

Использование корреляционно-регрессионного анализа в управлении предприятием



- **Корреляционно-регрессионный анализ** как общее понятие включает в себя измерение тесноты, направления связи и установление аналитического выражения (формы) связи (регрессионный анализ). Этот метод содержит две составляющие части - корреляционный анализ и регрессионный анализ.

Корреляционный анализ - это количественный метод определения тесноты и направления взаимосвязи между выборочными переменными величинами.

- В статистике принято различать следующие варианты зависимостей:
- **Парная корреляция** - связь между двумя признаками (результативным и факторным или двумя факторными).
- **Частная корреляция** - зависимость между результативным и одним факторным признаками при фиксированном значении других факторных признаков.
- **Множественная корреляция** - зависимость результативного и двух или более факторных признаков, включенных в исследование.

- Относительно формы связи различают:
- А) линейную корреляцию - характеризует тесноту и направление связи между двумя коррелируемыми признаками, в случае наличия между ними линейной зависимости.
- Б) нелинейную - корреляция, при которой отношение степени изменения одной переменной к степени изменения другой переменной является изменяющейся величиной.

- **Регрессионный анализ** - заключается в определении аналитического выражения связи, в котором изменение одной величины обусловлено влиянием одной или нескольких величин, а множество всех прочих факторов, также оказывающих влияние на зависимую величину, принимается за постоянные и средние значения. **Регрессия может быть однофакторной(парной) и многофакторной(множественной).**

- Относительно формы зависимости различают:
- А) **линейную регрессию**, выражаемую линейной функцией. При этой форме зависимости между исследуемыми переменными объективно существуют линейные соотношения. Выражается уравнением прямой вида:

$$\bar{Y}_x = a_0 + a_1x \quad \bar{Y}_x = a_0 + a_1x$$

- Б) **нелинейную регрессию**, выражаемую нелинейной функцией. В этом случае между исследуемыми экономическими явлениями объективно существуют нелинейные соотношения. Выражается уравнением вида:

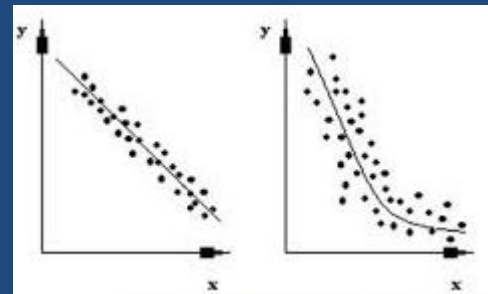
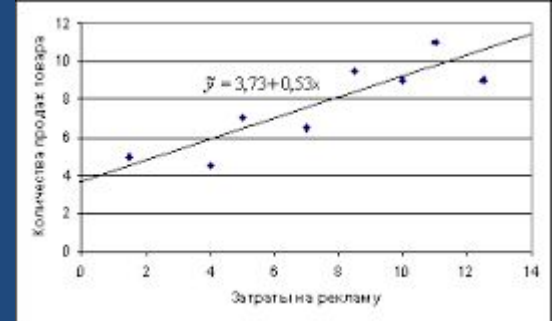
- Парабола -

$$\bar{Y}_x = a_0 + a_1x + a_2x^2 \quad \bar{Y}_x = a_0 + a_1x + a_2x^2$$

$$\bar{Y}_x = a_0 + \frac{a_1}{x} \text{ и т. д.} \quad \bar{Y}_x = a_0 + \frac{a_1}{x} \text{ и т. д.}$$

- Гипербола -

- По направлению связи различают:
- **прямую регрессию (положительную)**, возникающую при условии, если с увеличением или уменьшением независимой величины значения зависимой также соответственно увеличиваются или уменьшаются;
- **обратную (отрицательную) регрессию**, появляющуюся при условии, что с увеличением или уменьшением независимой величины зависимая соответственно уменьшается или увеличивается.



- **Требования**, при которых соблюдается адекватность уравнения регрессии
- **Совокупность** исследуемых исходных данных должна быть однородной.
- **Возможность** описания моделируемого явления одним или несколькими уравнениями причинно-следственных связей.
- **Все факторные признаки** должны иметь количественное (цифровое) выражение.
- **Наличие** достаточно большого объёма исследуемой выборочной совокупности.
- **Причинно-следственные связи** между явлениями и процессами следует описывать линейной или приводимой к линейной формами зависимости.
- **Отсутствие** количественных ограничений на параметры моделей связи.
- **Постоянство** территориальной и временной структуры изучаемой продукции.

- **Основной предпосылкой корреляционного анализа** является необходимость подчинения совокупности значений всех факторных (x_1, x_2, \dots, x_k) и результативного (Y) признаков k -мерному нормальному закону распределения или близость к нему.
- **Целью регрессионного анализа** является оценка функциональной зависимости условного среднего значения результативного признака (Y) от факторных (x_1, x_2, \dots, x_k).
- **Основной предпосылкой регрессионного анализа** является то, что только результативный признак подчиняется нормальному закону распределения, а факторные признаки могут иметь произвольный закон распределения.

- Рассмотрим метод линейного коэффициента корреляции более обширней. Линейный коэффициент корреляции разработали Карл Пирсон, Фрэнсис Эджуорт и Рафаэль Уэлдон в 90-х годах XIX века и рассчитывается по формуле:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] * [n \sum y^2 - (\sum y)^2]}} r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] * [n \sum y^2 - (\sum y)^2]}}$$

- Где X - факторный признак Y - результативный
- Коэффициент корреляции изменяется по модулю от -1 до 1.
- **1** - идеальная положительная связь Все точки данных располагаются строго на прямой линии, направленной вверх и в право.
- **Близко к 1** - сильная положительная взаимосвязь. Точки данных плотно сгруппированы вокруг прямой линии, направленной вверх и вправо.
- **Близко к 0 (положительно)** - отсутствие взаимосвязи. Случайное облако точек данных. Не имеет чёткой направленности ни вверх, ни вниз при движении вправо.
- **Близко к 0 (отрицательно)** - незначительная отрицательная взаимосвязь. Точки данных образуют случайное облако с незначительной ориентацией вниз и вправо.
- **Близко к -1** - сильная отрицательная взаимосвязь. Точки данных плотно сгруппированы вокруг прямой линии, направленной вниз и вправо.
- **-1** - Идеальная взаимосвязь, все точки располагаются строго на прямой.
- **Не определено** - точки данных располагаются строго на горизонтали или на вертикальной линии.

- **Коэффициент ранговой корреляции Кендалла**

- Применяется для выявления взаимосвязи между количественными или качественными показателями, если их можно ранжировать. Значения показателя X выставляют в порядке возрастания и присваивают им ранги. Ранжируют значения показателя Y и рассчитывают коэффициент Кендалла:

$$\tau = \frac{2S}{n(n-1)},$$

- где $S = P - Q$
- P - суммарное число наблюдений, следующих за текущими наблюдениями с **БОЛЬШИМ** значением рангов Y.
- Q- суммарное число наблюдений, следующих за текущими наблюдениями с **МЕНЬШИМ** значением рангов Y (равные ранги не

- Коэффициент ранговой корреляции Спирмена
- Каждому показателю X и Y присваивается ранг. На основе полученных рангов рассчитываются их разности и вычисляется коэффициент корреляции Спирмена:

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

- **Коэффициент корреляции знаков Фехнера**

- Подсчитывается количество совпадений и несовпадений знаков отклонений значений показателей от их среднего значения.

$$i = \frac{C - H}{C + H}$$

- С - число пар, у которых знаки отклонений значений от их средних совпадают.
- Н - число пар, у которых знаки отклонений значений от их средних не совпадают.

- Построение моделей и использование их на практике
- Результаты деятельности промышленных предприятий

Номер предприятия	Среднесписочная численность рабочих, чел.	Среднесписочная численность служащих, чел.	Затраты на сырьё, материалы, тыс. руб.	Фонд заработной платы (без учета выплат в различные фонды), тыс. р.	Объем товарной продукции, тыс. р.
1	130	22	3500	5973	13246
2	146	30	3800	3737	14871
3	156	26	3900	1489	16278
4	178	31	4400	3222	14322
5	184	36	4500	5450	11129
6	119	30	3100	2838	12780
7	127	22	3200	3682	13400
8	203	45	4600	3214	10431
9	170	38	4300	2864	11399
10	146	30	3600	1870	7641
11	161	38	4000	2565	18036
12	174	41	4200	2940	9819
13	150	27	3700	1189	7463
14	142	28	3400	2021	6560
15	153	31	3300	1524	7841
16	215	37	4700	5379	16696
Всего	2554	512	62200	49957	191912

- Расчет относительных показателей

Номер предприятия	Среднесписочная численность рабочих, чел.	Среднемесячная заработная плата рабочего	Материалоёмкость
1	108	4,61	0,26
2	116	2,68	0,26
3	130	0,95	0,24
4	147	1,83	0,31
5	148	3,07	0,40
6	89	2,66	0,24
7	105	2,92	0,24
8	158	1,70	0,44
9	132	1,81	0,38
10	116	1,34	0,47
11	123	1,74	0,22
12	133	1,84	0,43
13	123	0,81	0,50
14	114	1,48	0,52
15	122	1,04	0,42
16	178	2,52	0,28
Всего	2042	32,99	5,61

- **Определение тесноты взаимосвязи между показателями с помощью коэффициента ранговой корреляции.**
- Определим тесноту связи между показателями: фонд заработной платы, среднесписочная численность рабочих, используя коэффициент ранговой корреляции. Этот коэффициент представляет собой показатель, характеризующий статистическую связь двух признаков, измеряемых в порядковой шкале. Для признаков, измеренных в порядковых шкалах, наиболее известным является коэффициент ранговой корреляции Спирмена:

$$\rho_{\text{Сп}} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

- d_i^2 - квадрат разности рангов

Номер предприятия	Среднесписочная численность работающих, чел. (X)	Фонд заработной платы (без учета выплат в различные фонды), тыс. Р(Y)	Ранг X	Ранг Y	di^2	Rx	Ry	Q	P
1	108	3222	3	11	64	1	3	2	13
2	116	3737	5,5	13	56,25	2	16	14	0
3	130	1870	10	4	36	3	11	9	5
4	147	5379	13	14	1	4	2	1	11
5	148	2864	14	8	36	5,5	13	9	2
6	89	1524	1	3	4	5,5	9	6	4
7	105	5973	2	16	196	7	12	7	2
8	158	2565	15	6	81	8,5	15	8	0
9	132	2838	11	7	16	8,5	5	2	5
10	116	2940	5,5	9	12,25	10	4	1	5
11	123	5450	8,5	15	42,25	11	7	2	3
12	133	3214	12	10	4	12	10	3	1
13	123	2021	8,5	5	12,25	13	14	3	0
14	114	1489	4	2	4	14	8	2	0
15	122	3682	7	12	25	15	6	1	0
16	178	1189	16	1	225	16	1	0	0
Всего	2042	49957			815			70	51

$$P_{xy} = 1 - \frac{6 * 815}{16(16^2 - 1)} = -0,21$$

- Из полученного значения можно сделать вывод, что взаимосвязи практически нет. Точки данных образуют случайное облако с незначительной ориентацией вниз и вправо. Иначе говоря среднесписочная численность рабочих не влияет на увеличение фонда заработной платы из-за внешних или внутренних различных явлений и факторов.

- **Определение тесноты парной связи и формы для всей статистической совокупности**

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] * [n \sum y^2 - (\sum y)^2]}}$$

Номер предприятия	Количество изделий (x)	затраты, тыс. руб. (y)	X*Y	X^2	Y^2
1	22	3,5	77	484	12,25
2	30	3,8	114	900	14,44
3	26	3,9	101,4	676	15,21
4	31	4,4	136,4	961	19,36
5	36	4,5	162	1296	20,25
6	30	3,1	93	900	9,61
7	22	3,2	70,4	484	10,24
8	45	4,6	207	2025	21,16
9	38	4,3	163,4	1444	18,49
10	30	3,6	108	900	12,96
11	38	4	152	1444	16
12	41	4,2	172,2	1681	17,64
13	27	3,7	99,9	729	13,69
14	28	3,4	95,2	784	11,56
15	31	3,3	102,3	961	10,89
16	37	4,7	173,9	1369	22,09
Всего	512	62,2	2028,1	17038	245,84

- Рассчитаем линейный коэффициент корреляции:

$$r_{xy} = \frac{16 * 2028,1 - 512 * 62,2}{\sqrt{[(5 * 17038) - (512)^2] * [(16 * 245,84) - (62,2)^2]}} = 0,001$$

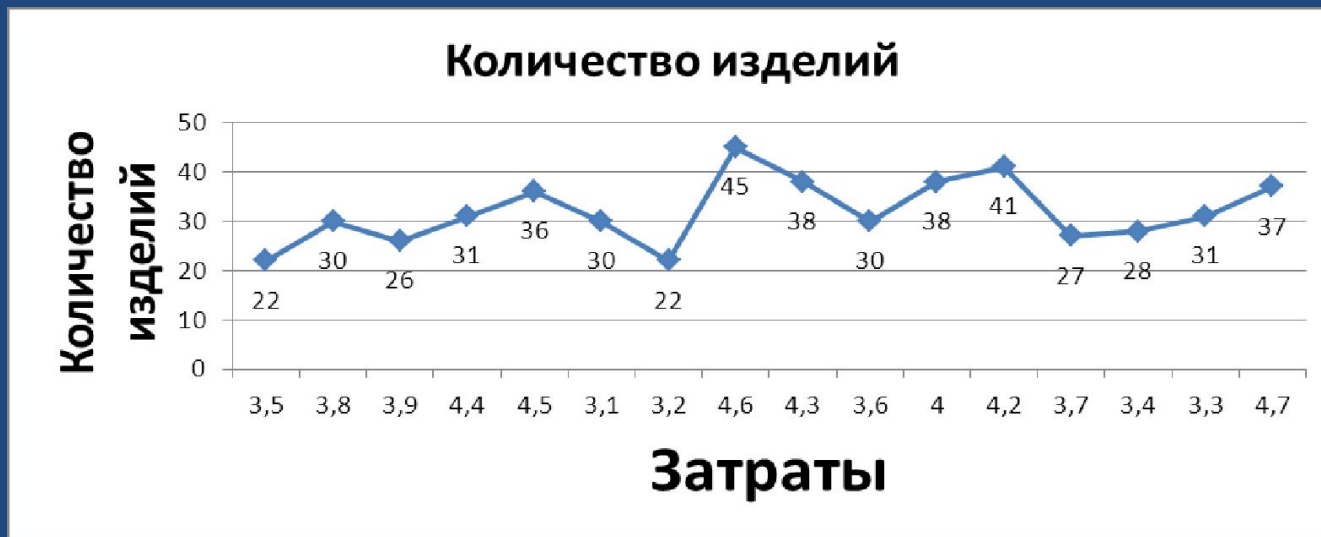
- Так как коэффициент положительный, следовательно, связь прямая. Можно сделать вывод, что с увеличением выпуска продукции, затраты увеличиваются, а с уменьшением - уменьшаются.

- График уравнений линейной регрессии для данных статистической совокупности
- Найдём параметры уравнения линейной регрессии:

$$\overline{Y_x} = a_0 + a_1 x$$

$$S = \sum_{i=1}^n (y_i - \overline{Y_x})^2 \rightarrow \min \sum_{i=1}^n (y_i - \overline{Y_x})^2 \rightarrow \min$$

$$\begin{cases} n a_0 + a_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i y_i \end{cases}$$

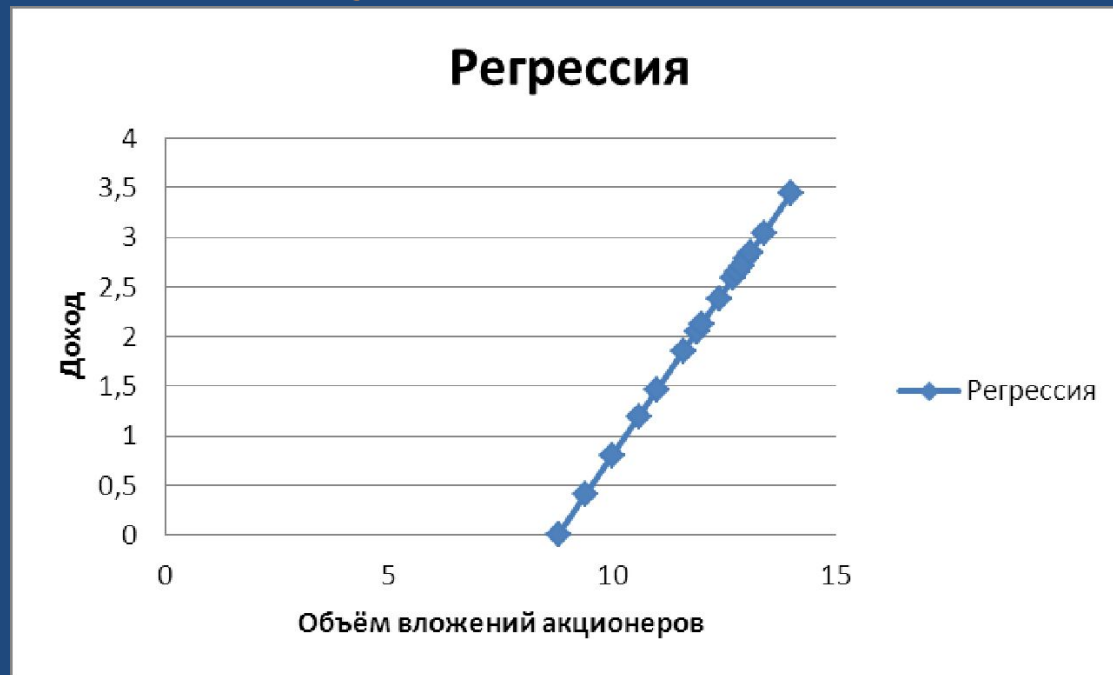


- a_0 - параметр отражающий количественную характеристику факторов, не включённых в данную модель.
- a_1 - коэффициент регрессии. Показывает как изменяется результирующий признак, при изменении факторного признака на единицу измерения.

Таблица промежуточных расчётов

Номер предприятия	Чистый доход, млрд.	Объём вложений акционеров, млрд.	x^2	xy	Y_x
1	0,1	8,8	77,44	0,88	0,008
2	1,3	9,4	88,36	12,22	0,404
3	0,1	10	100	1	0,8
4	2,6	10,6	112,36	27,56	1,196
5	0,1	11	121	1,1	1,46
6	0,3	11,9	141,61	3,57	2,054
7	4,6	12,7	161,29	58,42	2,582
8	2,5	12,8	163,84	32	2,648
9	1,7	13	169	22,1	2,78
10	3,2	12,4	153,76	39,68	2,384
11	0,9	12	144	10,8	2,12
12	1,7	11,6	134,56	19,72	1,856
13	3,3	12,9	166,41	42,57	2,714
14	2,8	13,1	171,61	36,68	2,846
15	2,3	13,4	179,56	30,82	3,044
16	4,4	14	196	61,6	3,44
Итого	31,9	189,6	2280,8	400,72	

- График уравнения регрессии.



- Из полученных значений можно сделать вывод, что при увеличении акционерных доходов на 1 млрд., чистый доход увеличится на 66 млрд.

- **Диаграмма рассеяния**
- позволяет увидеть структуру данных, наглядно демонстрирует взаимосвязь явлений, представляет каждое наблюдение в пространстве двух измерений, соответствующих двум факторам. По оси X располагается переменная, являющаяся «причиной» т.е. фактор, по оси Y - следствие (результат)



СПАСИБО ЗА
ВНИМАНИЕ!