

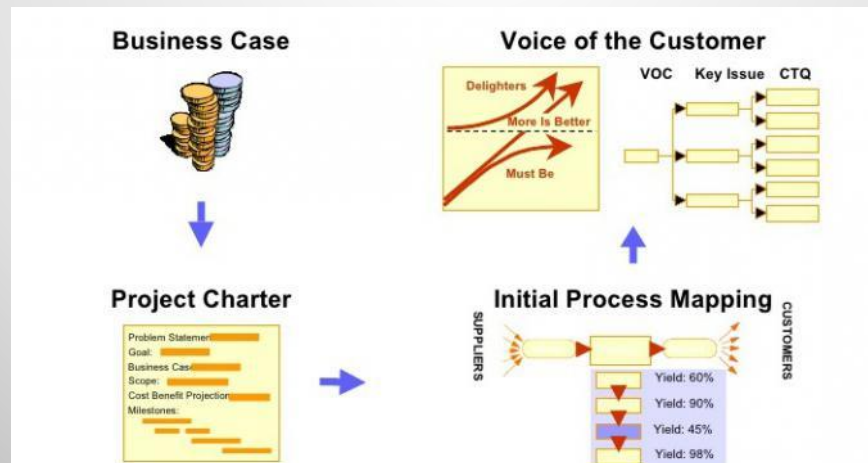
"Корреляционно- регрессионный анализ"

Выполнила: группа 14-ГУ-8А

Значение и основные этапы процесса корреляционно-регрессионного анализа экономических явлений.

- **Корреляционно-регрессионный анализ** — это один из способов решения задач и поиска информации.

Он позволяет определить совместное влияние множества взаимосвязанных и одновременно действующих признаков, а также отдельное влияние каждого признака на экономическое явление (процесс).



Этапы анализа:

1. Определение аргументов и предварительная обработка условной информации.
2. Определение тесноты и формы взаимосвязи между несколькими признаками.
3. Моделирование представленного экономического процесса и анализ полученной модели.
4. Применение конечных результатов для усовершенствования планирования и менеджмента модели.-

Simple Project Charter			
Business Case			Opportunity Statement
Goal statement			Project scope
• Goal • cost saving • metric			Process: Start: End: In scope: Out of scope:
Project Plan			Project Team
Phase	Start	End	Champion:
Define			Project Owner:
Measure			MBB:
Analyze			Finance Rep:
Improve			Green Belt:
Control			Member:

Корреляционно-регрессионный анализ: параметры созданной модели

- Парные коэффициенты корреляции

$$r_{x,y} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$$

С помощью парного линейного коэффициента корреляции выявляется связь между двумя признаками, один из которых можно рассматривать как результативный, другой — как факторный. Но в действительности на результат воздействуют несколько факторов.

- Частный, или чистый, коэффициент корреляции

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{x_1y} * r_{x_1x_2}}{\sqrt{(1 - r_{x_1y}^2) * (1 - r_{x_1x_2}^2)}}$$

остаточная дисперсия(остаточная сумма квадратов) = S2

$$S_{yx_1}^2 = \frac{\sum (y_i - \hat{y}_{x_1})^2}{n}.$$

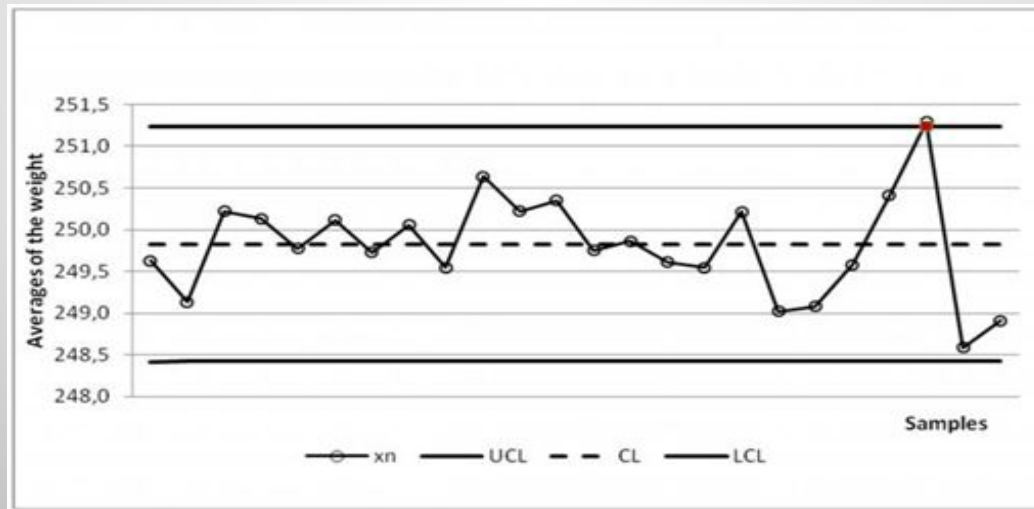
- Если выразить остаточную дисперсию через показатель детерминации $S^2_{\text{остат}} = \sigma^2_y(1 - r^2)$, то формула коэффициента частной корреляции примет вид:

$$r_{yx_1 \cdot x_2} = \sqrt{\frac{S^2_{yx_2} - S^2_{yx_1 x_2}}{S^2_{yx_2}}} = \sqrt{1 - \frac{S^2_{yx_1 x_2}}{S^2_{yx_2}}} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2}}{1 - r^2_{yx_2}}}.$$

- Коэффициент множественной корреляции (R) характеризует тесноту связи между результативным показателем и набором факторных показателей:

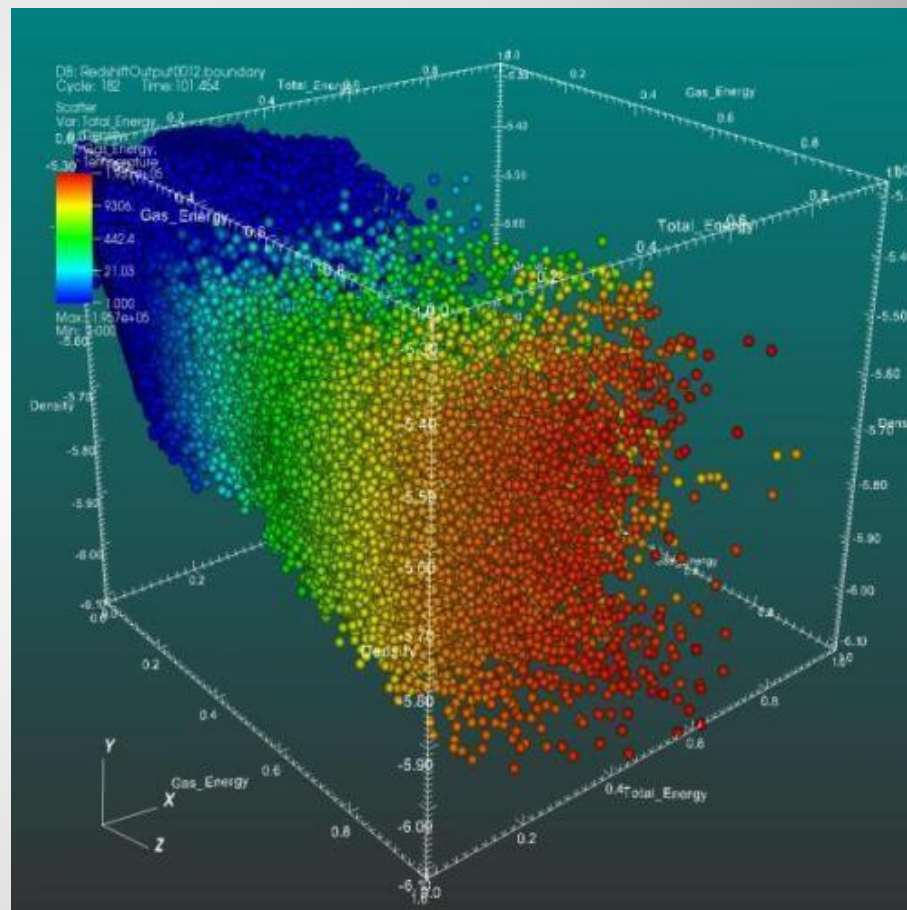
$$R = \sqrt{\frac{\sigma^2 - \sigma_{ост}^2}{\sigma^2}}, \quad \text{или} \quad R = \sqrt{1 - \frac{\sum (y - \bar{s})^2}{\sum (y - \bar{y})^2}},$$

- Коэффициенты частной детерминации (показывают влияние вариации аргумента на вариацию искомого признака).
- Коэффициент множественной детерминации (показывает удельный вес всех аргументов на вариацию искомого признака).
- Частные коэффициенты эластичности (характеризуют влияние факторов на результат, выраженное в едином масштабе в процентах).

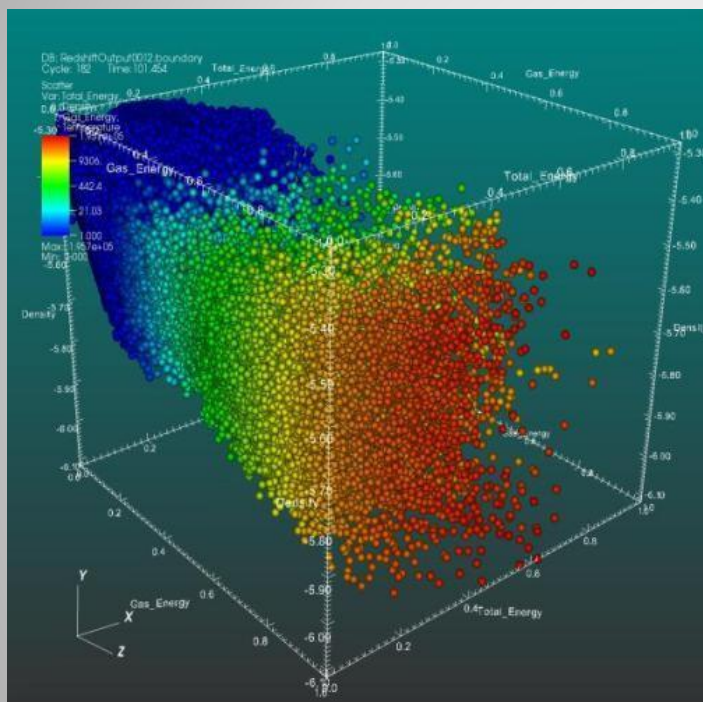


Цель анализа

- Основные задачи корреляционно-регрессионного анализа – это выявление факторов, существенно влияющих на экономический результат явления или процесса, и использование полученной информации для усовершенствования планирования экономического процесса или явления.



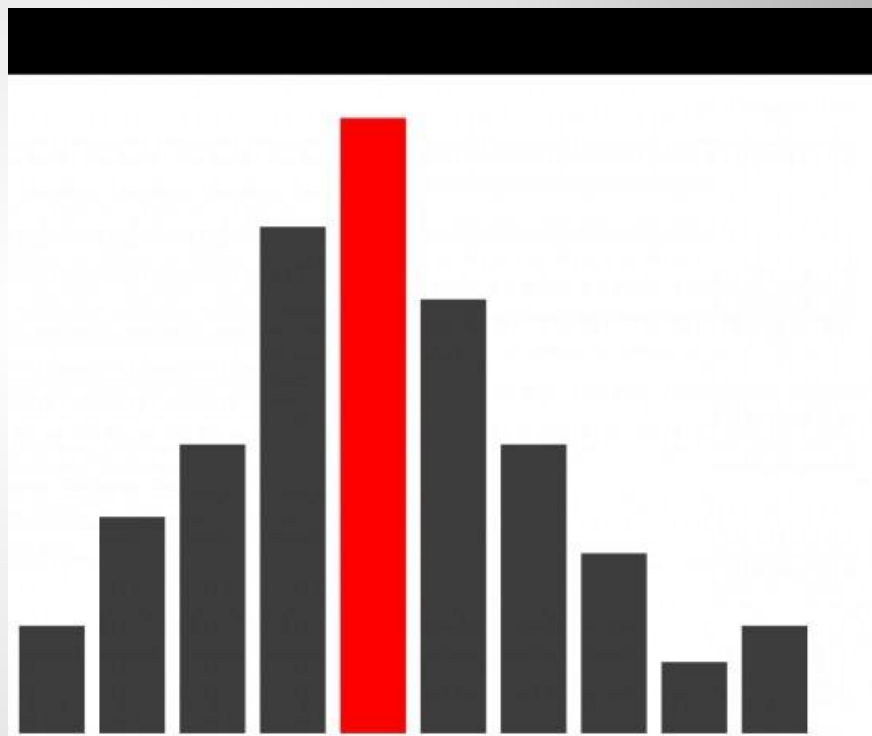
● Параметрические методы анализа.



Все производственные процессы находятся в тесной взаимосвязи. Эта взаимосвязь бывает стохастической (результат зависит от множества факторов) и функциональной (результат изменяться на такую же величину, как и фактор). Стохастическая зависимость чаще всего имеет корреляционный характер, то есть значению фактора одновременно соответствует несколько значений результата, имеющих абсолютно разные направления.

Корреляционная решетка

- Корреляционная взаимосвязь может иметь один или несколько факторов-признаков, обладать положительной или отрицательной направленностью, быть прямолинейной или криволинейной (в зависимости от выражения).



Непараметрические методы анализа

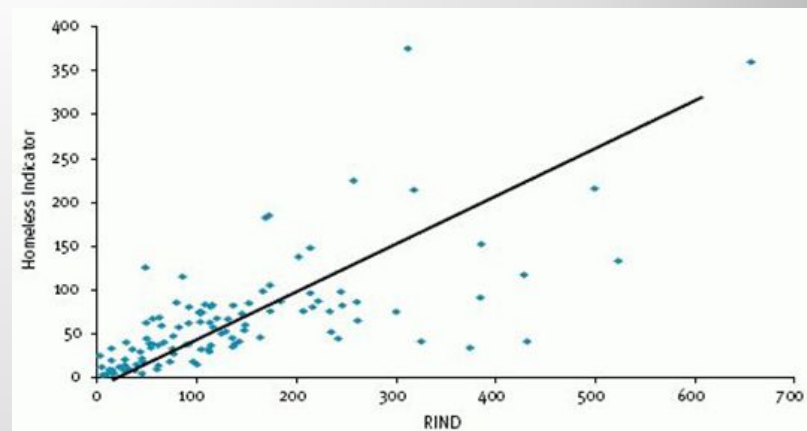
Таблица 12.1 Данные для расчета коэффициента Фехнера.

Магазин	Число работников, тыс. чел.	Товарооборот, у.е.	Отклонение от средних $\bar{x} = 1/5 = 0.2$ и $\bar{y} = 20/5 = 4$		Сравнение знаков ε_{xk} и ε_{yk}	
k	x_k	y_k	$\varepsilon_x = x - \bar{x}$	$\varepsilon_y = y - \bar{y}$	совпадение (C_k)	несовпадение (H_k)
1	0,2	3,1	+0,0	-0,9	0	1
2	0,1	3,1	-0,1	-0,9	1	0
3	0,4	5,0	+0,2	+1,0	1	0
4	0,2	4,4	+0,0	+0,4	1	0
5	0,1	4,4	-0,1	+0,4	0	1
Итого	1,0	20,0	-	-	3	2

По (1) имеем $K_{\Phi} = (3 - 2)/(3 + 2) = 0,20$. Направление взаимосвязи в вариациях II (Средняя численность работников)

● Коэффициент корреляции знаков Фехнера (определяет количество совпадений и несовпадений отклонений аргументов и результатов от их среднего значения).
Еще один немаловажный метод корреляционно-регрессионного анализа — Метод наименьших квадратов, позволяющий определить аналитическое выражение взаимосвязи результативного признака и его фактора. Он заключается в построении системы уравнений и определении параметров этих уравнений.

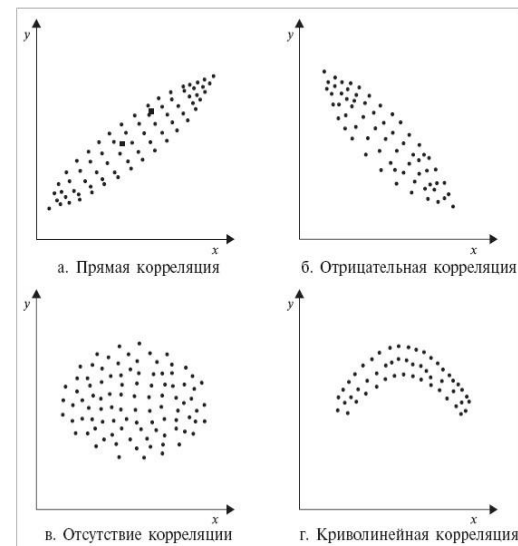
- Коэффициент ранговой корреляции Кендалла (определяет взаимосвязь количественных и качественных значений показателей, в случае если они подлежат ранжированию).
- Коэффициент ранговой корреляции Спирмена (присваивает ранги каждому аргументу и результату, на основе которых определяются разности и вычисляется показатель).



Статистические методы

- Основой статистических методов корреляционный и регрессионный анализ является один из семи простых инструментов контроля качества - диаграмма разброса (поле корреляции). Этот инструмент позволяет графически отобразить и в дальнейшем проанализировать вид и тесноту связи между исследуемыми факторами.

Диаграммы разброса



Построение поля корреляции сводится к следующим этапам:

- 1. Сбор не менее 25 пар данных исследуемых параметров в таблицу;
- 2. Нахождение максимальных и минимальных значений и . Выбор шкалы на горизонтальной и вертикальной оси так, чтобы длины рабочих областей были примерно равны.
- 3. Построение на отдельном листе координатной плоскости. Если исследуется влияние фактора на показатель качества, то фактор располагают по оси абсцисс, а показатель – по оси ординат; и нанесение собранных пар данных (в случае совпадения точек они либо располагаются максимально близко, либо обозначаются окружностями около первоначальной точки)
- 4. На диаграмму наносятся все необходимые обозначения:
 - название диаграммы;
 - интервал времени сбора данных;
 - число пар данных;
 - название и единицы для каждой оси;
 - идентифицирующая информация составителя диаграммы



- Если точки корреляционного поля образуют эллипс, главная диагональ которого имеет положительный угол наклона, то имеет место положительная корреляция (пример подобной ситуации можно видеть на рисунке 6.3).

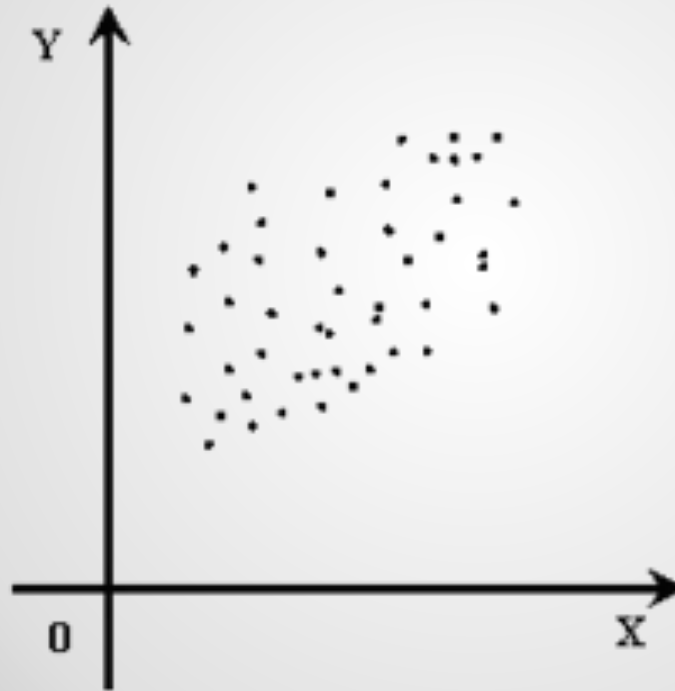


Рис. 6.3. Положительная корреляция

- Если точки корреляционного поля образуют эллипс, главная диагональ которого имеет отрицательный угол наклона, то имеет место отрицательная корреляция (пример изображен на рисунке 6.4).

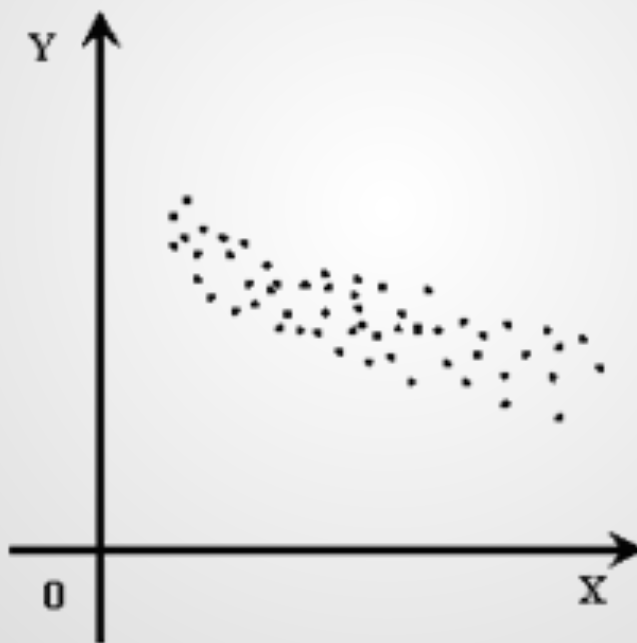


Рис. 6.4. Отрицательная корреляция

Экономический барометр

-это система экономических показателей, применяемых для анализа и прогнозирования конъюнктуры рынка.



Все показатели делятся на три группы: опережающие (лидирующие), совпадающие (синхронные) и запаздывающие.

Прогнозирование на основе регрессионных моделей

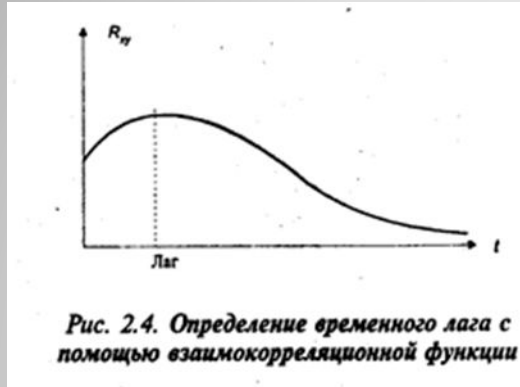
- Как и в случае парной модели, различают точечный и интервальный прогнозы. В первом случае прогнозируемая оценка – некоторое число, во втором – интервал, в котором находится истинное значение зависимой переменной с заданным уровнем значимости.

№	Удой молока на среднегодовую корову, кг	Расход кормов на 1 корову, корм. ед.	Удельный вес чистопородных коров в стаде, %	Себестоимость молока за 1 кг, руб.
1	3280	48,20	61	0,313
2	2920	43,10	54	0,413
3	5140	60,70	70	0,268
4	4630	60,10	67	0,310
5	4950	59,40	71	0,309
6	5000	52,50	74	0,288
7	2790	44,00	45	0,357
8	4340	54,20	68	0,247
9	4160	53,20	65	0,305
10	2660	46,40	51	0,376
11	2960	47,10	52	0,351
12	3230	46,10	57	0,356
13	3480	53,90	58	0,312
14	3230	53,40	52	0,415
15	2370	39,40	44	0,411
16	2610	40,20	50	0,380
17	3000	45,50	52	0,326
18	2960	41,40	49	0,341
19	3100	47,80	53	0,398
20	2720	46,30	57	0,405

Расчетные таблицы для определения коэффициента корреляции рангов Спирмэна

Удой молока на среднегодовую корову, кг	Себестоимость молока за 1 кг, руб.	Ранги		Разность рангов $d = N_x - N_y$	d^2
$\bar{x}$	$\bar{y}$	N_x	N_y		
3280	0,313	8	13	-5	25
2920	0,413	15	2	13	169
5140	0,268	1	19	-18	324
4630	0,31	4	15	-11	121
4950	0,309	3	16	-13	169
5000	0,288	2	18	-16	256
2790	0,357	16	8	8	64
4340	0,247	5	20	-15	225
4160	0,305	6	17	-11	121
2660	0,376	18	7	11	121
2960	0,351	13,5	10	3,5	12,25
3230	0,356	9,5	9	0,5	0,25
3480	0,312	7	14	-7	49
3230	0,415	9,5	1	8,5	72,25
2370	0,411	20	3	17	289
2610	0,38	19	6	13	169
3000	0,326	12	12	0	0
2960	0,341	13,5	11	2,5	6,25
3100	0,398	11	5	6	36
2720	0,405	17	4	13	169
$\bar{n} = 20$				$\sum d^2 =$	2398
				$\rho =$	-0,803

- Определение временного сдвига (временного лага) двух переменных



Учет временного сдвига при определении взаимозависимости результирующего параметра и независимого фактора позволяет получать более корректные выводы.

Модель авторегрессии в корреляционной теории

$$x[t] = \sum_{j=1}^p \Phi[j]x[t-j] + a[t]$$

где $\Phi[j]$ - коэффициенты АР, $a[t]$ - некоррелированные случайные отсчеты, p - порядок модели АР.
Величина

$$\tilde{x}[t] = \sum_{j=1}^p \Phi[j]x[t-j]$$

называется предсказанием случайной величины $x[t]$. Разность между текущим значением отсчета и его предсказанием называется ошибкой предсказания

$$R[i] - \sum_{j=1}^p \Phi[j]R[j-i] = 0 \quad i = 1 \div p$$

$$R[0] - \sum_{j=1}^p \Phi[j]R[j] = D_x$$

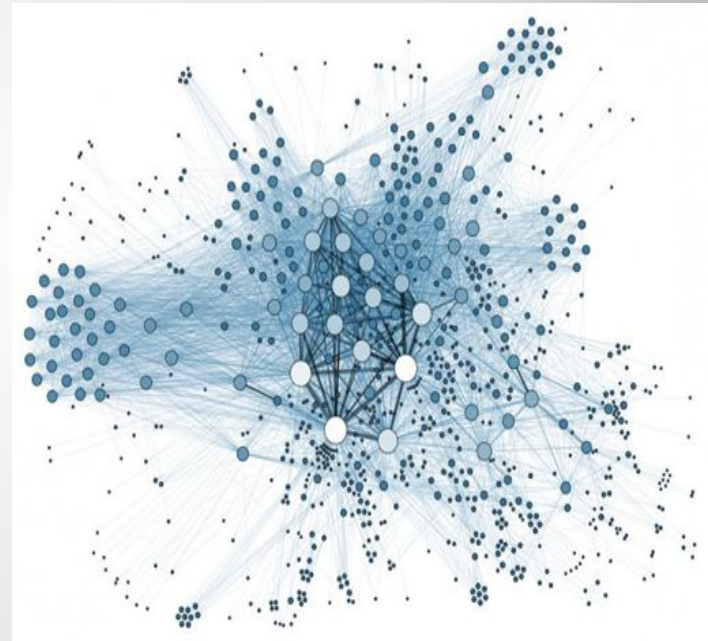
где $R[i] = E\{x[t]x[t-i]\}$ - значения функции корреляции случайного процесса

Корреляционно-регрессионный анализ: пример

- В статистике и экономике применяются самые разнообразные виды и объекты анализа. Статистические методы анализа направлены на изучение повторяющихся процессов, для того чтобы составить длительные прогнозы поведения экономических явлений. Например, для того чтобы проанализировать социально-экономическое развитие территории, необходимо изучить показатели уровня жизни населения. Корреляционно-регрессионный анализ в статистике позволяет создать уравнение регрессии и определить коэффициенты корреляции, демонстрирующие взаимосвязь между уровнем жизни и развитием территории. Уровень жизни определяется доходами, а основной источник доходов – зарплата. В таком случае фактором выступает уровень зарплаты, а результатом – численность населения с невысокими доходами.

Программное обеспечение анализа

- В качестве столбцов и строк используются корреляционные коэффициенты. На основе полученных данных таблицы необходимо будет провести корреляционный анализ. Пример последовательности проведения анализа:
 - В команде «Сервис» выбрать пункт «Анализ данных».
 - В качестве инструмента анализа выбрать пункт «Корреляция».
 - В появившемся окне в строке «Входной интервал» указать диапазон анализируемых данных, выбрать пункт «Группировка» в строке «Параметры вывода», ввести диапазон вывода результатов и нажать «ОК».



Спасибо за внимание!