Анализ и диагностика финансово-хозяйственной деятельности предприятия

Тема № 3

ФАКТОРНЫЙ АНАЛИЗ

Факторный анализ

1. Методика факторного анализа
2. Типы детерминированных факторных моделей
3. Способы измерения влияния факторов в детерминированных факторных моделях
4. Стохастические факторные модели
5. Способы измерения влияния факторов в стохастическом факторном анализе

Стохастический факторный анализ

Корреляционный метод

Регрессионный метод

Дисперсионный метод

Метод кластерного анализа

Другие методы

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

Целью *корреляционного анализа* является *оценка тесноты связи* между признаками.

Теснота связи *количественно* выражается величиной коэффициентов корреляции.

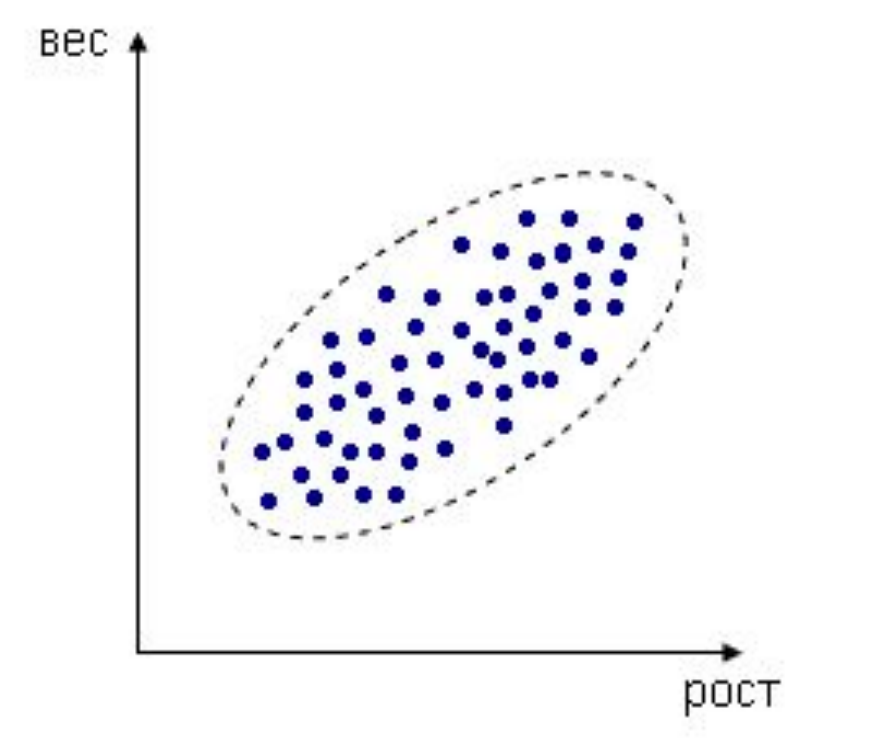
Корреляционный анализ:

1. **Парная корреляция** – связь между двумя признаками (результативным и факторным или двумя факторными).
2. **Частная корреляция** – зависимость между результативным и одним факторным признаками при фиксированном значении других факторных признаков.
3. **Множественная корреляция** – зависимость результативного и двух или более факторных признаков, включенных в исследование.

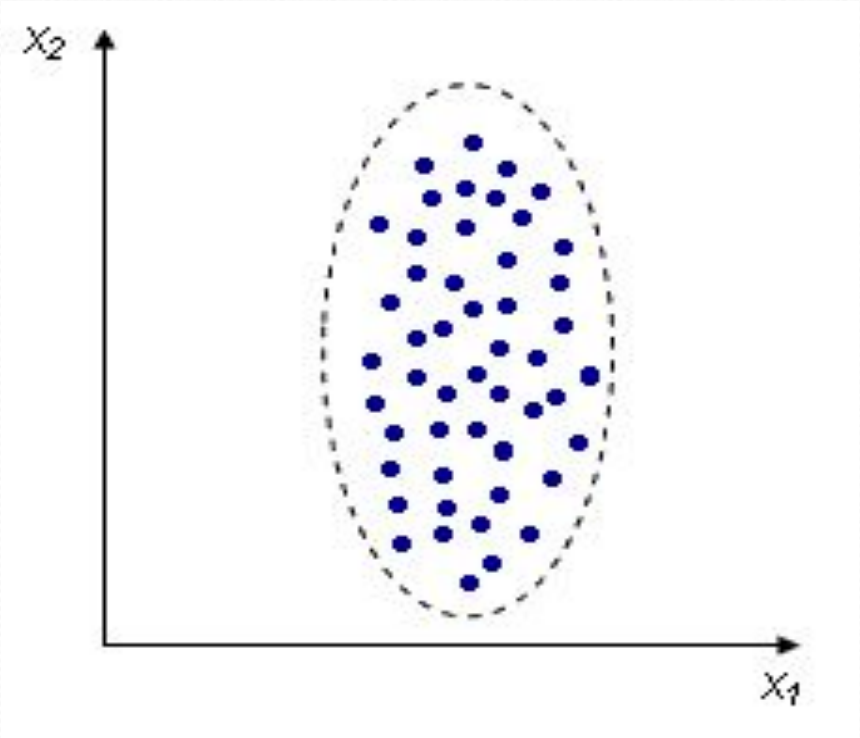
Виды корреляционных связей:

- **По форме** корреляционная связь может быть прямолинейной или криволинейной.
- **По направлению** корреляционная связь может быть положительной ("прямой") и отрицательной ("обратной").
- **По силе** корреляционная связь определяется шкалой Чеддока

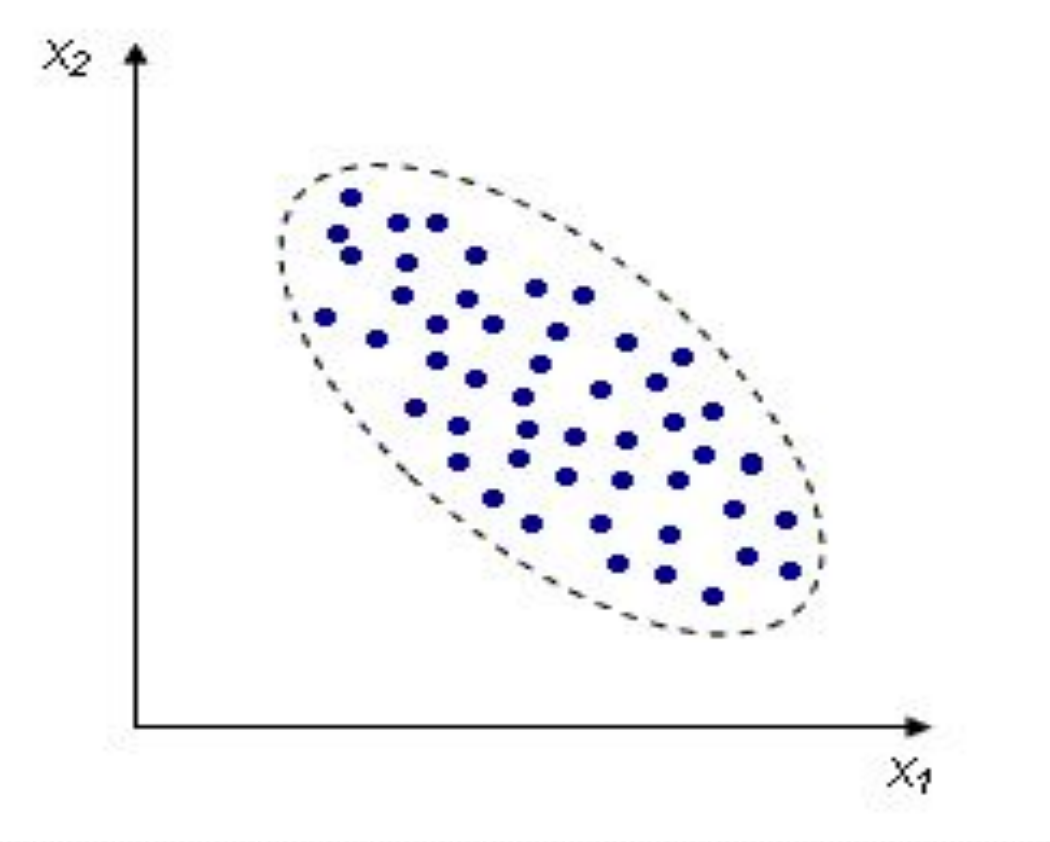
Положительная корреляция



Отсутствие корреляции



Отрицательная корреляция



Шкала Чеддока

Теснота связи	Значение коэффициента корреляции при наличии:	
	Прямой связи	Обратной связи
Слабая	0,1–0,3	(–0,3)–(–0,1)
Умеренная	0,3–0,5	(–0,5)–(–0,3)
Заметная	0,5–0,7	(–0,7)–(–0,5)
Высокая	0,7–0,9	(–0,9)–(–0,7)
Весьма высокая	0,9–1	(–1)–(–0,9)

Термин «корреляция» был введен в науку английским естествоиспытателем **Френсисом Гальтоном** в 1886 г.

Однако точную формулу для подсчета коэффициента корреляции разработал его ученик **Карл Пирсон**.

Формула коэффициента корреляции при линейной зависимости

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \cdot \sum (y_i - \bar{y})^2}}$$

Величина коэффициента линейной корреляции Пирсона не может превышать +1 и быть меньше чем -1. Эти два числа +1 и -1 — являются границами для коэффициента корреляции.

Когда при расчете получается величина большая +1 или меньшая -1 — следовательно произошла ошибка в вычислениях.

Регрессионный анализ — это количественный метод определения вида математической функции в причинно-следственной зависимости между переменными величинами.

Целью регрессионного анализа является установление формы зависимости.

Виды регрессий

Линейная* регрессия**: ***y=a+bx	***Регрессии, нелинейные по объясняющим переменным*:** **1) полиномы разных степеней *y=a+b1x+b2x2+…*;** **2) равносторонняя гипербола *y=a+b/x*.** ***Регрессии, нелинейные по оцениваемым параметрам*:** **1) степенная *y=axb*;** **2) показательная *y=abx*;** **3) экспоненциальная *y=ea+bx*.**

Оценки параметров *a и b находятся по формулам:*

$$a = \bar{y} - b \cdot \bar{x}, \qquad b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)},$$

$$\sigma_x^2 = \overline{x^2} - \overline{x}^2$$ – дисперсия признака x

$$\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \overline{xy} - \overline{x} \cdot \overline{y}$$

a и b

Формально *a* – *значение y при x =0.*

Если признак-фактор x не может иметь нулевого значения, то вышеуказанная трактовка свободного члена *a не имеет смысла, т.е. параметр a может не иметь экономического* содержания.

Параметр *b* *называется коэффициентом регрессии.*

Его величина показывает среднее изменение результата с изменением фактора на одну единицу.

Коэффициент детерминации

$$r_{xy}^2 = 1 - \frac{\sigma_{ост}^2}{\sigma_y^2},$$

$$\sigma_{ост}^2 = \frac{1}{n}\sum\left(y - \hat{y}_x\right)^2, \quad \sigma_y^2 = \frac{1}{n}\sum\left(y - \bar{y}\right)^2 = \overline{y^2} - \bar{y}^2$$

Пример:

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

Решение:

	x	y	xy	x^2	y^2
1	78	133	10374	6084	17689
2	82	148	12136	6724	21904
3	87	134	11658	7569	17956
4	79	154	12166	6241	23716
5	89	162	14418	7921	26244
6	106	195	20670	11236	38025
7	67	139	9313	4489	19321
8	88	158	13904	7744	24964
9	73	152	11096	5329	23104
10	87	162	14094	7569	26244
11	76	159	12084	5776	25281
12	115	173	19895	13225	29929
ИТОГО	**1027**	**1869**	**161808**	**89907**	**294377**
среднеее значение	*85,6*	*155,8*	*13484,0*	*7492,3*	*24531,4*

Решение:

$$r = \frac{\overline{yx} - \overline{y}\,\overline{x}}{\overline{x^2} - \overline{x}^2} = \frac{13484 - 155{,}8 \cdot 85{,}6}{7492{,}3 - 85{,}6^2} \approx 0{,}92$$

$$a = \overline{y} - b\overline{x} = 155{,}8 - 0{,}92 \cdot 85{,}6 \approx 77$$

Кластерный анализ

Кластерный анализ представляет собой класс методов, используемых для классификации объектов или событий в относительно однородные группы, которые называют **кластерами (clusters).**

Кластерный анализ

Объекты в каждом кластере должны быть похожи между собой и отличаться от объектов в других кластерах.

Кластерный анализ также называют классификационным анализом (classification analysis) или численной таксономией (систематикой) (numerical taxonomy).

Идеальная ситуация кластеризации

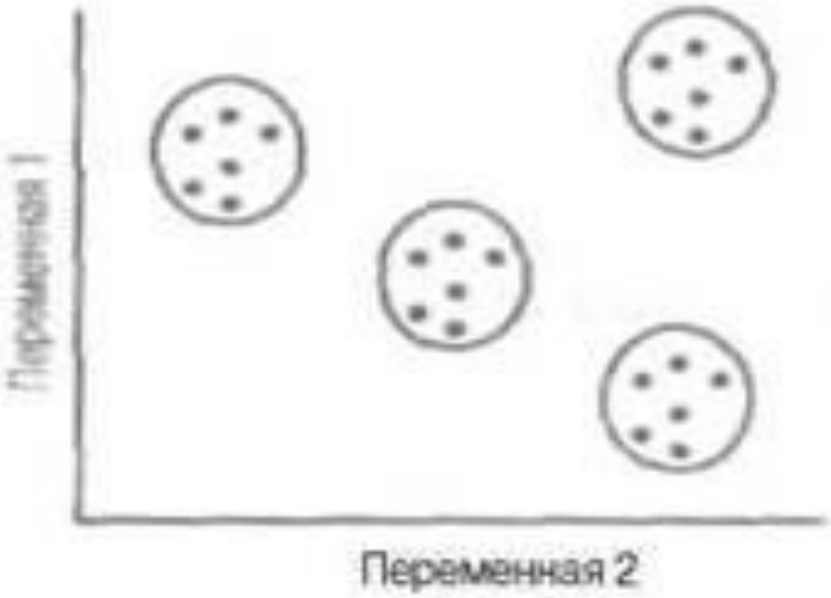


Рис. 20.1. Идеальная ситуация кластеризации

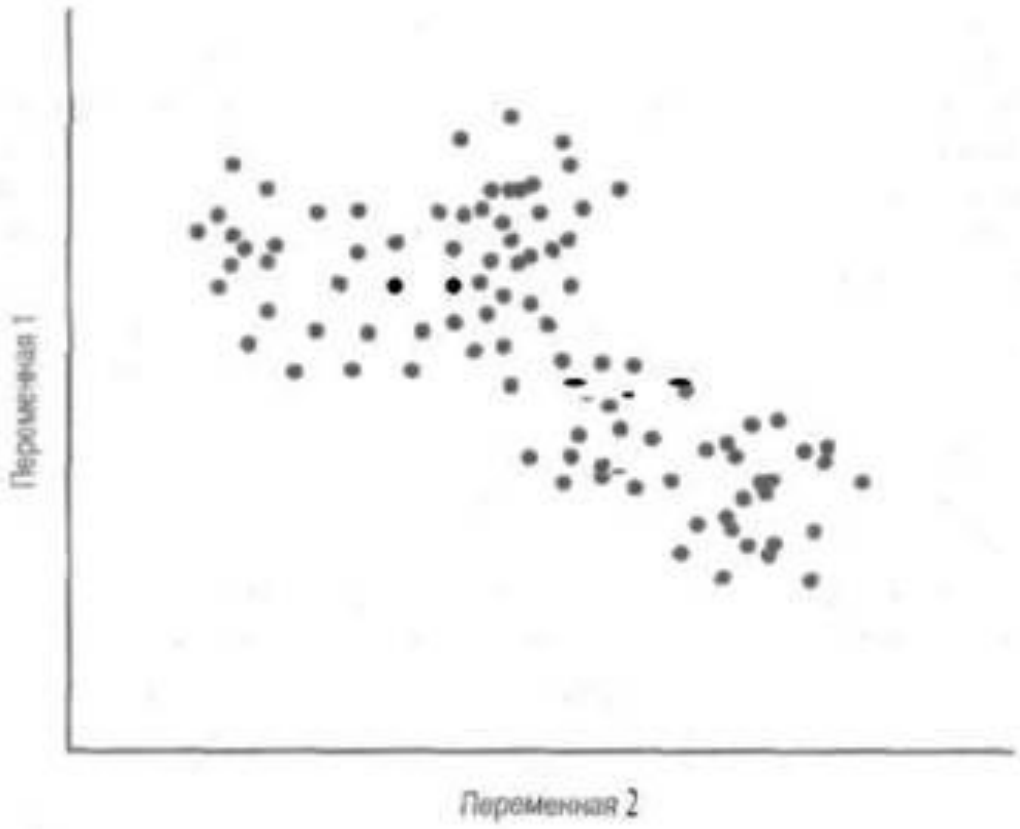


Рис. 20.2. Реальная ситуация кластеризации

Самостоятельная работа

1

- А - О

2

- П - Я

Понимание причинно-следственных связей между различными явлениями очень важно для эффективного принятия решений – как в области экономической политики, так и в быту. Однако сделать правильный вывод о причинно-следственной связи, наблюдая статистику, не так просто. Следующие примеры иллюстрируют ошибки, которые легко допустить при поверхностном анализе. Примеры, иллюстрирующие похожие ошибки, объединены в группы. Для каждой группы:

а) объясните для каждого из примеров в отдельности, почему сделанный вывод неверен;

б) сформулируйте, какой общий тип ошибки иллюстрируют примеры данной группы;

в) приведите другой пример, иллюстрирующий тот же тип ошибки.

1) Дворник Василий Иваныч несколько лет подряд наблюдал такую картину: вскоре после того как десяток-другой жителей дома № 48 выходили на крышу, начинался фейерверк. Так было и на Новый год, и на День Победы, и на День города; закономерность ни разу не была нарушена. Василий Иваныч ложится спать рано и не любит шум. Наконец он понял, как решить эту проблему: в преддверии очередного праздника он повесил замок на дверь, ведущую на крышу. Теперь он может быть уверен, что фейерверк не потревожит его сон.

2) *«Как только я нанял телохранителя, страховая компания стала требовать более высокую плату за страхование моей жизни, объясняя это тем, что теперь они исходят из более высокой вероятности моей смерти в ближайшем будущем. Это произвело на меня глубокое впечатление. У них большой опыт и накопленная статистика, им виднее; я хочу пожить подольше, поэтому я решил поскорее уволить телохранителя».*

1) Одна кинокомпания заметила, что снятые ею сиквелы (продолжения) её же фильмов почти всегда нравятся зрителям меньше, чем первый фильм. Поэтому, если её цель – снять как можно более интересный для зрителей фильм, ей следует снять фильм на новый сценарий, не связанный с предыдущими фильмами, так как при прочих равных условиях ожидаемый кинокомпанией зрительский успех у такого фильма выше, чем у сиквела.

2) Замечено, что, в большинстве случаев, после того как родители ругают ребёнка за полученную оценку, его средний балл растёт, а после того как родители хвалят его за полученную оценку, его средний балл падает. Значит, если родители хотят, чтобы ребёнок имел как можно более высокий средний балл, за полученные оценки его нужно ругать, но нельзя хвалить.