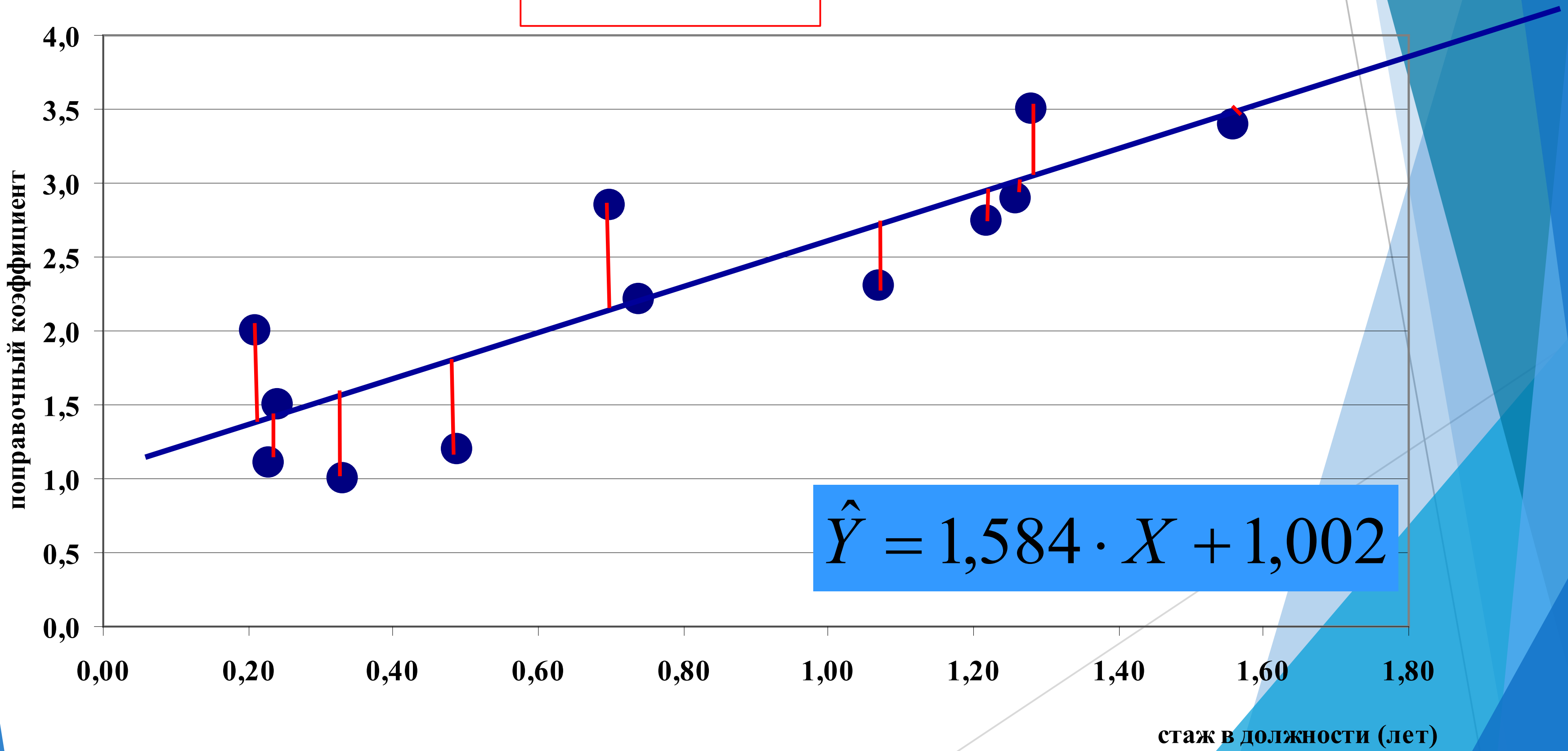
$$\sum e_t = 0$$



Метод наименьших квадратов

- линию регрессии подбирают такую, чтобы общая сумма квадратов ошибок (residuals) была наименьшей.

$$\sum e_t^2 \rightarrow \min$$



$$\hat{Y} = 1{,}584 \cdot X + 1{,}002$$

Линейная регрессия в MS Excel

- **Способ 1.** Линия тренда на графике
- Построить точечный график по имеющимся данным вида (y_i, x_i)
- Щелкнуть правой кнопкой мыши на серии точек и выбрать «добавить линию тренда»
- Нагляден, но не проводятся F-тест и оценка доверительных интервалов коэффициентов регрессии

- **Способ 2.** Использование пакета анализа данных
- Выбрать вкладку данные, щелкнуть по пункту меню «анализ данных»
- Из предлагаемых опций выбрать регрессию
- Указать входные данные и изучить результат
- Нагляден, содержит F-тест и оценку доверительных интервалов коэффициентов регрессии

Описание регрессии

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,996
R-квадрат	0,992
Нормированный R-квадрат	0,991
Стандартная ошибка	121,674
Наблюдения	25

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*ачимость F*
Регрессия	1	39785221,53	39785221,53	2687,367	2,5E-25
Остаток	23	340504,3072	14804,5351		
Итого	24	40125725,84			

	Коэффициен	*андартная ошиб*	*t-статистика*	*P-Значение*	*ижние 95*%	*рхние 95*%	*ижние 95,0*	*рхние 95,0%*
Y-пересечение	-141,41	67,96105028	-2,080753151	0,0488	-281,998	-0,82203	-281,998	-0,82203
Средний доход людей, грн.	0,77	0,014880843	51,83982221	0,0000	0,74064	0,8022	0,74064	0,8022

Виды тренда

Линейный $f(t) = a_0 + a_1 t$

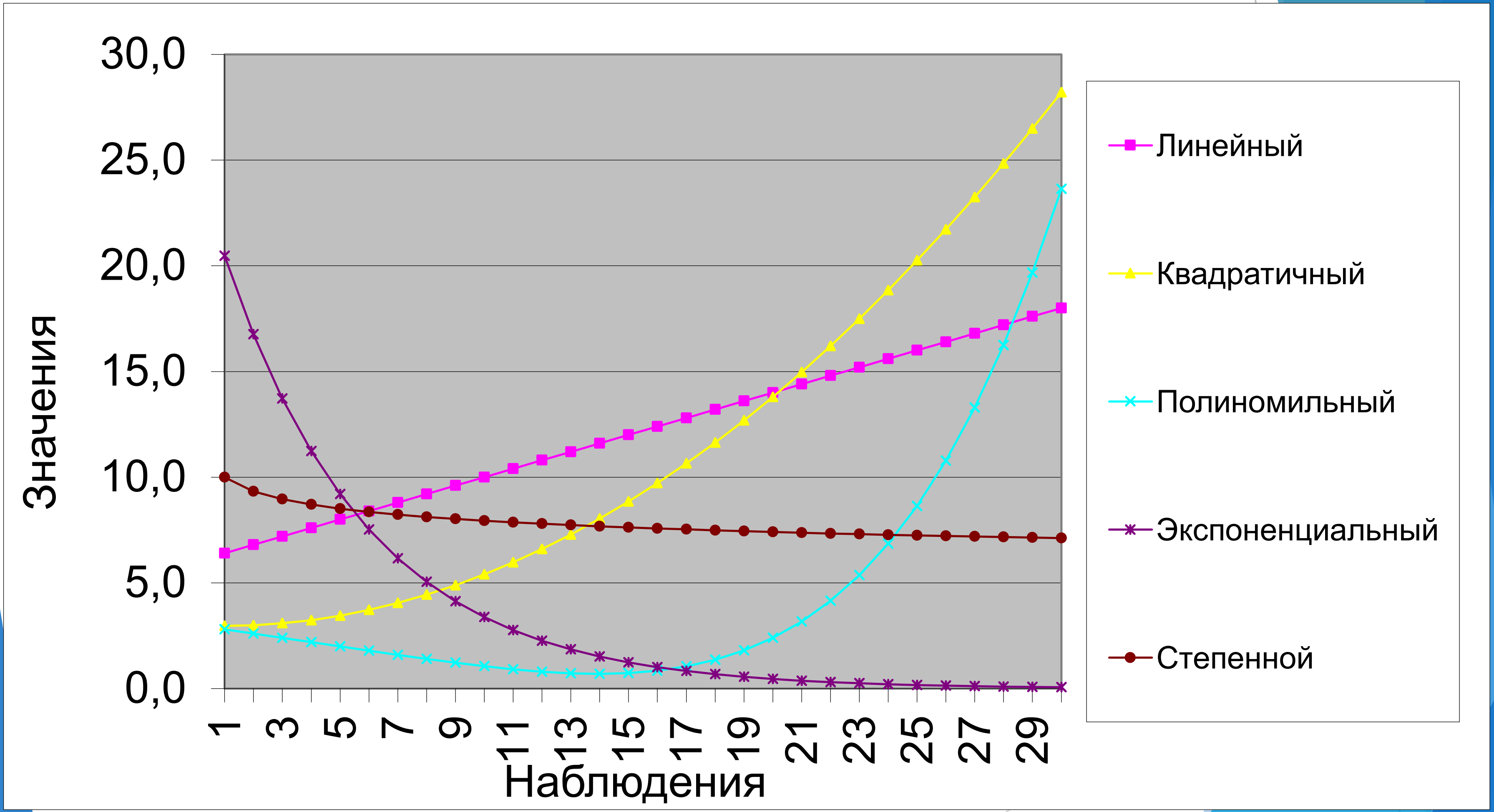
Квадратичный $f(t) = a_0 + a_1 t + a_2 t^2$

Полиномиальный $f(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n$

Экспоненциальный $f(t) = a_0 e^{a_1 t}$

Степенной $f(t) = a_0 t^{a_1}$

Пример



30,0
25,0
20,0
15,0
10,0
5,0
0,0
Значения
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
Наблюдения
Линейный
Квадратичный
Полиномильный
Экспоненциальный
Степенной

Множественная регрессия

Изучение:

- проблем спроса,
- доходности акций,
- изучение функции издержек производства, в микроэкономических расчетах.

Учет влияния нескольких факторов

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_{k-1} x_{k-1t} + \varepsilon_t, t = \overline{1, n}$$

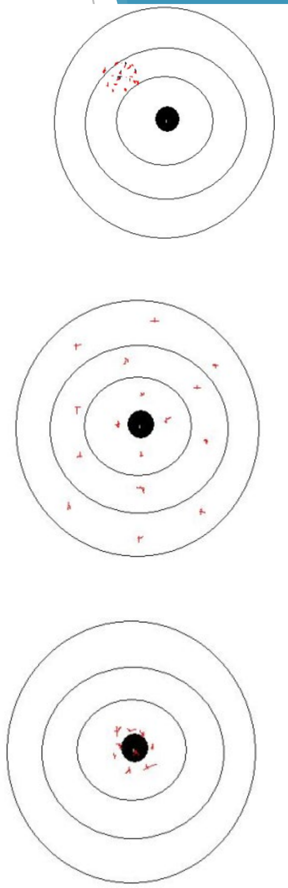
y_t - зависимая переменная;

$x_{1t}, x_{2t}, \ldots, x_{k-1t}$ - независимые переменные;

ε_t - остатки.

Желательные характеристики

- **Отсутствие смещения**
- **Эффективность**
  - Стандартные ошибки будут минимальными
- **Консистентность**
  - При увеличении числа наблюдений стандартная ошибка уменьшается.



Основная цель множественной регрессии

- построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

Пример

- Функция спроса чаще всего рассматривается как модель вида

$$C = f(y, P, M, Z),$$

- C – спрос;
- y – доход;
- P – уровень цен;
- M – наличные деньги;
- Z – ликвидные активы;

Спецификация

- Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели.

Условия включения факторов при построении множественной регрессии

- Они должны быть количественно измеримы. Если необходимо включить модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность.
- Факторы не должны быть интеркоррелированы. Если между факторами существует высокая корреляция, то нельзя определить их изолированное влияние на результативный показатель и параметры уравнения регрессии оказываются неинтерпретируемыми.

Отбор факторов при построении множественной регрессии

- На первой стадии подбираются факторы исходя из сущности проблемы;

- на второй – на основе матрицы показателей корреляции определяют существенность включения в уравнение регрессии каждого из факторов. Если факторы явно зависимы, то они дублируют друг друга и один из них рекомендуется исключить из регрессии.

Пример

	Y	X1	X2	X3
Y	1			
X1	0,8	1		
X2	0,7	0,8	1	
X3	0,6	0,5	0,2	1

Очевидно, что факторы X1 и X2 дублируют друг друга.

Мультиколлинеарность – 1

- Для оценки мультиколлинеарности факторов может использоваться определитель матрицы парных коэффициентов корреляции между факторами.
- Если бы факторы не коррелировали между собой, то матрица парных коэффициентов корреляции была бы единичной матрицей т.е.

$$Det|R| = \begin{vmatrix} r_{x_1x_1} & r_{x_2x_1} & r_{x_3x_1} \\ r_{x_1x_2} & r_{x_2x_2} & r_{x_3x_2} \\ r_{x_1x_3} & r_{x_2x_3} & r_{x_3x_3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1,$$

Мультиколлинеарность – 2

- Если же, наоборот, между факторами существует полная линейная зависимость и все коэффициенты корреляции равны единице, то определитель такой матрицы равен нулю:

$$Det|R| = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

Задание:

- На основе приводимых данных (закладка «Мороженое») построить регрессионную модель
- Выделить значимые факторы (объяснить почему)
- Построить функцию спроса с учетом только значимых параметров и на ее основе определить уровень цен, при которых доход фирмы будет максимальным
- Рассчитать значения эластичности по цене для уровня цены=25

Эластичность

$$E_p = \frac{\partial Q}{\partial P} \cdot \frac{P}{Q}$$

Определение цены

$$\Pi = Q \cdot P = (b_0 + b_1 P) P \to \max$$

$$\frac{\partial \Pi}{\partial P} = b_0 + 2b_1 P = 0$$

$$P = -\frac{b_0}{2b_1}$$

Выделение сезонности с помощью фиктивных переменных

- Фиктивная переменная – величина, которая принимает значений 1 при наличии некоторого признака, 0 – при его отсутствии.

Регрессии

$$y_t = \beta_0 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_3 + \varepsilon_t$$

$$y_t = \beta_0 + \beta_1 q_1 + \beta_2 q_2 + \beta_3 q_3 + \beta_4 t + \varepsilon_t$$

Фиктивные переменные

$$q_1 = (1,0,0,0,1,0,0,0,1,0,\ldots)^/$$

$$q_2 = (0,1,0,0,0,1,0,0,0,1,0,\ldots)^/$$

$$q_3 = (0,0,1,0,0,0,1,0,0,0,1,0,\ldots)^/$$

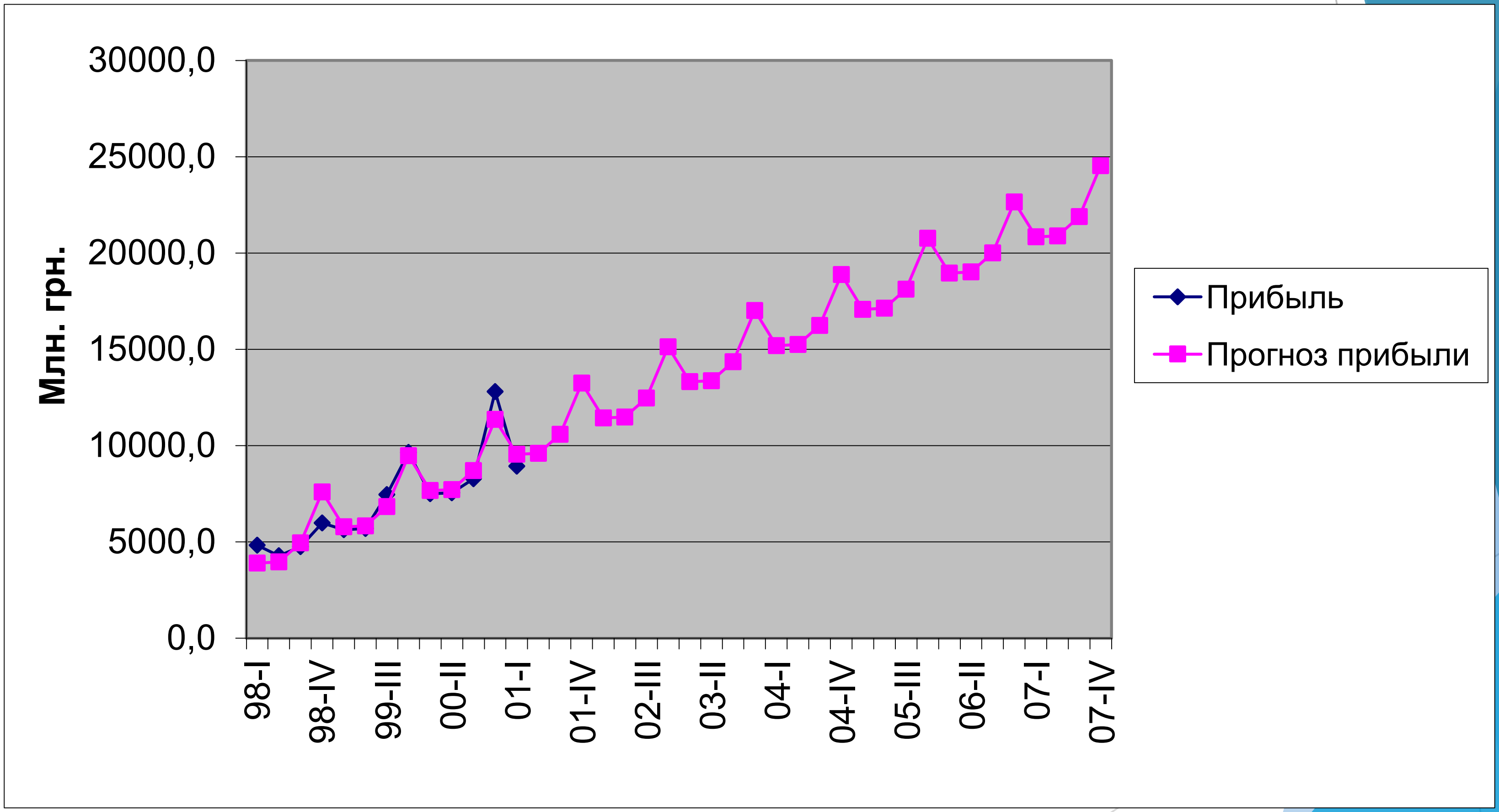
1 квартал $\quad y_t = \beta_0 + \beta_1 + \varepsilon_t$

2 квартал $\quad y_t = \beta_0 + \beta_2 + \varepsilon_t$

3 квартал $\quad y_t = \beta_0 + \beta_3 + \varepsilon_t$

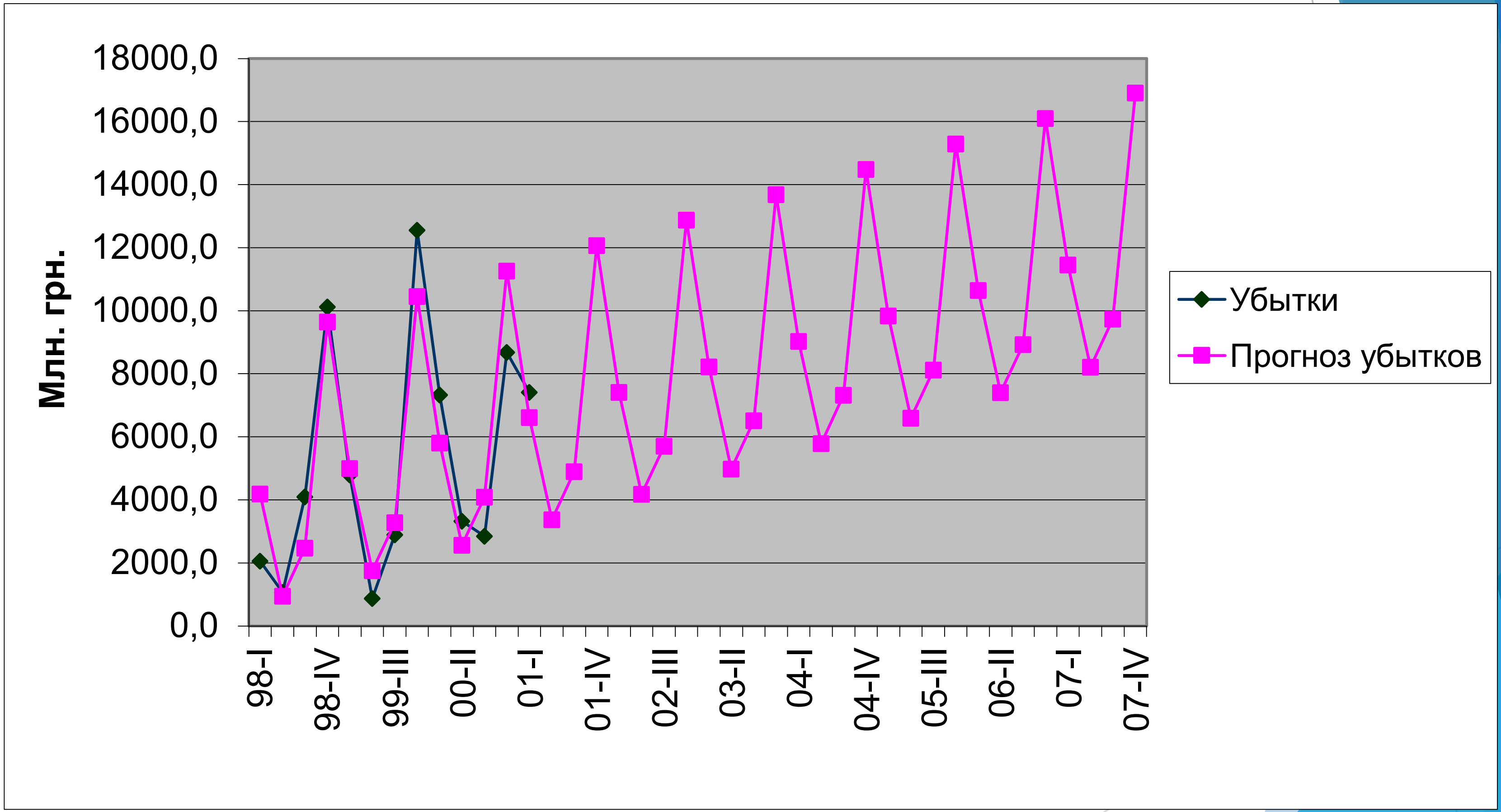
4 квартал $\quad y_t = \beta_0 + \varepsilon_t$

Пример – 1



30000,0
25000,0
20000,0
15000,0
10000,0
5000,0
0,0
Млн. грн.
98-I
98-IV
99-III
00-II
01-I
01-IV
02-III
03-II
04-I
04-IV
05-III
06-II
07-I
07-IV
Прибыль
Прогноз прибыли

Пример – 2



18000,0
16000,0
14000,0
12000,0
10000,0
8000,0
6000,0
4000,0
2000,0
0,0
Млн. грн.
98-I
98-IV
99-III
00-II
01-I
01-IV
02-III
03-II
04-I
04-IV
05-III
06-II
07-I
07-IV
Убытки
Прогноз убытков

Применение фиктивных переменных

- Выделение сезонных колебаний
- Нивелирование кризисных явлений
- Моделирование качественных показателей

Создание фиктивных сезонных переменных

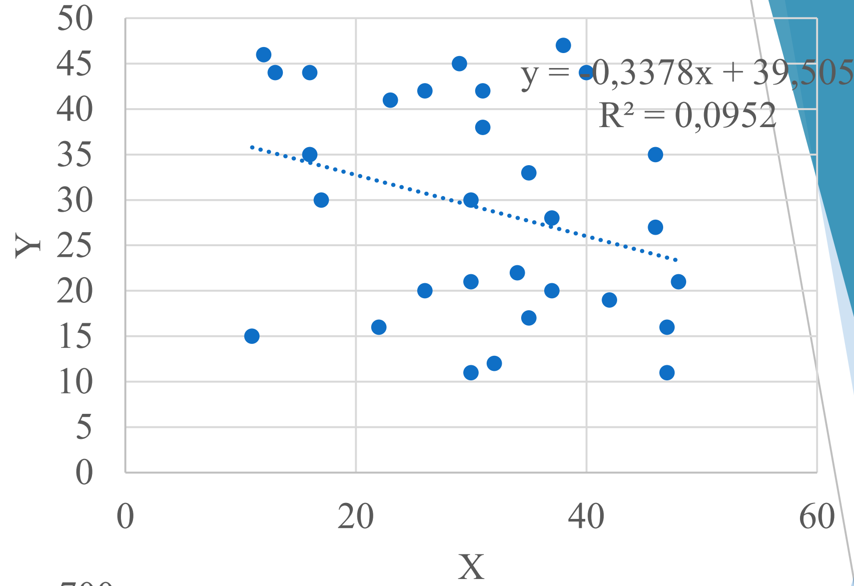
- Функции Month/Месяц, WeekDay/ДеньНед
- Вручную

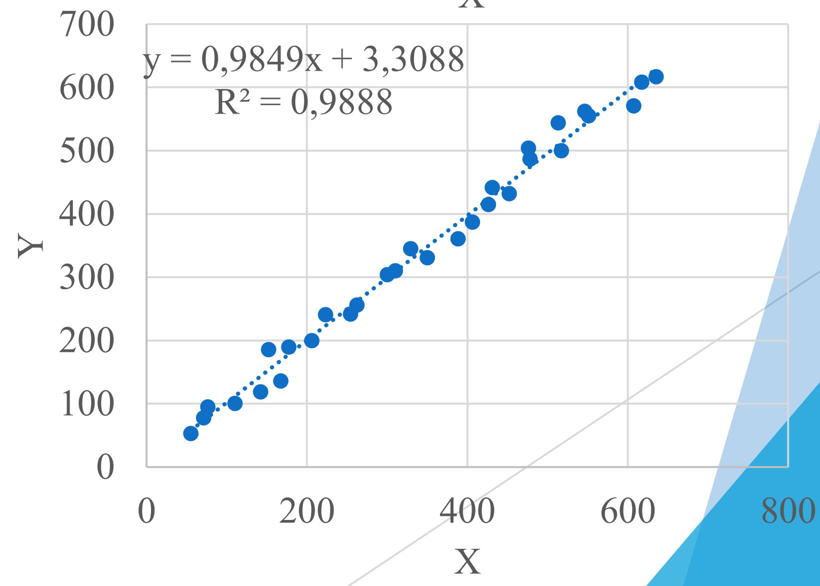
Стационарный процесс

- Случайный процесс X (t) называется стационарным, если все его вероятностные характеристики не меняются со временем t.

Почему стационарность важна?

№	X	Y	T	X*	Y*
1	35	33	20	55	53
2	31	38	40	71	78
3	16	35	60	76	95
4	30	21	80	110	101
5	42	19	100	142	119
6	47	16	120	167	136
7	12	46	140	152	186
8	17	30	160	177	190
9	26	20	180	206	200
10	23	41	200	223	241
11	34	22	220	254	242
12	22	16	240	262	256
13	40	44	260	300	304
14	30	30	280	310	310
15	29	45	300	329	345
16	30	11	320	350	331
…					
30	35	17	600	635	617





Тест на стационарность

- В общем случае модель сводим к виду

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$$

- и проверяем гипотезу

$$H^0: \quad \gamma = 0$$

$$H^1: \quad \gamma < 0$$

Возможные проблемы

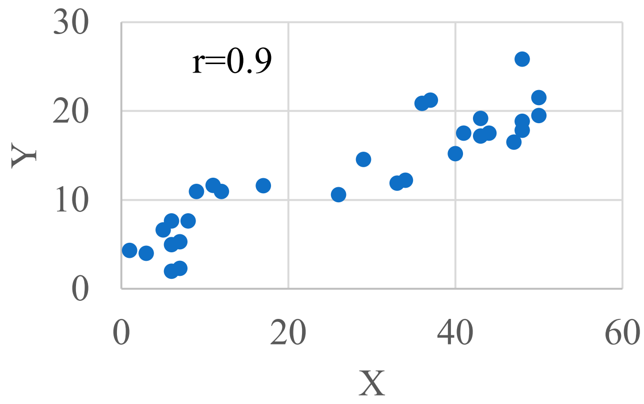
- Стойкость модели
- Распределение остатков
- Использование лагов
- Пропуски данных

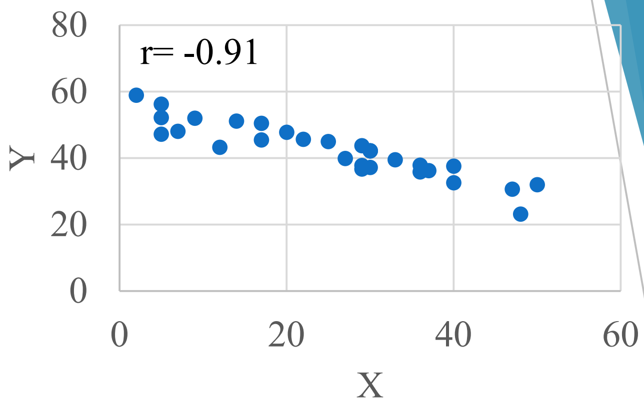
Сравнение моделей и выбор оптимальной

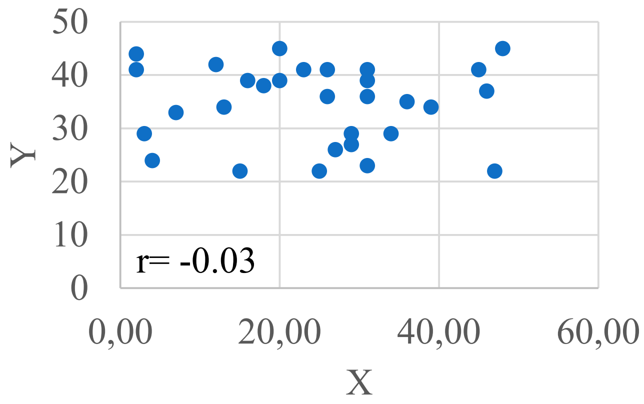
- В основном, сравниваем скорректированный коэффициент детерминации.
- Он равен квадрату коэффициента корреляции.

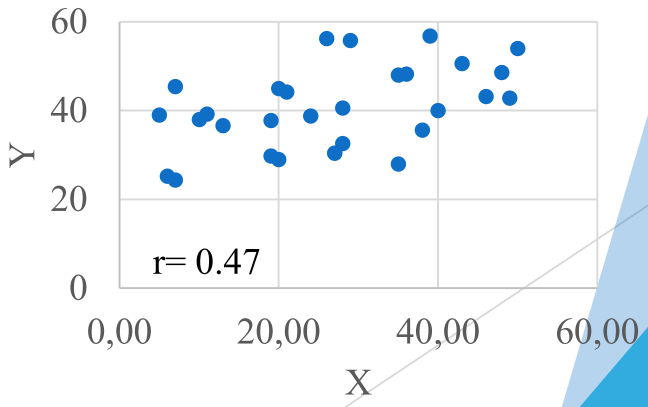
Коэффициент корреляции

- Может принимать значения от -1 до +1
- Знак коэффициента показывает направление связи (прямая или обратная)
- Абсолютная величина показывает силу связи
- Всегда основан на парах чисел


r=0.9
30
20
10
0
0
20
40
60
Y
X

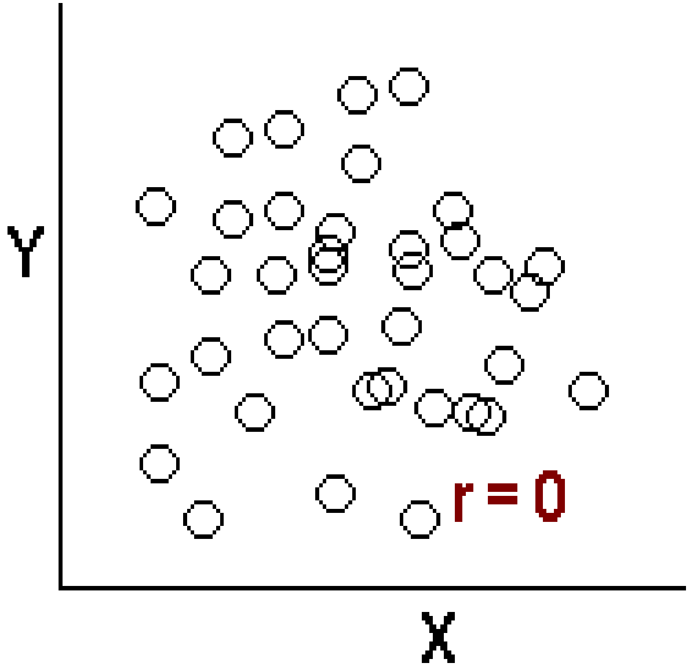

r= -0.91
80
60
40
20
0
0
20
40
60
Y
X


r= -0.03
50
40
30
20
10
0
0,00
20,00
40,00
60,00
Y
X


r= 0.47
60
40
20
0
0,00
20,00
40,00
60,00
Y
X

Но!

- Создается впечатление, что близкий к нулю коэффициент корреляции говорит о том, что связи между переменными нет или почти нет.

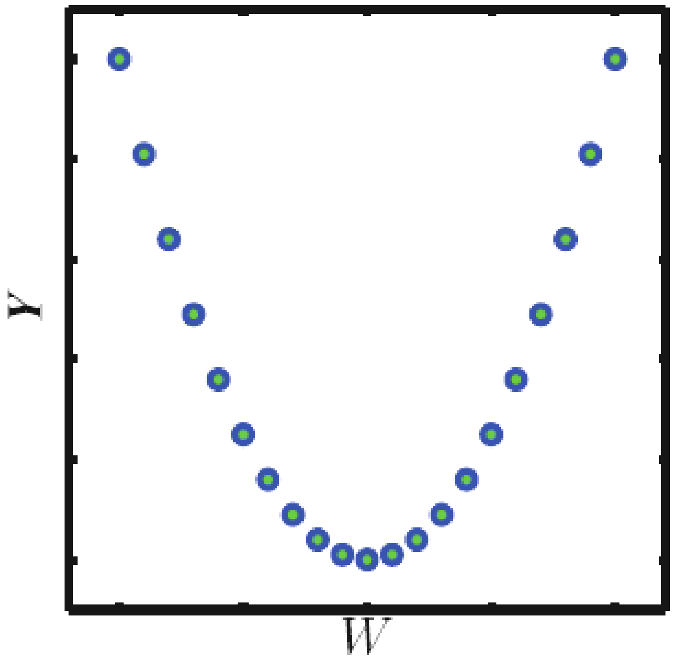


Здесь ее нет

Но это не всегда так, есть исключения!

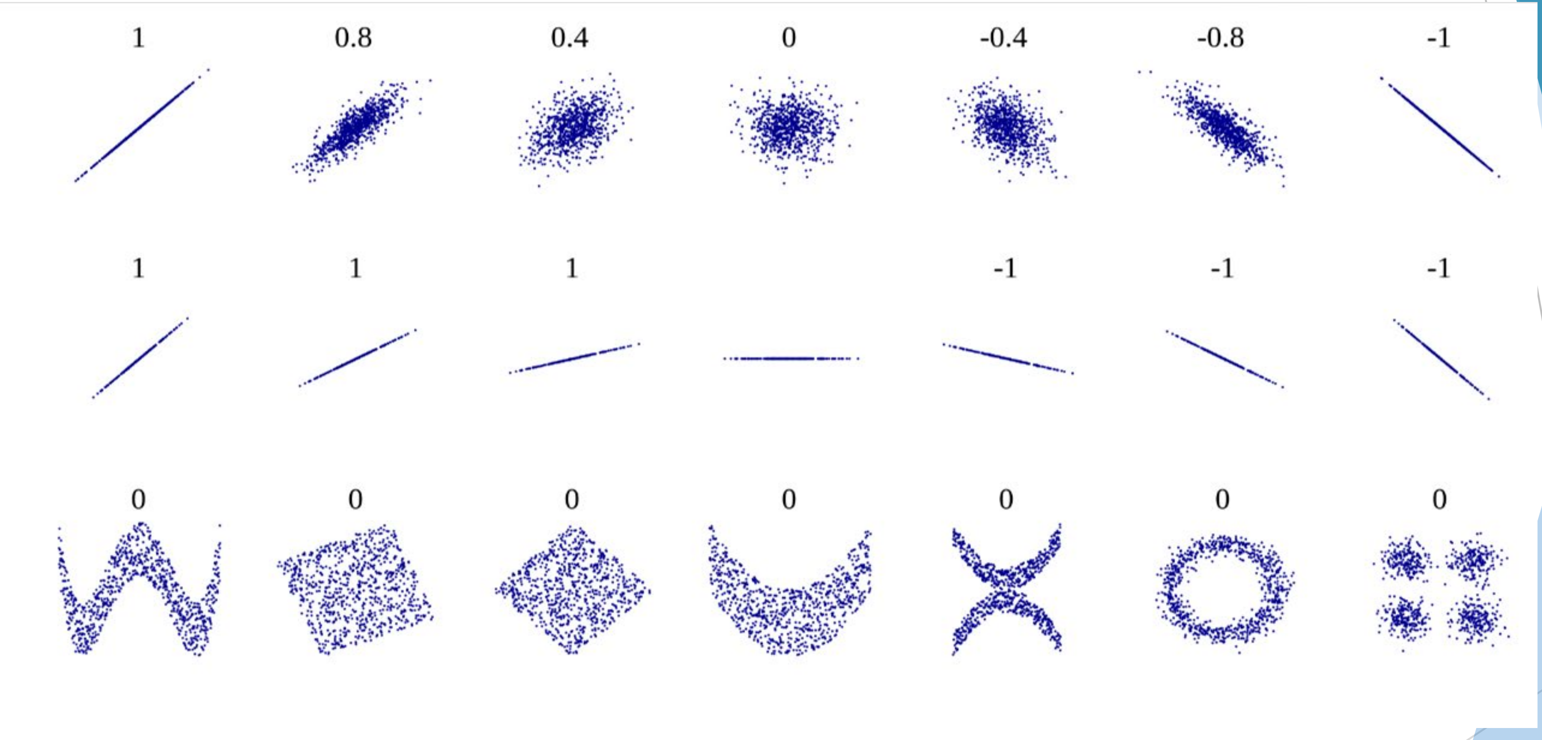
Недостаток коэффициента корреляции

- Коэффициент корреляции оценивает только линейная связь переменных!
- Он не показывает наличие нелинейной связи!



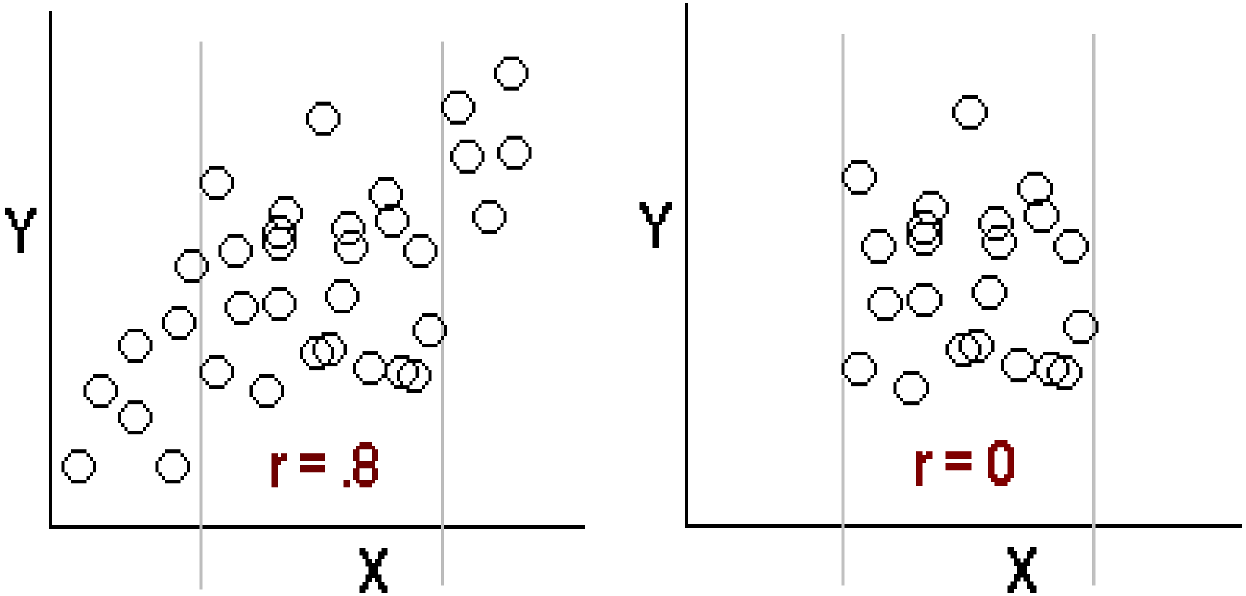
Здесь связь переменных очень сильная, но r = 0.00

Примеры

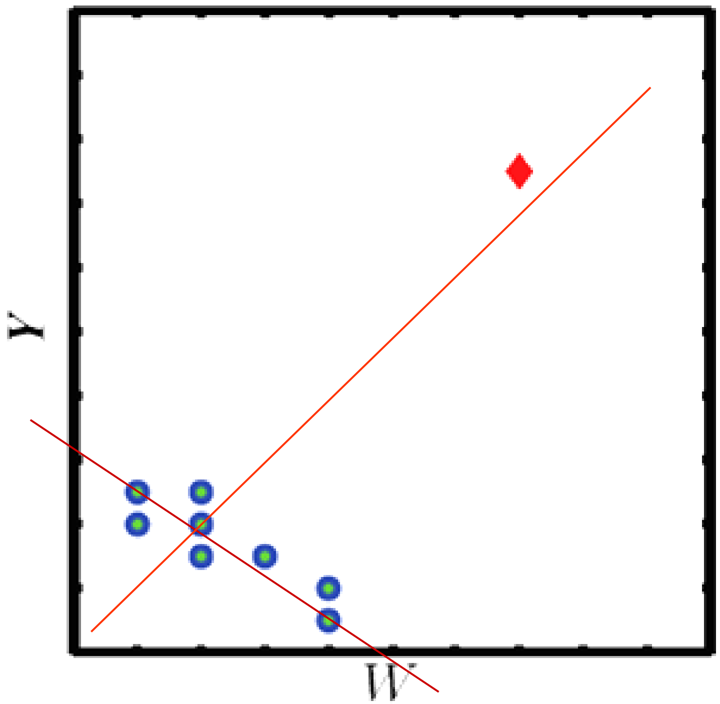


Помним!

- Необходимо, чтобы в переменных была значительная изменчивость! Если сформировать выборку из однотипных лиц, надеяться обнаружить там корреляцию не стоит.



Выброс



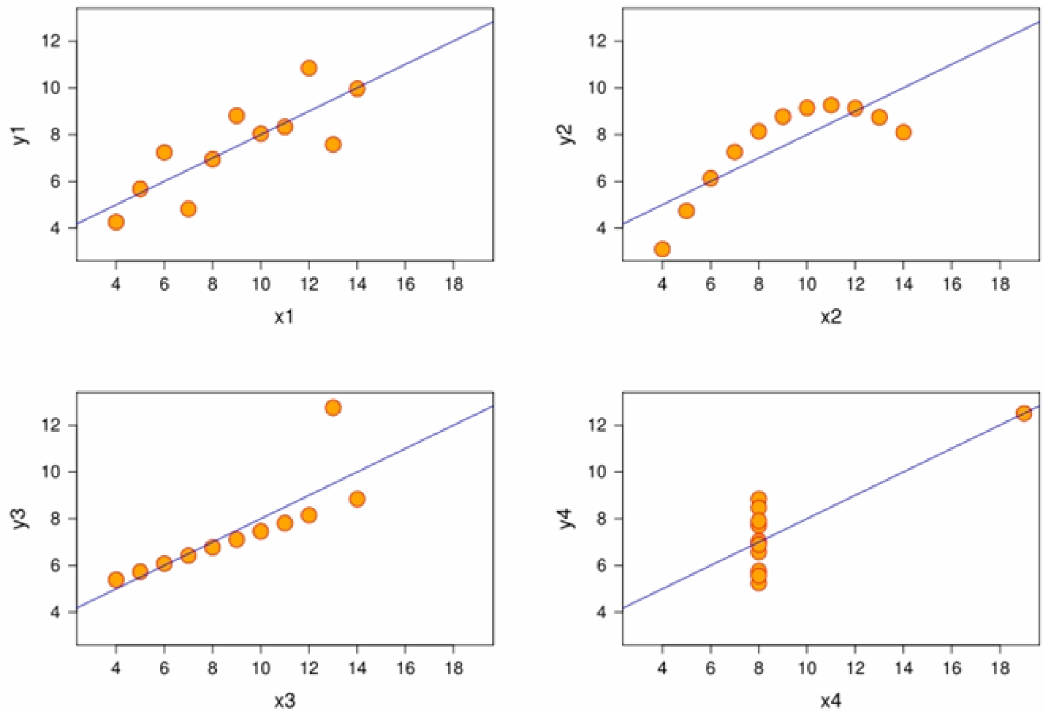
Важно!

- Корреляция совершенно не подразумевает наличие причинно-следственной связи!

- Она ВООБЩЕ НИЧЕГО о нем НЕ ГОВОРИТ (даже очень большой r)

Решение о зависимости

- Коэффициент корреляции - параметр выборки. Можно ли на его основе судить о популяции?
- Просто глядя на коэффициент - НЕТ.



Корреляция между переменными везде r = 0.816

Прогнозирование на основании ведущих экономических индикаторов

- опережающие
- совпадающие
- запаздывающие индикаторы

Использование рыночных тенденций для прогнозирования

- Индексы настроения потребителей
- Учетная ставка НБУ
- Рост денежной массы
- Изменение внутреннего долга
- Изменение платежного баланса
- Baltic Dry Index (BADI)
- Oil WTI Futures
- US Dollar Index Futures

Прогнозирование спроса нескольких товаров

- Условное форматирование
- Написание макросов
- Формирование листа формул

Что мы сделали?

- Прогнозирование – важный момент для формирования закупок.

Рассмотрены методы:

- методы анализа и прогнозирования временных рядов;
- каузальные (причинно-следственные) методы (попытка найти факторы, определяющие поведение прогнозируемого показателя);
- Анализ качественных показателей.

Спасибо за внимание!

Доц., к.э.н. Ставицкий А.В.

Email: a.stavytskyy@gmail.com