

КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Кафедра математической статистики

Методические разработки
по специальному курсу

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

Часть I

ТЕОРЕТИКО-ВЕРОЯТНОСТНЫЕ АСПЕКТЫ.

РЕГРЕССИЯ. КОРРЕЛЯЦИЯ.

МЕТОД ГЛАВНЫХ КОМПОНЕНТ. ФАКТОРНЫЙ АНАЛИЗ.

МНОГОМЕРНАЯ НОРМАЛЬНАЯ МОДЕЛЬ.

Казань – 2006

П Е Ч А Т А Е Т С Я
ПО РЕШЕНИЮ СЕКЦИИ НАУЧНО-МЕТОДИЧЕСКОГО СОВЕТА
КАЗАНСКОГО УНИВЕРСИТЕТА

С о с т а в и т е л ь: доцент С.В. Симушкин

Никто не обнимет необъятного.¹

В практике статистических обследований обычно проводятся измерения большого числа показателей, по которым планируется построение тех или иных выводов об объектах исследования. В курсе многомерного статистического анализа изучаются основные методы обработки таких данных: регрессионный анализ, корреляционный анализ, метод выделения главных компонент, факторный анализ, дискриминантный анализ. В основе построения большинства из этих методов лежит идея представления вектора наблюдений посредством наилучшей в некотором смысле линейной комбинации его компонент – линейная регрессионная модель.

В первой части данного пособия мы рассмотрим теоретические характеристики вероятностных распределений, позволяющие судить о степени взаимосвязи случайных величин. Собственно методы многомерного статистического анализа представлены во второй части.

Вспомогательный материал, включающий в себя необходимые сведения из векторной алгебры, теории распределения случайных векторов и ряд полезных интегральных соотношений, собран в двух приложениях.

¹Все цитаты принадлежат славному перу незабвенного поэта-мыслителя Козьмы Пруткова.

Оглавление

Где начало того конца, которым
оканчивается начало?

I.	Характеристики взаимосвязи	
§ 1.	Линейная регрессия	5
	Остаточная дисперсия	8
§ 2.	Меры зависимости	10
	Множественный коэффициент корреляции. . .	10
	Частный коэффициент корреляции	12
§ 3.	Каноническая корреляция	16
	Пример интерпретации канонических величин.	20
§ 4.	Эллипсоид рассеяния	21
§ 5.	Нелинейная регрессия	25
	Доказательства	28
II.	Метод главных компонент	
§ 1.	Определение и способ построения главных компонент	44
§ 2.	Свойства главных компонент	46
§ 3.	Понижение размерности наблюдаемого вектора	47
§ 4.	Интерпретация главных компонент	48
	Доказательства	50

III.	Факторный анализ	
§ 1.	Модель факторного анализа	56
§ 2.	Методы построения факторной модели	61
	Оценка общностей.	62
	Построение матрицы нагрузок.	65
§ 3.	Проблема вращения	68
IV.	Нормальное распределение	
§ 1.	Плотность, характеристическая функция, моменты	71
§ 2.	Линейные преобразования нормального вектора	74
§ 3.	Квадратичные формы	77
§ 4.	Условные распределения	77
	Доказательства	80
A.	Случайные векторы	
	Элементы векторной алгебры	85
	Случайные векторы	89
B.	Интегральные соотношения	95
	Предметный указатель	98

Глава I. ХАРАКТЕРИСТИКИ ВЗАИМОСВЯЗИ
И ИЗМЕНЧИВОСТИ
СЛУЧАЙНЫХ ВЕЛИЧИН

§ 1. Линейная регрессия

Коэффициент счастья в обратном
содержании к достоинству.

Большинство методов многомерного анализа эксплуатируют в своей основе идею линейной связи компонент случайного вектора. Поэтому первая задача, которую мы рассмотрим, – задача линейной аппроксимации значений одной случайной величины (с.в.) по реализациям остальных составляющих p -мерного случайного вектора $\vec{X} = (X_1, \dots, X_p)'$. Вспомогательный материал, содержащий сведения об основных характеристиках распределений с.в., приведен в Приложении А.

О п р е д е л е н и е 1. Линейная функция

$$x_{1*} = x_{1*(2\dots p)} = \beta_0 + \beta_2 x_2 + \dots + \beta_p x_p$$

называется линейной среднеквадратической регрессией (сокращенно, ЛСКР) с.в. X_1 на вектор с.в. (X_2, \dots, X_p) , если

$$\mathbf{E}(X_1 - X_{1*})^2 = \min_{b_0, \dots, b_p} \mathbf{E}(X_1 - (b_0 + b_2 X_2 + \dots + b_p X_p))^2.$$

З а м е ч а н и е 1. Существенным в этом определении является линейная зависимость от коэффициентов b_0, b_2, \dots, b_p , а не от компонентов вектора \vec{X} . Если требуется построить регрессию с функциональной зависимостью $f(x)$ от переменной X_i , то простая замена $f(X_i) \rightarrow X_i$ снова приводит к задаче построения ЛСКР. Единственным препятствием на этом пути является то обстоятельство, что при такой замене распределение вектора \vec{X} перестает быть нормальным, конечно, если оно таковым было.

Пояснительные выражения
объясняют темные мысли.

Т е о р е м а I.1. Если вектор \vec{X} имеет средние $\vec{\mu} = (\mu_1, \dots, \mu_p)'$ и невырожденную матрицу ковариаций $\Lambda > 0$, то ЛСКР случайной величины X_1 на (X_2, \dots, X_p) равна

$$x_{1*} = \mu_1 + \sum_{i=2}^p (-1)^i \frac{|\Lambda_{1i}|}{|\Lambda_{11}|} (x_i - \mu_i). \quad (I.1)$$

З а м е ч а н и е 2. Если матрица Λ_{11} вырождена, то вектор $(X_2, \dots, X_p)'$ состоит из линейно зависимых случайных величин (см. теорему А.1, стр. 91), то есть в этом случае часть случайных величин представляется посредством линейных комбинаций остальных составляющих вектора $(X_2, \dots, X_p)'$. Поэтому достаточно ограничиться рассмотрением только линейно независимой части этого вектора.

Уравнение регрессии j -й компоненты вектора \vec{X} получается аналогичным образом. Приведем здесь одну интерес-

ную форму записи всех уравнений регрессии, удобную для запоминания.

Правило. Для того, чтобы записать уравнение ЛСКР с.в. X_j относительно компонент вектора \vec{X} , необходимо j -ую строку соотношения

$$\Lambda^{-1}(\vec{x} - \vec{\mu}) = \vec{0}$$

разрешить относительно переменной x_j .

Действительно, если 1-ую строку этого соотношения

$$(-1)^{1+1} \frac{|\Lambda_{11}|}{|\Lambda|} (x_1 - \mu_1) + \dots + (-1)^{1+p} \frac{|\Lambda_{1p}|}{|\Lambda|} (x_p - \mu_p) = 0$$

умножить на $|\Lambda| / |\Lambda_{11}|$ и разрешить последнее равенство относительно x_1 , то получим уравнение I.1.

Проведя аналогичные преобразования, легко получить уравнение ЛСКР с.в. X_j на остальные компоненты вектора \vec{X} :

$$x_{j*} = \mu_j + \sum_{\substack{i=1 \\ i \neq j}}^p \beta_{ji*} (x_i - \mu_i).$$

Коэффициент

$$\beta_{ji*} = - \frac{(-1)^{j+i} |\Lambda_{ji}|}{|\Lambda_{jj}|}$$

называется коэффициентом линейной регрессии X_j на X_i за вычетом влияния всех остальных случайных переменных. Знак этого коэффициента указывает направление изменения X_j в среднем при возрастании значений X_i и фиксированных значениях всех остальных компонент. Если обозначить через $\tilde{\lambda}_{ij}$ элементы обратной матрицы Λ^{-1} , то коэффициент

регрессии можно представить в виде

$$\beta_{ji*} = -\tilde{\lambda}_{ji} / \tilde{\lambda}_{jj}.$$

В случае $p = 2$ уравнение регрессии X_1 на X_2 имеет вид

$$x_1 = \mu_1 + \rho_{12} \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2). \quad (I.2)$$

Лемма I.1. Пусть X_{1*} – линейная среднеквадратическая регрессия с.в. X_1 на (X_2, \dots, X_p) , тогда

1) ошибка линейного прогноза $\varepsilon_{1*} = \varepsilon_{1*(2\dots p)} = X_1 - X_{1*}$ не коррелирует с величинами X_2, \dots, X_p ;

2) ковариация ε_{1*} с X_1 равна

$$\text{Cov}(X_1, \varepsilon_{1*}) = \frac{|\Lambda|}{|\Lambda_{11}|}.$$

З а м е ч а н и е 3. Первое утверждение этой леммы можно интерпретировать в том смысле, что на разность $X_1 - X_{1*}$ не влияют с.в. X_2, \dots, X_p . Другими словами, всё влияние вектора с.в. (X_2, \dots, X_p) на X_1 содержится в функции линейной среднеквадратической регрессии. В дальнейшем эти соображения лягут в основу определений двух коэффициентов корреляции.

Остаточная дисперсия

Степень влияния вектора с.в. (X_2, \dots, X_p) на X_1 и, соответственно, качество прогноза X_1 описывается величиной

среднеквадратической ошибки $\mathbf{E} \varepsilon_{1*}^2 = \mathbf{E}(X_1 - X_{1*})^2$, которую обозначают σ_{1*}^2 или $\sigma_{1*(2\dots p)}^2$ и называют остаточной дисперсией.

Теорема 1.2.

а) Остаточная дисперсия равна

$$\sigma_{1*}^2 = \mathbf{E}(X_1 - X_{1*})^2 = \frac{|\Lambda|}{|\Lambda_{11}|} = \sigma_1^2 \frac{|\mathcal{P}|}{|\mathcal{P}_{11}|} = \frac{1}{\tilde{\lambda}_{11}},$$

где $\tilde{\lambda}_{11}$ – элемент обратной матрицы Λ^{-1} ;

б) дисперсия X_1 равна $\sigma_1^2 = \sigma_{1*}^2 + \mathbf{D} X_{1*}$.

З а м е ч а н и е 4. Второе утверждение теоремы показывает, что разброс с.в. X_1 складывается из случайной составляющей X_1 и случайных колебаний, обусловленных вектором (X_2, \dots, X_p) .

З а м е ч а н и е 5. Понятно, что сама по себе остаточная дисперсия не дает полной информации о качестве ЛСКР. Наибольший интерес представляет её отношение к полной дисперсии с.в. X_1 , которую можно воспринимать как остаточную дисперсию при линейном прогнозе, когда отсутствует какая-либо информация о значениях других случайных величин (этот прогноз равен среднему μ_1).

Пример. Матрица ковариаций и обратная к ней равны (при $p = 3$)

$$\Lambda = \begin{pmatrix} 5 & 4 & 3 \\ 4 & 4 & 3 \\ 3 & 3 & 3 \end{pmatrix}, \quad \Lambda^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 4/3 \end{pmatrix}.$$

Следовательно, уравнения регрессий (без учета средних значений) и остаточные дисперсии равны

$$\begin{aligned} x_1 &= x_2, & \sigma_{1*}^2 &= 1, \\ x_2 &= 0.5(x_1 + x_3), & \sigma_{2*}^2 &= 0.5, \\ x_3 &= 0.75x_2. & \sigma_{3*}^2 &= 0.75. \end{aligned}$$

Таким образом, ошибки прогноза X_1, X_2 и X_3 уменьшаются в 5 раз, в 8 раз и в 4 раза, соответственно.

Заметим, что величины X_1 и X_3 не связаны в уравнениях регрессии, хотя имеют высокий коэффициент корреляции $\rho_{13} = 3/\sqrt{15} \approx 0,77$. Причины такого явления будут объяснены ниже при рассмотрении частного коэффициента корреляции.

§ 2. Меры зависимости

Еще раз скажу, никто не обнимет
необъятного.

Множественный коэффициент корреляции.

Как было замечено после леммы I.1, стр. 8, функция ЛСКР включает в себя всё влияние величин X_2, \dots, X_p на X_1 . Вероятностной характеристикой такого влияния может служить коэффициент корреляции. Этим объясняется введение следующего определения.

О п р е д е л е н и е 2. Коэффициент корреляции $\rho_{1*(2\dots p)}$ (или ρ_{1*}) между X_1 и ЛСКР с.в. X_1 на (X_2, \dots, X_p)

$$\rho_{1*(2\dots p)} = \mathbf{Corr}(X_1, X_{1*(2\dots p)})$$

называется множественным коэффициентом корреляции между случайной величиной X_1 и вектором (X_2, \dots, X_p) .

Теорема 1.3. Множественный коэффициент корреляции вычисляется по одной из следующих формул:

$$\rho_{1*} = \sqrt{1 - \frac{\sigma_{1*(2 \dots p)}^2}{\sigma_1^2}} = \sqrt{1 - \frac{|\Lambda|}{\sigma_1^2 |\Lambda_{11}|}} = \sqrt{1 - \frac{|\mathcal{P}|}{|\mathcal{P}_{11}|}}.$$

З а м е ч а н и е 6. Учитывая последнее соотношение этой теоремы, легко показать, что для компоненты X_j квадрат множественного коэффициента корреляции равен

$$\rho_{j*}^2 = 1 - 1/\tilde{\rho}_{jj}, \quad (1.3)$$

где $\tilde{\rho}_{jj}$ — j -й диагональный элемент обратной матрицы корреляций \mathcal{P}^{-1} .

З а м е ч а н и е 7. Из первого соотношения этой теоремы, некоррелируемости случайных величин ε_{1*} и X_{1*} , а также из очевидного представления $X_1 = X_{1*} + \varepsilon_{1*}$ получаем, что квадрат множественного коэффициента корреляции

$$\rho_{1*}^2 = \frac{\mathbf{D} X_1 - \mathbf{D} \varepsilon_{1*}}{\mathbf{D} X_1} = \frac{\mathbf{D} X_{1*}}{\mathbf{D} X_1}. \quad (1.4)$$

Величину $\rho_{1*}^2 \cdot 100\%$ называют коэффициентом детерминации и в соответствии со вторым равенством (1.4) интерпретируют как долю дисперсии X_1 , обусловленную влиянием переменных X_2, \dots, X_p .

В рамках рассмотренного выше **примера** множественные коэффициенты корреляции и соответствующие коэффи-

циенты детерминации равны

$$\begin{aligned}\rho_{1*} &= 0.89, & \rho_{1*}^2 \cdot 100\% &= 80\%, \\ \rho_{2*} &= 0.94, & \rho_{2*}^2 \cdot 100\% &= 87.5\%, \\ \rho_{3*} &= 0.87, & \rho_{3*}^2 \cdot 100\% &= 75\%.\end{aligned}$$

Свойства множественного коэффициента корреляции.

- 1) $0 \leq \rho_{1*} \leq 1$.
- 2) $\rho_{1*} = 1$ только в том случае, если случайная величина X_1 стохастически линейно зависит от случайных величин X_2, \dots, X_p .
- 3) $\rho_{1*} = 0$ только в том случае, если случайная величина X_1 не коррелирует ни с одной из случайных величин X_2, \dots, X_p .
- 4) Множественный коэффициент корреляции представляет собой максимальный коэффициент корреляции между X_1 и всевозможными линейными комбинациями X_2, \dots, X_p . Другими словами, этот коэффициент показывает степень максимально достижимой связи между X_1 и вектором (X_2, \dots, X_p) .
- 5) В двумерном случае множественный коэффициент корреляции равен по абсолютной величине полному коэффициенту корреляции: $\rho_{1*2} = |\rho_{12}|$.
- 6) Для трех случайных величин

$$\rho_{1*(23)} = \sqrt{\frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{12}\rho_{13}\rho_{23}}{1 - \rho_{23}^2}}.$$

Частный коэффициент корреляции

Воспользуемся снова первым утверждением леммы I.1, стр. 8, применительно к остаткам $\varepsilon_{1*(3\dots p)}$ и $\varepsilon_{2*(3\dots p)}$ линейного прогноза X_1 и X_2 по X_3, \dots, X_p . Как следует из этой леммы,

указанные остатки содержат в себе только сами случайные составляющие X_1, X_2 , а также их возможное взаимодействие друг с другом, причем без влияния всех остальных величин. Эти соображения приводят к следующему определению.

О п р е д е л е н и е 3. Коэффициент корреляции между остатками $\varepsilon_{1*(3\dots p)}$ и $\varepsilon_{2*(3\dots p)}$ линейной среднеквадратической регрессии случайных величин X_1 и X_2 по значениям случайного вектора (X_3, \dots, X_p) называется частным коэффициентом корреляции между X_1 и X_2 за вычетом влияния (X_3, \dots, X_p) и обозначается $\rho_{12*} = \rho_{12*(3\dots p)}$.

Т е о р е м а 1.4. Частный коэффициент корреляции равен

$$\rho_{12*} = \frac{|\Lambda_{12}|}{\sqrt{|\Lambda_{11}| |\Lambda_{22}|}} = \frac{|\mathcal{P}_{12}|}{\sqrt{|\mathcal{P}_{11}| |\mathcal{P}_{22}|}}.$$

З а м е ч а н и е 8. Формула для вычисления частного коэффициента корреляции очень похожа на формулу для полного парного коэффициента корреляции $\rho_{12} = \lambda_{12}/\sqrt{\lambda_{11}\lambda_{22}}$.

Свойства частного коэффициента корреляции.

- 1) $-1 \leq \rho_{12*} \leq 1$;
- 2) $\rho_{12*} = \pm 1$ тогда и только тогда, когда полный коэффициент корреляции $\rho_{12} = \pm 1$;
- 3) при $p = 3$

$$\rho_{12*3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1 - \rho_{13}^2)(1 - \rho_{23}^2)}}; \quad (1.5)$$

4) для многомерного нормального распределения частный коэффициент корреляции совпадает с условным коэффициентом корреляции между X_1 и X_2 при фиксированных значениях всех остальных величин (см. теорему IV.10, стр. 79).

З а м е ч а н и е 9. Формула для вычисления частного коэффициента корреляции между i -ой и j -ой компонентами вектора \vec{X} легко получается из теоремы I.4 перестановкой строк и столбцов матрицы Λ :

$$\begin{aligned} \rho_{ij*} &= (-1)^{i+j-1} \frac{|\Lambda_{ij}|}{\sqrt{|\Lambda_{ii}| |\Lambda_{jj}|}} = (-1)^{i+j-1} \frac{|\mathcal{P}_{ij}|}{\sqrt{|\mathcal{P}_{ii}| |\mathcal{P}_{jj}|}} = \\ &= -\frac{\tilde{\lambda}_{ij}}{\sqrt{\tilde{\lambda}_{ii} \tilde{\lambda}_{jj}}}, \end{aligned}$$

где $(\tilde{\lambda}_{ij})_{i,j=1}^p$ – обратная матрица Λ^{-1} .

Зри в корень.

З а м е ч а н и е 10. Элементарными преобразованиями формул для частных коэффициентов корреляции и остаточных дисперсий легко получается следующее представление коэффициента регрессии

$$\beta_{ji*} = \rho_{ji*} \frac{\sigma_{j*}}{\sigma_{i*}}$$

(сравните с формулой I.2, стр. 8, для коэффициента регрессии X_1 на X_2). Отсюда можно сделать вывод, что влияние одной случайной величины на другую описывает не полный,

а частный коэффициент корреляции.

З а м е ч а н и е 11. Вычисление всех частных коэффициентов корреляции можно производить с использованием соотношения (1.5). Например, если вычислены все коэффициенты ρ_{ij*k} за вычетом одной переменной, то частные коэффициенты корреляции за вычетом двух величин равны

$$\rho_{ij*kl} = \frac{\rho_{ij*k} - \rho_{il*k}\rho_{jl*k}}{\sqrt{(1 - \rho_{il*k}^2)(1 - \rho_{jl*k}^2)}}.$$

З а м е ч а н и е 12. Свойство 4) является определяющим при практической интерпретации частных коэффициентов корреляции. Считается (даже вне рамок нормальной модели), что этот коэффициент несет информацию о связи между двумя показателями при фиксированных всех остальных.

Продолжая рассмотрение **примера**, легко находим полные парные коэффициенты корреляции

$$\begin{aligned} \rho_{12} &= 4/\sqrt{20} \approx 0.89, & \rho_{13} &= 3/\sqrt{15} \approx 0.77, & \rho_{23} &= 3/\sqrt{12} \approx 0.87 \\ \text{и частные коэффициенты корреляции} \\ \rho_{12*3} &= \sqrt{2}/2 \approx 0.71, & \rho_{13*2} &= 0, & \rho_{23*1} &= \sqrt{3/8} \approx 0.61. \end{aligned}$$

Исчезновение корреляции между X_1 и X_3 показывает, что их высокое взаимовлияние (высокое значение полного коэффициента корреляции) обусловлено исключительно влиянием на них третьей величины X_2 . При увеличении X_2 происходит одновременное увеличение как X_1 , так и X_3 , что ложно воспринимается как увеличение X_1 при увеличении X_3 .

В следующих примерах приводятся два варианта изменения выводов о взаимосвязи величин при переходе от

полных коэффициентов корреляции к частным. Конкретные цифры в этих примерах выдуманы и не обязательно соответствуют действительности.

Пример ложной корреляции. Исследования величины ущерба от пожара (X_1) выявили сильную зависимость значения X_1 от факта приезда или не приезда пожарной машины. Коэффициент корреляции между X_1 и случайной величиной X_2 , принимающей два значения 1 и 0 в зависимости от того, приехала или нет машина пожарной помощи, $\rho_{12} = 0.7$, то есть вызов пожарных способствует увеличению ущерба. Парадоксальность ситуации объясняется тем, что обе величины имеют высокую положительную корреляцию с третьей неучтенной здесь характеристикой, равной стоимости объекта до пожара. При значениях $\rho_{13} = \rho_{23} = 0.9$ по формуле (I.5), стр. 13, получаем $\rho_{12*3} = -0.58$.

Пример скрытой корреляции. Постепенное увеличение количества контролеров (X_1) на трамвайных маршрутах не дало соответствующего увеличения поступлений от штрафов (X_2), налагаемых на „зайцев“ (коэффициент корреляции $\rho_{12} \simeq 0$). В данном случае это можно объяснить высокой противоположно направленной корреляцией обеих величин с количеством транспортных „зайцев“ : $\rho_{13} = -0.7$, $\rho_{23} = 0.7$. В результате имеем $\rho_{12*3} = 0.96$.

§ 3. Каноническая корреляция

В спертom воздухе при всем старании не отдышишься.

На практике часто оказывается, что вектор наблюдаемых характеристик может быть разделен на две группы связанных между собой показателей, при этом перед исследователем

стоит задача изучения взаимосвязей между показателями в различных группах. Например, при изучении заболеваемости на предприятиях химической промышленности представляет интерес влияние всего спектра примесей в окружающей среде (первая группа показателей) на весь спектр возможных легочных заболеваний (вторая группа показателей). Если число показателей достаточно велико (что весьма характерно для исследований, когда не знают, что ищут), то попарное сравнение показателей приводит к слишком большому массиву корреляционных отношений, недоступному для конструктивной интерпретации. В этой ситуации возникает задача вычисления в некотором смысле обобщающих характеристик взаимосвязи. Следующая концепция корреляции между двумя совокупностями случайных величин принадлежит известному специалисту в области математической статистики Гарольду Хотеллингу. Он предложил рассматривать максимальное значение коэффициента корреляции между линейными комбинациями компонент двух исследуемых совокупностей случайных величин. Если одна из групп содержит всего одну компоненту, то эта концепция естественным образом приводит к определению множественного коэффициента корреляции.

О п р е д е л е н и е 4. Пусть $\vec{X} = (\vec{X}'_1, \vec{X}'_2)'$ – p -мерный вектор, составленный из двух векторов \vec{X}_1, \vec{X}_2 размерностей p_1 и p_2 , соответственно ($p = p_1 + p_2$). Для удобства будем считать, что $p_1 \leq p_2$.

1) Пара линейных комбинаций $U_1 = \vec{a}'_1 \vec{X}_1$ и $V_1 = \vec{b}'_1 \vec{X}_2$ определяет первую пару канонических величин, если каждая из этих с.в. имеет единичную дисперсию и коэффициент кор-

реляции $r_1 = \mathbf{Corr}(U_1, V_1)$ максимален среди всех линейных комбинаций векторов \vec{X}_1, \vec{X}_2 .

к) Пара линейных комбинаций $U_k = \vec{a}'_k \vec{X}_1$ и $V_k = \vec{b}'_k \vec{X}_2$ определяет *k-ую пару канонических величин*, если каждая из этих с.в. имеет единичную дисперсию и коэффициент корреляции $r_k = \mathbf{Corr}(U_k, V_k)$ максимален среди всех линейных комбинаций векторов \vec{X}_1, \vec{X}_2 , не коррелирующих с первыми $k - 1$ парами канонических величин.

k -й по порядку коэффициент r_k называется *k-ой канонической корреляцией*.

Таким образом, канонические величины не коррелируют между собой и позволяют находить максимальные корреляционные связи между двумя группами случайных величин.

Для решения задачи построения канонических величин рассмотрим разбиение матрицы ковариаций

$$\Lambda = \mathbf{Cov}(\vec{X}) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

соответствующее разбиению случайного вектора \vec{X} , где подматрица Σ_{ii} – матрица ковариаций \vec{X}_i размерности $p_i \cdot p_i$, $i = 1, 2$, подматрицы Σ_{12} и $\Sigma_{21} = \Sigma'_{12}$ – матрицы взаимных ковариаций \vec{X}_1 и \vec{X}_2 размерностей $p_1 \cdot p_2$ и $p_2 \cdot p_1$, соответственно.

Теорема I.5. 1) Для каждой пары случайных векторов \vec{X}_1, \vec{X}_2 размерностей p_1 и p_2 ($p_1 \leq p_2$) найдется не более p_1 ненулевых коэффициентов канонической корреляции.

2) Коэффициент k -ой канонической корреляции ($k = 1, \dots, p_1$) равен k -ому по порядку решению уравнения p -ой степени относительно переменной r

$$\begin{vmatrix} -r\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -r\Sigma_{22} \end{vmatrix} = 0 .$$

Это уравнение имеет не более $2p_1$ ненулевых решений, причем p_1 корень положителен, а остальные p_1 получаются из них умножением на (-1) .

3) Векторы, определяющие k -ую пару канонических величин, удовлетворяют уравнению

$$\begin{pmatrix} -r_k \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -r_k \Sigma_{22} \end{pmatrix} \begin{pmatrix} \vec{a} \\ \vec{b} \end{pmatrix} = 0$$

при условиях $\vec{a}'\Sigma_{11}\vec{a} = 1, \quad \vec{b}'\Sigma_{22}\vec{b} = 1 .$

Т е о р е м а 1.6. 1) Квадраты канонических корреляций r_k^2 равны собственным значениям матрицы

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

2) Векторы \vec{a}_k , определяющие первые компоненты пар канонических величин, равны соответствующим собственным векторам при условиях

$$\vec{a}'_k \Sigma_{11} \vec{a}_k = 1,$$

вторые компоненты вычисляются по формуле

$$\vec{b}_k = \frac{1}{r_k} \Sigma_{22}^{-1} \Sigma_{21} \vec{a}_k.$$

Пример интерпретации канонических величин.

Коэффициенты первых пар канонических величин дают некоторую информацию о показателях, участвующих в формировании того или иного коэффициента корреляции. Рассмотрим принципы выделения этой информации на конкретном примере.

Исследовалось влияние химического состава удобрений на потребительские качества одной из разновидностей ягодных культур. Качество ягод оценивалось по их весу (*Ves*), объему (*Vol*), содержанию витаминов А и С, а также по содержанию белков (*B*). В показатели химического состава удобрений входили: содержание солей калия (*Ka*), магния (*Mg*), кальция (*Ca*), редкоземельных элементов (*L*) и тяжелых металлов (*Met*). Было установлено, что первые два канонических коэффициента корреляции имеют высокое значение $r_1 = 0.87$, $r_2 = 0.73$. Канонические величины задаются уравнениями

$$\begin{aligned}U_1 &= 1.12 * Ves + 1.51 * Vol + 0.21 * A + 0.15 * C + 0.47 * B, \\V_1 &= 0.98 * Ka + 0.34 * Mg - 0.25 * Ca - 0.23 * L + 1.36 * Met,\end{aligned}$$

$$\begin{aligned}U_2 &= 0.36 * Ves - 0.11 * Vol + 0.86 * A + 1.47 * C + 1.38 * B, \\V_2 &= -0.14 * Ka + 0.74 * Mg + 0.89 * Ca - 1.04 * L - 0.14 * Met.\end{aligned}$$

Таким образом, анализ первой пары показывает, что вес и объем ягод (– показатели с наибольшими по абсолютной величине коэффициентами в компоненте U_1) сильно зависят от количества тяжелых металлов и солей калия (соответствующие показатели компоненты V_1). Другими словами, можно утверждать, что „товарный вид“ ягод определяется содержанием металлов в применяемом удобрении.

С другой стороны, вторая пара канонических величин показывает, что для улучшения „питательных качеств“ ягод необходимо применять удобрения с повышенным содержанием редкоземельных элементов и солей кальция и магния.

§ 4. Эллипсоид рассеяния

Бросая в воду камешки, смотри на круги, ими образуемые; иначе такое бросание будет пустою забавою.

Эллипсоид рассеяния служит геометрической характеристикой разброса значений случайного вектора.

О п р е д е л е н и е 5. Эллипсоид $(\vec{x} - \vec{m})'A(\vec{x} - \vec{m}) \leq c^2$ называется эллипсоидом рассеяния случайного вектора \vec{X} , если случайный вектор \vec{Y} , равномерно распределенный внутри этого эллипсоида, имеет те же первые и вторые моменты, что и вектор \vec{X} :

$$\mathbf{E} \vec{Y} = \mathbf{E} \vec{X}, \quad \mathbf{Cov}(\vec{Y}) = \mathbf{Cov}(\vec{X}).$$

Т е о р е м а 1.7. Эллипсоид рассеяния p -мерного случайного вектора с невырожденной матрицей ковариаций Λ и вектором средних $\vec{\mu}$ задается соотношением

$$(\vec{x} - \vec{\mu})' \Lambda^{-1} (\vec{x} - \vec{\mu}) \leq p + 2.$$

З а м е ч а н и е 13. Если матрица ковариаций вырождена, то, как следует из теоремы А.1, стр. 91, случайный вектор

\vec{X} сосредоточен в пространстве меньшей размерности, где и надо определять эллипсоид рассеяния.

З а м е ч а н и е 14. В теореме IV.8, стр. 77, устанавливается, что для нормально распределенного вектора \vec{X} квадратичная форма $(\vec{X} - \vec{\mu})' \Lambda^{-1} (\vec{X} - \vec{\mu})$ имеет хи-квадрат распределение с p степенями свободы. Поэтому вероятность попадания нормального вектора в соответствующий эллипсоид рассеяния равна $K_p(p+2)$, где K_p - функция хи-квадрат распределения с p степенями свободы.

З а м е ч а н и е 15. Для нормального случайного вектора можно также показать, что условное распределение \vec{X} при условии, что он лежит на границе эллипсоида, подобного эллипсоиду рассеяния, будет равномерным. Поэтому такие эллипсоиды называют эллипсоидами равных вероятностей.

В качестве меры разброса случайного вектора может быть выбрана величина, равная объему эллипсоида рассеяния, который обратно пропорционален корню квадратному из определителя матрицы Λ^{-1} (см. соотношение В.3, стр. 96).

О п р е д е л е н и е 6. Обобщенной дисперсией случайного вектора \vec{X} с ковариационной матрицей Λ называется величина

$$\mathfrak{S}^2 = |\Lambda|.$$

Изучим свойства эллипсоида в двух простейших частных случаях.

[$p = 1$] В одномерном случае эллипсоид представляет собой отрезок $|x - \mu| \leq \sqrt{3}\sigma_1$. Обобщенная дисперсия совпа-

дает с дисперсией X_1 , $-\mathfrak{E}^2 = \sigma_1^2$, а вероятность попадания в этот эллипсоид (отрезок) нормального случайного вектора приблизительно равна 0.917.

[$p = 2$] Если средние значения и дисперсии двумерного случайного вектора (Y, X) равны $\mu_y, \mu_x, \sigma_y^2, \sigma_x^2$, соответственно, а коэффициент корреляции равен ρ , то эллипс рассеяния можно записать в виде

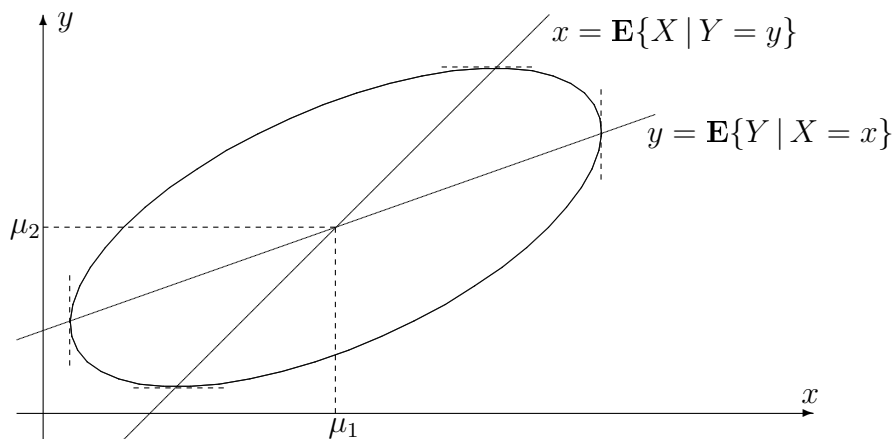
$$\frac{(y - \mu_y)^2}{\sigma_y^2} - 2\rho \frac{(y - \mu_y)(x - \mu_x)}{\sigma_y \sigma_x} + \frac{(x - \mu_x)^2}{\sigma_x^2} = 4(1 - \rho^2). \quad (\text{I.6})$$

Таким образом,

1) если $\rho = 0$, то оси эллипса параллельны осям координат, а если, вдобавок, сл.в. X и Y имеют одинаковые дисперсии ($\sigma_x^2 = \sigma_y^2 = \sigma^2$), то эллипс превращается в окружность радиуса 2σ ;

2) если $|\rho| \rightarrow 1$, то эллипс вытягивается вдоль линии регрессии $y = \mu_y + \text{sign}(\rho)(x - \mu_x)\sigma_y / \sigma_x$

3) вероятность попадания двумерного нормального сл.вектора в эллипс рассеяния равна 0.865.



Приведенный здесь рисунок иллюстрирует связь меж-

ду эллипсом рассеяния и обеими линиями регрессии. Точная формулировка этой связи приведена в следующей теореме.

Т е о р е м а 1.8.

1) Линия регрессии X на Y проходит через точки касания с эллипсом рассеяния двух прямых линий, параллельных оси OX .

2) Линия регрессии Y на X проходит через точки касания с эллипсом рассеяния двух прямых линий, параллельных оси OY .

З а м е ч а н и е 16. Главная ось эллипса рассеяния представляет собой ещё одну линию регрессии. Эта линия строится из условия достижения минимума среднеквадратического расстояния от реализаций двумерного случайного вектора (X, Y) перпендикулярно линии регрессии:

$$\min_{b,c} \frac{\mathbf{E}(Y - bX - c)^2}{1 + b^2} .$$

Такая линия называется линией ортогональной регрессии. Легко показать, что уравнение этой линии имеет вид

$$\frac{y - \mu_y}{\sigma_y} = \text{sign}(\rho) \frac{x - \mu_x}{\sigma_x} .$$

Для любителей курса аналитической геометрии не составит большого труда установить совпадение этой линии с главной

осью эллипса рассеяния.

§ 5. Нелинейная регрессия

Если не ограничиваться множеством линейных функций при аппроксимации значений одной с.в. по реализациям ряда других с.в., приходим к определению (нелинейной) среднеквадратической регрессии.

О п р е д е л е н и е 7. Функция $\mu_1^*(x_2, \dots, x_p)$, зависящая от реализаций с.в. X_2, \dots, X_p , называется среднеквадратической регрессией (СКР) X_1 на X_2, \dots, X_p , если

$$\mathbf{E}(X_1 - \mu_1^*(X_2, \dots, X_p))^2 = \min_h \mathbf{E}(X_1 - h(X_2, \dots, X_p))^2,$$

где минимум берется по всем функциям $h(x_2, \dots, x_p)$.

Т е о р е м а 1.9. СКР с.в. X_1 на (X_2, \dots, X_p) равна

$$\mu_{1*}(x_2, \dots, x_p) = \mathbf{E}\{X_1 \mid X_2 = x_2, \dots, X_p = x_p\} -$$

условному математическому ожиданию X_1 при фиксированных значениях вектора (X_2, \dots, X_p) .

Если $f_p(x_1, \dots, x_p)$ – совместная функция плотности вектора (X_1, \dots, X_p) , а $f_{p-1}(x_2, \dots, x_p)$ – функция плотности вектора (X_2, \dots, X_p) , то

$$\mu_{1*}(x_2, \dots, x_p) = \int_{-\infty}^{\infty} x_1 f_p(x_1, \dots, x_p) dx_1 / f_{p-1}(x_2, \dots, x_p).$$

Эта теорема оправдывает обозначение функции СКР

символом μ , закрепленным за средним значением.

З а м е ч а н и е 17. Как видно из теоремы, для построения функции СКР необходимо полностью знать функцию распределения вектора \vec{X} . На практике такое требование выполняется только в случае, когда по каким-либо причинам можно предположить, что это распределение нормально. Ниже будет показано, что в этом случае СКР линейна.

Если в формуле для множественного коэффициента корреляции линейную регрессию заменить на нелинейную, то придем к определению коэффициента функциональной зависимости X_1 от X_2, \dots, X_p . Здесь принято рассматривать нелинейный аналог коэффициента детерминации (см. теорему I.3, стр. 11, и формулу (I.4), стр. 11).

О п р е д е л е н и е 8. Коэффициент

$$\eta_{1*(2\dots p)} = \sqrt{1 - \frac{\mathbf{E}(X_1 - \mu_{1*}(X_2, \dots, X_p))^2}{\mathbf{D} X_1}}$$

называется корреляционным отношением.

Т е о р е м а I.10. Если μ_{1*} – функция СКР X_1 на (X_2, \dots, X_p) , то квадрат корреляционного отношения равен

$$\eta_{1*(2\dots p)}^2 = \frac{\mathbf{D} \mu_{1*}}{\mathbf{D} X_1}.$$

З а м е ч а н и е 18. Величина $\eta_{1*}^2 \cdot 100\%$ показывает долю разброса значений с.в. X_1 , обусловленную влиянием на неё

сл. величин X_2, \dots, X_p .

Свойства корреляционного отношения.

Если ρ_{1*} – множественный коэффициент корреляции X_1 с вектором (X_2, \dots, X_p) , тогда

1) $\rho_{1*}^2 \leq \eta_{1*}^2 \leq 1$;

2) если с.в. X_1 не зависит от (X_2, \dots, X_p) , тогда $\eta_{1*}^2 = 0$ (обратное не всегда верно);

3) $\eta_{1*}^2 = 1$ тогда и только тогда, когда случайная величина X_1 функционально представима через величины X_2, \dots, X_p ;

4) $\eta_{1*}^2 = \rho_{1*}^2$ только в том случае, если среднеквадратическая регрессия линейна;

5) если вектор $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \Lambda)$, то СКР линейна.

З а м е ч а н и е 19. Свойства 1) и 4) показывают, что отношение ρ_{1*}^2/η_{1*}^2 может служить мерой „нелинейности“ регрессии. Очень часто это отношение используют для проверки гипотезы линейности регрессии.

З а м е ч а н и е 20. Свойство 1) часто переносят на выборочные оценки коэффициентов корреляции и детерминации. Однако это не вполне верно, так как при стандартном методе построения оценки для η_{1*}^2 исходные данные группируются и заменяются центрами соответствующих групп. Отсюда понятно, что свойство 1) заведомо будет выполняться, только если коэффициент множественной корреляции также вычисляется по группированным данным.

Доказательства

Доказательство Теоремы I.1.

Очевидно, задача построения ЛСКР эквивалентна задаче минимизации функции

$$\begin{aligned} H(b_0, b_2, \dots, b_p) &= \\ &= \mathbf{E}((X_1 - \mu_1) - b_0 - b_2(X_2 - \mu_2) - \dots - b_p(X_p - \mu_p))^2 = \\ &= b_0^2 + \lambda_{11} - 2 \sum_{i=2}^p b_i \lambda_{1i} + \sum_{i=2}^p b_i^2 \lambda_{ii} + 2 \sum_{i=3}^p b_i \sum_{2 \leq k < i} b_k \lambda_{ki}, \end{aligned}$$

где учтено, что среднее $\mathbf{E}(X_i - \mu_i) = 0$, $i = \overline{1, p}$, а среднее произведения $\mathbf{E}(X_k - \mu_k)(X_i - \mu_i) = \lambda_{ki}$, $k, i = \overline{1, p}$.

Имеем, во-первых,

$$H'_{b_0} = 2b_0.$$

Следовательно, уравнение $H'_{b_0} = 0$ приводит к значению свободного параметра $b_0 = 0$. Другие производные равны

$$H'_{b_i} = -2(\lambda_{1i} - b_2 \lambda_{2i} - \dots - b_p \lambda_{pi}).$$

Таким образом, для нахождения экстремальных точек H необходимо решить систему уравнений

$$\lambda_{1i} = b_2 \lambda_{2i} + \dots + b_p \lambda_{pi}, \quad i = \overline{2, p}. \quad (\text{I.7})$$

Запишем эту систему в матричной форме, для чего рассмотрим матрицу Λ_{11} , полученную из матрицы Λ вычеркиванием первой строки и первого столбца, а также $(p-1)$ -мерные векторы $\vec{b} = (b_2, \dots, b_p)'$ и $\vec{\lambda}_{1*} = (\lambda_{12}, \dots, \lambda_{1p})'$. В этих обозначениях система (I.7) имеет вид

$$\vec{\lambda}_{1*} = \Lambda_{11} \vec{b}.$$

Ее решение при условии $\Lambda > 0$ равно

$$\vec{\beta} = (\Lambda_{11})^{-1} \vec{\lambda}_{1*}. \quad (I.8)$$

Докажем сначала, что решение полученной системы доставляет минимум функции H . Для этого покажем, что матрица вторых производных H положительно определена.

Из предыдущих выкладок получаем

$$H''_{b_0 b_0} = 2, \quad H''_{b_0 b_j} = 0, \quad H''_{b_i b_j} = 2\lambda_{ij}.$$

Следовательно, матрица вторых производных функции H равна

$$2 \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & \Lambda_{11} & & \\ 0 & & & \end{pmatrix},$$

и ее положительная определенность есть следствие положительной определенности Λ_{11} .

Теперь выведем более удобную формулу для вычисления коэффициентов линейной регрессии. Общий член матрицы $(\Lambda_{11})^{-1}$ равен

$$(-1)^{i+j} \frac{|(\Lambda_{11})_{ij}|}{|\Lambda_{11}|},$$

где индексы $i, j = 1, \dots, p-1$, указывают на номера столбцов и строк матрицы Λ_{11} . В исходной матрице Λ они будут соответствовать $(i+1)$ -ой строке и $(j+1)$ -ому столбцу. Обратим внимание также на то, что i -ые элементы векторов $\vec{\beta}$ и $\vec{\lambda}_{1*}$ равны β_{i+1} и $\lambda_{1(i+1)}$ соответственно. Поэтому

$$\beta_{i+1} = \sum_{j=1}^{p-1} (-1)^{i+j} \frac{|(\Lambda_{11})_{ij}|}{|\Lambda_{11}|} \lambda_{1(j+1)}.$$

Заметим, что при замене порядка вычеркивания строк в матрице Λ справедливо равенство

$$(\Lambda_{11})_{ij} = (\Lambda_{(i+1)1})_{1j}.$$

В матрице $\Lambda_{(i+1)1}$ элемент $\lambda_{1(j+1)}$ находится в j -ом столбце, поэтому его алгебраическое дополнение равно $(-1)^{1+j}(\Lambda_{(i+1)1})_{1j}$. Таким образом, разлагая определитель матрицы $\Lambda_{(i+1)1}$ по его первой строке, получаем соотношение

$$\|\Lambda_{(i+1)1}\| = \sum_{j=1}^{p-1} (-1)^{1+j} \|(\Lambda_{(i+1)1})_{1j}\| \lambda_{1(j+1)},$$

которое дает окончательное выражение для коэффициентов регрессии

$$\beta_i = (-1)^i \frac{\|\Lambda_{1i}\|}{\|\Lambda_{11}\|}, \quad i = \overline{2, p}. \quad (I.9)$$

⊠

Доказательство Леммы I.1.

Запишем разность

$$\varepsilon_{1*} = X_1 - X_{1*} = \sum_{j=1}^p (-1)^{j+1} \frac{\|\Lambda_{1j}\|}{\|\Lambda_{11}\|} (X_j - \mu_j). \quad (I.10)$$

Ковариация

$$\text{Cov}(X_l, \varepsilon_{1*}) = \sum_{j=1}^p (-1)^{j+1} \frac{\|\Lambda_{1j}\|}{\|\Lambda_{11}\|} \lambda_{lj}.$$

Утверждения леммы следуют теперь из формулы разложения определителя матрицы Λ по первой строке (см. Приложение А).

⊠

Доказательство Теоремы I.2.

Представим квадрат ошибки следующим образом: $\varepsilon_{1*}^2 = \varepsilon_{1*}(X_1 - \mu_1) - \varepsilon_{1*}(X_{1*} - \mu_1)$. Так как среднее $\mathbf{E} \varepsilon_{1*} = 0$ и разность $(X_{1*} - \mu_1)$ зависит только от случайных величин X_2, \dots, X_p , то в силу первого утверждения леммы I.1 среднее значение $\mathbf{E} \varepsilon_{1*}(X_{1*} - \mu_1) = 0$. Поэтому остаточная дисперсия

$$\sigma_{1*}^2 = \mathbf{E} \varepsilon_{1*}^2 = \mathbf{Cov}(X_1, \varepsilon_{1*}) = \frac{|\Lambda|}{|\Lambda_{11}|},$$

по второму утверждению леммы I.1, стр. 8.

Последнее равенство теоремы устанавливается простыми алгебраическими соотношениями.

⊠

Доказательство Теоремы I.3.

Пусть X_{1*} – функция линейной регрессии X_1 на (X_2, \dots, X_p) , тогда ковариация

$$\begin{aligned} \mathbf{Cov}(X_1, X_{1*}) &= \mathbf{E}(X_1 - \mu_1)(X_{1*} - \mu_1) = \\ &= \mathbf{E}(X_1 - \mu_1)[(X_1 - \mu_1) - (X_1 - X_{1*})] = \\ &= \mathbf{D} X_1 - \mathbf{Cov}(X_1, \varepsilon_{1*}). \end{aligned}$$

С учетом леммы I.1, стр. 8, получаем

$$\mathbf{Cov}(X_1, X_{1*}) = \sigma_1^2 - \frac{|\Lambda|}{|\Lambda_{11}|}.$$

Дисперсия X_{1*} равна

$$\begin{aligned} \mathbf{D} X_{1*} &= \mathbf{E}(X_{1*} - \mu_1)^2 = \mathbf{E}[(X_1 - \mu_1) - (X_1 - X_{1*})]^2 = \\ &= \mathbf{E}(X_1 - \mu_1)^2 - 2\mathbf{E}(X_1 - \mu_1)\varepsilon_{1*} + \mathbf{E}\varepsilon_{1*}^2 = \\ &= \mathbf{D} X_1^2 - 2\mathbf{Cov}(X_1, \varepsilon_{1*}) + \mathbf{D} \varepsilon_{1*} = \\ &= \sigma_1^2 - \frac{|\Lambda|}{|\Lambda_{11}|}. \end{aligned}$$

Поэтому коэффициент корреляции

$$\mathbf{Corr}(X_1, X_{1*}) = \frac{\mathbf{Cov}(X_1, X_{1*})}{\sqrt{\mathbf{D}(X_1) \mathbf{D}(X_{1*})}} = \sqrt{1 - \frac{|\Lambda|}{\sigma_1^2 |\Lambda_{11}|}}.$$

⊠

Доказательство свойств ρ_{1} .*

Первые два свойства очевидным образом вытекают из определения ρ_{1*} как простого коэффициента корреляции случайной величины X_1 с линейной функцией, а также из формул для вычисления.

Третье свойство следует из того, что $\rho_{1*} = 0$ тогда и только тогда, когда определитель матрицы $|\mathcal{P}| = |\mathcal{P}_{11}|$, что для положительно определенных матриц возможно лишь в случае, если равны нулю все элементы первой строки и первого столбца матрицы \mathcal{P} , за исключением первого элемента, который равен 1.

Для доказательства свойства 4) заметим, что любой коэффициент корреляции не изменяется при умножении рассматриваемых величин на положительную константу. Поэтому задача максимизации

$$\mathbf{Corr}(X_1, \vec{a}' \vec{X}_*) = \frac{\mathbf{Cov}(X_1, \vec{a}' \vec{X}_*)}{\sigma_1 \sqrt{\mathbf{D} \vec{a}' \vec{X}_*}},$$

где $\vec{a}' = (a_2, \dots, a_p)$, $\vec{X}_* = (X_2, \dots, X_p)'$, эквивалентна задаче максимизации $\mathbf{Cov}(X_1, \vec{a}' \vec{X}_*)$ при условии, что $\mathbf{D} \vec{a}' \vec{X}_* = \vec{a}' \Lambda_{11} \vec{a} = 1$.

Ковариация $\mathbf{Cov}(X_1, \vec{a}' \vec{X}_*) = \vec{a}' \vec{\lambda}_{1*}$, где $\vec{\lambda}_{1*} = (\lambda_{12}, \dots, \lambda_{1p})'$. Рассмотрим функцию Лагранжа

$$H(\vec{a}) = \vec{a}' \vec{\lambda}_{1*} - \gamma \sum_{i,j=2}^p a_i a_j \lambda_{ij}.$$

Производная H по переменной a_i равна

$$H'_{a_i} = \lambda_{1i} - 2\gamma \sum_{j=2}^p a_j \lambda_{ij}.$$

Поэтому система уравнений для нахождения максимума коэффициента корреляции в матричной форме приобретает вид

$$\begin{cases} 2\gamma \vec{a}' \Lambda_{11} = \vec{\lambda}'_{1*}, \\ \vec{a}' \Lambda_{11} \vec{a} = 1. \end{cases}$$

Решение этой системы доставляют вектор

$$\vec{a} = \frac{\Lambda_{11}^{-1} \vec{\lambda}_{1*}}{2\gamma}$$

и коэффициент Лагранжа $2\gamma = \pm \sqrt{\vec{\lambda}'_{1*} \Lambda_{11}^{-1} \vec{\lambda}_{1*}}$.

Легко видеть, что матрица вторых производных функции H равна $-2\gamma \Lambda_{11}$. Поэтому максимум коэффициента корреляции достигается при выборе положительного значения γ . Величина максимального коэффициента корреляции равна

$$\text{Corr}(X_1, \vec{a}' \vec{X}_*) = \frac{\vec{a}' \vec{\lambda}_{1*}}{\sigma_1} = \frac{\vec{\lambda}'_{1*} \Lambda_{11}^{-1} \vec{\lambda}_{1*}}{\sigma_1 2\gamma} = \frac{\sqrt{\vec{\lambda}'_{1*} \Lambda_{11}^{-1} \vec{\lambda}_{1*}}}{\sigma_1}.$$

С выражением $\Lambda_{11}^{-1} \vec{\lambda}_{1*}$ мы уже встречались (см. (I.8), стр.29) при вычислении коэффициентов регрессии. Как было показано в (I.9), стр. 30,

$$\Lambda_{11}^{-1} \vec{\lambda}_{1*} = \left((-1)^i \frac{|\Lambda_{1i}|}{|\Lambda_{11}|} \right)_{i=2, \dots, p}.$$

Таким образом, в числителе искомого коэффициента корреляции под корнем стоит выражение

$$\vec{\lambda}'_{1*} \Lambda_{11}^{-1} \vec{\lambda}_{1*} = \frac{1}{|\Lambda_{11}|} \sum_{i=2}^p (-1)^i |\Lambda_{1i}| \lambda_{1i} = \lambda_{11} - \frac{|\Lambda|}{|\Lambda_{11}|}$$

(здесь мы снова использовали разложение определителя $|\Lambda|$ по первой строке). Замечая, что $\lambda_{11} = \sigma_1^2$, получаем доказательство свойства 4).

⊠

Доказательство Теоремы I.4.

Дисперсии остатков равны соответствующим остаточным дисперсиям

$$\mathbf{D} \varepsilon_{1*(3\dots p)} = \frac{|\Lambda_{22}|}{|(\Lambda_{22})_{11}|}, \quad \mathbf{D} \varepsilon_{2*(3\dots p)} = \frac{|\Lambda_{11}|}{|(\Lambda_{11})_{11}|},$$

причем, легко видеть, что матрица $(\Lambda_{22})_{11} = (\Lambda_{11})_{11}$.

Ковариация между остатками равна

$$\begin{aligned} \mathbf{Cov}(\varepsilon_{1*(3\dots p)}, \varepsilon_{2*(3\dots p)}) &= \mathbf{E}(X_1 - X_{1*(3\dots p)})\varepsilon_{2*(3\dots p)} = \\ &= \mathbf{E}[(X_1 - \mu_1) - (X_{1*(3\dots p)} - \mu_1)]\varepsilon_{2*(3\dots p)} = \\ &= \mathbf{E}(X_1 - \mu_1)\varepsilon_{2*(3\dots p)}, \end{aligned}$$

так как $\varepsilon_{2*(3\dots p)}$ не коррелирует с величинами X_3, \dots, X_p , входящими линейно в регрессию $X_{1*(3\dots p)}$. Далее, запишем остаток $\varepsilon_{2*(3\dots p)}$ в виде (I.10), стр. 30, заменив матрицу Λ на матрицу Λ_{11} :

$$\varepsilon_{2*(3\dots p)} = \sum_{j=1}^{p-1} (-1)^{j+1} \frac{|(\Lambda_{11})_{1j}|}{|(\Lambda_{11})_{11}|} (X_{j+1} - \mu_{j+1}).$$

Следовательно,

$$\mathbf{Cov}(\varepsilon_{1*(3\dots p)}, \varepsilon_{2*(3\dots p)}) = \sum_{j=1}^{p-1} (-1)^{j+1} \frac{|(\Lambda_{11})_{1j}|}{|(\Lambda_{11})_{11}|} \lambda_{1(j+1)}. \quad (\text{I.11})$$

Очевидно, при замене порядка вычеркивания строк матрица $(\Lambda_{11})_{1j} = (\Lambda_{21})_{1j}$. Отсюда видно, что сумма в формуле (I.11)

представляет собой (с точностью до знаменателя $|\!(\Lambda_{11})_{11}|\!$) разложение определителя матрицы $(\Lambda_{21})_{1j}$ по первой строке. Таким образом,

$$\mathbf{Cov}(\varepsilon_{1*(3\dots p)}, \varepsilon_{2*(3\dots p)}) = \frac{|\!(\Lambda_{21})_{1j}|\!}{|\!(\Lambda_{11})_{11}|\!}.$$

Подставляя теперь полученные значения в формулу для коэффициента корреляции, получаем доказательство теоремы:

$$\mathbf{Corr}(\varepsilon_{1*(3\dots p)}, \varepsilon_{2*(3\dots p)}) = \frac{\mathbf{Cov}(\varepsilon_{1*(3\dots p)}, \varepsilon_{2*(3\dots p)})}{\sqrt{\mathbf{D}(\varepsilon_{1*(3\dots p)}) \mathbf{D}(\varepsilon_{2*(3\dots p)})}}.$$

⊠

Доказательство Теоремы 1.5.

Рассмотрим произвольные линейные комбинации $U = \vec{a}' \vec{X}_1$ и $V = \vec{b}' \vec{X}_2$ компонент векторов \vec{X}_1 и \vec{X}_2 . Поскольку коэффициент корреляции не изменяется при умножении случайных величин на положительные константы, то однозначное решение задачи отыскания линейных комбинаций с максимальным коэффициентом корреляции возможно лишь при некоторых условиях нормированности. Будем рассматривать линейные комбинации с единичной дисперсией:

$$\begin{aligned} \mathbf{D}U &= \vec{a}' \Sigma_{11} \vec{a} = 1, \\ \mathbf{D}V &= \vec{b}' \Sigma_{22} \vec{b} = 1. \end{aligned} \tag{I.12}$$

Коэффициент корреляции между U и V равен

$$\mathbf{Corr}(U, V) = \mathbf{E}UV = \mathbf{E} \vec{a}' \vec{X}_1 \vec{X}_2' \vec{b} = \vec{a}' \Sigma_{12} \vec{b}. \tag{I.13}$$

Введем функцию Лагранжа

$$H = \vec{a}' \Sigma_{12} \vec{b} - \frac{1}{2}t \vec{a}' \Sigma_{11} \vec{a} - \frac{1}{2}s \vec{b}' \Sigma_{22} \vec{b},$$

где t и s – множители Лагранжа. Приравняв к нулю соответствующие производные H , получаем следующую систему уравнений

$$-t\Sigma_{11}\vec{a} + \Sigma_{12}\vec{b} = \vec{0}, \quad (\text{I.14})$$

$$\Sigma'_{12}\vec{a} - s\Sigma_{22}\vec{b} = \vec{0}. \quad (\text{I.15})$$

Умножив выражение (I.14) слева на \vec{a}' и (I.15) слева на \vec{b}' и учитывая условия (I.12), получим

$$t = \vec{a}'\Sigma_{12}\vec{b} = \mathbf{Corr}(U, V), \quad (\text{I.16})$$

$$s = \vec{b}'\Sigma'_{12}\vec{a} = \mathbf{Corr}(U, V). \quad (\text{I.17})$$

Таким образом, $t = s = \vec{a}'\Sigma_{12}\vec{b}$, и соотношения (I.14) и (I.15) можно переписать в виде

$$\begin{pmatrix} -t\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -t\Sigma_{22} \end{pmatrix} \begin{pmatrix} \vec{a} \\ \vec{b} \end{pmatrix} = \vec{0}. \quad (\text{I.18})$$

Для существования ненулевых решений необходимо, чтобы матрица слева была вырожденной, т.е.

$$\begin{vmatrix} -t\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -t\Sigma_{22} \end{vmatrix} = 0. \quad (\text{I.19})$$

Легко установить, что наименьшая степень t в этом уравнении равна $(p_2 - p_1)$. Поэтому оно имеет не более $2p_1$ ненулевых корней: $t_1 \geq t_2 \geq \dots \geq t_{p_1} > 0$.

В силу соотношения (I.16) и ввиду задачи определения максимального коэффициента корреляции, следует положить $t = t_1$. Пусть \vec{a}_1 и \vec{b}_1 – нормированные решения системы (I.18). Тогда случайные величины $U_1 = \vec{a}'_1 \vec{X}_1$ и $V_1 = \vec{b}'_1 \vec{X}_2$ суть нормированные линейные комбинации с максимальным коэффициентом корреляции.

Предположим, что мы определили уже k пар канонических корреляций, причем значения канонических корреляций являются корнями уравнения (I.19). Будем искать такие линейные комбинации $U = \vec{a}' \vec{X}_1$ и $V = \vec{b}' \vec{X}_2$, которые имеют максимальный коэффициент корреляции, нормированы (I.12) и не коррелируют с $U_1, V_1, \dots, U_k, V_k$. Условие некоррелированности U и U_i имеет вид

$$\mathbf{Cov}(U, U_i) = \mathbf{E} U U_i = \mathbf{E} \vec{a}' \vec{X}_1 \vec{X}_1' \vec{a}_i = \vec{a}' \Sigma_{11} \vec{a}_i = 0. \quad (\text{I.20})$$

Покажем, что в этом случае U не будет коррелировать и с V_i . Действительно, в силу равенства (I.14) $\Sigma_{12} \vec{b}_i = t_i \Sigma_{11} \vec{a}_i$ и ковариация

$$\mathbf{E} U V_i = \vec{a}' \Sigma_{12} \vec{b}_i = t_i \vec{a}' \Sigma_{11} \vec{a}_i = 0.$$

Аналогично, условие некоррелированности V и V_i –

$$\mathbf{E} V V_i = \vec{b}' \Sigma_{22} \vec{b}_i = 0,$$

обеспечивает некоррелированность V и U_i .

Рассмотрим функцию

$$H = \vec{a}' \Sigma_{12} \vec{b} - \frac{1}{2} t \vec{a}' \Sigma_{11} \vec{a} - \frac{1}{2} s \vec{b}' \Sigma_{22} \vec{b} + \sum_{i=1}^k (g_i \vec{a}' \Sigma_{11} \vec{a}_i + w_i \vec{b}' \Sigma_{22} \vec{b}_i),$$

где t, s, \vec{g}, \vec{w} – множители Лагранжа. Приравняв к нулю производные H по компонентам \vec{a} и \vec{b} , получаем

$$H'_a = \Sigma_{12} \vec{b} - t \Sigma_{11} \vec{a} + \sum_i g_i \Sigma_{11} \vec{a}_i = \vec{0},$$

$$H'_b = \Sigma_{21} \vec{a} - s \Sigma_{22} \vec{b} + \sum_i w_i \Sigma_{22} \vec{b}_i = \vec{0}.$$

Умножая эти соотношения слева на \vec{a}'_j и \vec{b}'_j соответственно и учитывая условия некоррелированности, получим, что

коэффициенты g_j и w_j равны нулю:

$$0 = g_j \vec{a}'_j \Sigma_{11} \vec{a}_j = g_j, \quad 0 = w_j \vec{b}'_j \Sigma_{22} \vec{b}_j = w_j.$$

Таким образом, задача поиска экстремума функции H снова приводит нас к уравнениям (I.18) и (I.19). Можно показать, что невырожденность матрицы Σ_{11} обеспечивает существование ровно p_1 различных решений этих уравнений. Эти решения и есть пары канонические переменных.

Рассмотрим матрицы $\mathbf{A} = (\vec{a}_1, \dots, \vec{a}_{p_1})$ и $\mathbf{B}_1 = (\vec{b}_1, \dots, \vec{b}_{p_1})$. Достроим матрицу \mathbf{B}_1 справа до матрицы $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2)$ размера $p_2 \cdot p_2$ матрицей \mathbf{B}_2 , удовлетворяющей следующим уравнениям:

$$\mathbf{B}'_2 \Sigma_{22} \mathbf{B}_1 = 0, \quad \mathbf{B}'_2 \Sigma_{22} \mathbf{B}_1 = \mathbf{I}.$$

Эту матрицу можно построить последовательно, столбец за столбцом; вектор \vec{b}_{p_1+1} ортогонален вектору $\Sigma_{22} \mathbf{B}_1$ и нормирован так, что $\vec{b}'_{p_1+1} \Sigma_{22} \vec{b}_{p_1+1} = 1$, вектор \vec{b}_{p_1+2} ортогонален вектору $\Sigma_{22} \mathbf{B}_1 \vec{b}_{p_1+1}$ и нормирован так, что $\vec{b}'_{p_1+2} \Sigma_{22} \vec{b}_{p_1+2} = 1$, и так далее.

Поскольку, матрицы \mathbf{A} и \mathbf{B} невырождены ($\mathbf{A}' \Sigma_{11} \mathbf{A} = \mathbf{I}$, $\mathbf{B}' \Sigma_{22} \mathbf{B} = \mathbf{I}$), то любые линейные комбинации векторов \vec{X}_1 и \vec{X}_2 могут быть представлены в виде линейных комбинаций векторов \vec{U} и \vec{V} соответственно. Коэффициент корреляции между нормированными линейными комбинациями $\vec{a}' \vec{U}$ и $\vec{b}' \vec{V}$ равен

$$\text{Corr}(\vec{a}' \vec{U}, \vec{b}' \vec{V}) = \vec{a}' \text{Corr}(\vec{U}, \vec{V}') \vec{b} = \vec{a}' (T, 0) \vec{b} = \sum_{i=1}^{p_1} t_i a_i b_i,$$

где матрица T – диагональная матрица из элементов t_1, \dots, t_{p_1} . Положим $c_i = t_i a_i / \sqrt{\sum (t_i a_i)^2}$. Тогда корреляция

$$\text{Corr}(\vec{a}' \vec{U}, \vec{b}' \vec{V}) = \sqrt{\sum (t_i a_i)^2} \sum_{i=1}^{p_1} c_i b_i$$

и её максимальное значение по \vec{b} достигается при $b_i = c_i$ и $\sum c_i b_i = 1$, так как последняя сумма представляет собой косинус угла между векторами \vec{b} и $(c_1, \dots, c_{p_1}, 0, \dots, 0)$.

С другой стороны, вектор \vec{a} нормирован: $\sum a_i^2 = 1$, что можно записать как $a_1^2 = 1 - \sum_{i=2}^{p_1} a_i^2$. Поэтому выражение

$$\sqrt{\sum (t_i a_i)^2} = \sqrt{t_1^2 + \sum_{i=2}^{p_1} (t_i^2 - t_1^2) a_i^2}$$

с $t_i^2 \leq t_1^2$ будет достигать своего максимума при $a_i = 0, i = 2, \dots, p_1$. В этом случае все $b_i = 0, i \geq 2$, и максимальный коэффициент корреляции достигается при выборе U_1 и V_1 в качестве линейных комбинаций. При построении следующей пары канонических величин мы должны рассмотреть линейные комбинации $\vec{a}' \vec{U}$ и $\vec{b}' \vec{V}$, не коррелирующие с U_1 и V_1 , что приводит к значениям $a_1 = 0, b_1 = 0$:

$$0 = \mathbf{Corr}(U_1, \vec{a}' \vec{U}) = \sum a_i \mathbf{Corr}(U_1, U_i) = a_1,$$

$$0 = \mathbf{Corr}(V_1, \vec{b}' \vec{V}) = \sum b_i \mathbf{Corr}(V_1, V_i) = b_1.$$

Дальнейшее доказательство проводится по аналогии с предыдущим.

⊠

Доказательство Теоремы I.6.

Умножив уравнение (I.14) на t , а уравнение (I.15) (с $s = t$) слева на Σ_{22}^{-1} , получим, что

$$t \Sigma_{12} \vec{b} = t^2 \Sigma_{11} \vec{b}, \quad \Sigma_{22}^{-1} \Sigma_{21} \vec{a} = t \vec{b}.$$

Подстановка $t \vec{b}$ из второго уравнения в первое приводит к соотношению

$$\Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \vec{a} = t^2 \Sigma_{11} \vec{a},$$

которое совпадает с характеристическим уравнением для матрицы, заявленной в формулировке теоремы.

⊠

Доказательство Теоремы I.7.

Рассмотрим эллипс $\Omega = \{\vec{y} : (\vec{y} - \vec{m})' A (\vec{y} - \vec{m}) \leq c^2\}$, и пусть V_Ω – его объем. Тогда плотность равномерного распределения внутри Ω равна $f(\vec{y}) = 1/V_\Omega$, если $\vec{y} \in \Omega$. Поэтому среднее

$$\mathbf{E} \vec{Y} = \frac{1}{V_\Omega} \int_{\Omega} \vec{y} d\vec{y} = \frac{1}{V_\Omega} \int_{\Omega} (\vec{y} - \vec{m}) d\vec{y} + \vec{m} \frac{1}{V_\Omega} \int_{\Omega} d\vec{y} = \vec{m}$$

в силу симметричности эллипса относительно своего центра. Таким образом, центр эллипсоида рассеяния $\vec{m} = \vec{\mu}$.

Для вычисления $\mathbf{Cov}(\vec{Y})$ воспользуемся соотношением (B.3), стр. 96. После очевидной замены получаем

$$\mathbf{Cov}(\vec{Y}) = \frac{1}{V_\Omega} \int_{\Omega} (\vec{y} - \vec{\mu})(\vec{y} - \vec{\mu})' d\vec{y} = \frac{c^2}{p+2} A^{-1}.$$

Заметим, что выражение, задающее эллипсоид, содержит лишнюю константу c^2 , которой можно присвоить произвольное значение, разделив на соответствующую величину обе части равенства. Очевидно, для наших целей удобно положить $c^2 = p + 2$. В этом случае матрица, определяющая эллипсоид рассеяния, равна $A = \Lambda^{-1}$.

⊠

Доказательство Теоремы I.8.

Угол наклона касательных, параллельных оси OX , равен нулю, поэтому производная функции $y = y(x)$, описывающей эллипс, в точках касания равна 0. Для отыскания этой производной продифференцируем уравнение эллипса (I.6), стр. 23,

по x , при условии, что y есть функция x . Имеем

$$2\frac{y - \mu_1}{\sigma_1^2}y' - 2\rho\frac{x - \mu_2}{\sigma_1\sigma_2}y' - 2\rho\frac{y - \mu_1}{\sigma_1\sigma_2} + 2\frac{x - \mu_2}{\sigma_2^2} = 0.$$

Как уже отмечалось, в точках касания прямых, параллельных оси OX , производная $y' = 0$. Подстановка этого значения в предыдущее уравнение прямым приводит нас к уравнению регрессии X на Y .

⊠

Доказательство Теоремы I.9.

Представим квадрат ошибки в следующем виде:

$$\begin{aligned} (X_1 - h(X_2, \dots, X_p))^2 &= (X_1 - \mu_{1*} + \mu_{1*} - h)^2 = \\ &= (X_1 - \mu_{1*})^2 + 2(X_1 - \mu_{1*})(\mu_{1*} - h) + (\mu_{1*} - h)^2 \end{aligned}$$

(при очевидных сокращениях записи). По свойству условного математического ожидания, функцию, зависящую только от величин, входящих в условие, можно вынести за знак условного ожидания. Поэтому

$$\begin{aligned} \mathbf{E}(X_1 - \mu_{1*})(\mu_{1*} - h) &= \\ &= \mathbf{E}[\mathbf{E}\{(X_1 - \mu_{1*})(\mu_{1*} - h) \mid X_2, \dots, X_p\}] = \\ &= \mathbf{E}[(X_1 - \mu_{1*}) \mathbf{E}\{(\mu_{1*} - h) \mid X_2, \dots, X_p\}] = 0. \end{aligned}$$

Таким образом, среднеквадратическая ошибка равна

$$\mathbf{E}(X_1 - h(X_2, \dots, X_p))^2 = \mathbf{E}(X_1 - \mu_{1*})^2 + \mathbf{E}(\mu_{1*} - h)^2.$$

Очевидно, ее минимум достигается при $h = \mu_{1*}$.

⊠

Доказательство Теоремы I.10.

Очевидно, среднее значение $\mu_1 = \mathbf{E} X_1 = \mathbf{E} \mu_{1*}$. Поэтому

$$\begin{aligned}
\mathbf{E}(X_1 - \mu_{1*})^2 &= \mathbf{E}[(X_1 - \mu_1) - (\mu_{1*} - \mu_1)]^2 = \mathbf{E}(X_1 - \mu_1)^2 - \\
&\quad - 2\mathbf{E}(X_1 - \mu_1)(\mu_{1*} - \mu_1) + \mathbf{E}(\mu_{1*} - \mu_1)^2 = \\
&= \mathbf{D} X_1 - \mathbf{D} \mu_{1*}, \tag{I.21}
\end{aligned}$$

поскольку среднее значение

$$\begin{aligned}
\mathbf{E}(X_1 - \mu_1)(\mu_{1*} - \mu_1) &= \\
&= \mathbf{E}[(\mu_{1*} - \mu_1) \mathbf{E}\{(X_1 - \mu_1) \mid X_2, \dots, X_p\}] = \\
&= \mathbf{E}(\mu_{1*} - \mu_1)^2 = \mathbf{D} \mu_{1*}.
\end{aligned}$$

Подстановка полученных соотношений в определение корреляционного отношения доказывает теорему.

⊠

Доказательство Свойств корреляционного отношения.

Пусть H_L – функция линейной среднеквадратической регрессии X_1 на вектор (X_2, \dots, X_p) , тогда, поскольку при определении линейной регрессии минимум берется по более узкому множеству функций, то

$$\mathbf{E}(X_1 - \mu_{1*})^2 \leq \mathbf{E}(X_1 - H_L)^2 = \mathbf{D}(X_1)(1 - \rho_{1*}^2),$$

причем знак равенства в неравенстве достигается только при линейности среднеквадратической регрессии.

Воспользовавшись равенством (I.21), получаем

$$\mathbf{E}(X_1 - \mu_{1*})^2 = \mathbf{D} X_1 - \mathbf{D} \mu_{1*} \geq 0$$

со знаком равенства только при $X_1 \equiv \mu_{1*}$. Таким образом, имеем

$$0 \leq 1 - \eta_{1*}^2 \leq 1 - \rho_{1*}^2$$

при соответствующих условиях достижения знака равенства.

⊠

Глава II. МЕТОД ГЛАВНЫХ КОМПОНЕНТ

Если хочешь быть красивым,
поступи в гусары.

Введение

Во многих исследовательских работах число показателей, которые требуется обработать, слишком велико. В таких изысканиях интерес зачастую представляют не значения самих показателей, а их разброс между изучаемыми объектами. Так, например, антрополог-физиолог проводит измерения нескольких десятков характеристик, таких как длина уха, ширина уха, длина лица, ширина лица и т.п., у большого числа индивидов. Его может интересовать описание и анализ различий индивидов по подобному рода характеристикам (скажем, для дальнейшего изучения причин этих различий). Поэтому, естественно, вначале он должен будет отбросить характеристики или комбинации характеристик с малыми различиями а затем выделить характеристики или комбинации характеристик, обобщающие разделительные свойства индивидов. Для этих целей и служит метод главных компонент.

§ 1. Определение и способ построения главных компонент

О п р е д е л е н и е 1. Пусть \vec{X} — p -мерный случайный вектор с вектором средних $\vec{\mu}$. Вектор $\vec{\zeta} = (\zeta_1, \dots, \zeta_p)'$ называется вектором главных компонент для \vec{X} , если

(Гк1) все компоненты $\vec{\zeta}$ получены как нормированные линейные комбинации центрированного вектора \vec{X} :

$$\zeta_j = \vec{c}_j'(\vec{X} - \vec{\mu}) = c_{j1}(X_1 - \mu_1) + \dots + c_{jp}(X_p - \mu_p),$$

$$\vec{c}_j' \vec{c}_j = c_{j1}^2 + \dots + c_{jp}^2 = 1, \quad j = 1, \dots, p;$$

(Гк2) дисперсия 1-ой компоненты ζ_1 максимальна среди всех нормированных линейных комбинаций \vec{X} ;

(Гк3) дисперсия j -ой компоненты ζ_j максимальна среди всех нормированных линейных комбинаций вектора \vec{X} , не коррелирующих с компонентами $\zeta_1, \dots, \zeta_{j-1}$, $j = 1, \dots, p$.

Решение проблемы главных компонент дает следующая теорема.

Теорема II.1. Пусть Λ – матрица ковариаций \vec{X} . Тогда вектор-столбцы \vec{c}_j , определяющие главные компоненты, совпадают с нормированными собственными векторами матрицы Λ :

$$\Lambda \vec{c}_j = \gamma_j \vec{c}_j, \quad j = 1, \dots, p,$$

причем собственные числа выбраны так, что $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_p$, а все собственные векторы ортогональны.

Нормирование измеряемых характеристик

Метод главных компонент весьма чувствителен к изменению единиц измерения характеристик исследуемого объекта. Поэтому применение этого метода будет наиболее плодотворным в ситуациях, когда показатели X_1, \dots, X_p имеют общую физическую природу и измеряются в одних и тех же единицах. В противном случае обычно прибегают к предварительной нормировке показателей и в качестве нормирующей константы выбирают стандартное отклонение

$$X_j^* = \frac{X_j - \mu_j}{\sigma_j}.$$

Легко понять, что в этом случае ковариационная матрица \vec{X}^* совпадает с корреляционной матрицей \vec{X} . Таким образом, метод главных компонент, примененный к нормированным показателям, отличается от этого же метода, но примененного к ненормированным показателям, тем, что в этом случае решается характеристическое уравнение относительно матрицы корреляций, а не ковариаций. Заметим, кстати, что тогда

сумма дисперсий главных компонент равна p , так как на диагонали матрицы корреляций стоят единицы.

§ 2. Свойства главных компонент

Пусть $\vec{\zeta} = C' \vec{X}$ — вектор главных компонент \vec{X} .

1) Матрица C ортогональна:

$$C'C = CC' = \mathbf{I}.$$

2) Матрица ковариаций $\vec{\zeta}$ диагональна:

$$\mathbf{Cov}(\vec{\zeta}) = \text{diag}(\gamma_1, \dots, \gamma_p).$$

3) Дисперсия j -ой главной компоненты равна j -ому собственному числу матрицы Λ :

$$\mathbf{D} \zeta_j = \gamma_j, \quad j = 1, \dots, p.$$

4) Переход к главным компонентам сохраняет обобщенную дисперсию

$$\mathfrak{S}^2(\vec{X}) = |\Lambda| = \mathfrak{S}^2(\vec{\zeta}) = \prod_{j=1}^p \gamma_j.$$

5) Переход к главным компонентам сохраняет сумму дисперсий

$$\sum_{j=1}^p \mathbf{D} X_j = \sum_{j=1}^p \mathbf{D} \zeta_j = \sum_{j=1}^p \gamma_j.$$

6) Переход к главным компонентам представляет собой поворот осей системы координат параллельно осям эллипсоида рассеяния \vec{X} .

§ 3. Понижение размерности наблюдаемого вектора

Информативность первых главных компонент

Свойство 5) лежит в основе применения метода главных компонент на практике. Рассмотрим отношение

$$m_q = \frac{\gamma_1 + \dots + \gamma_q}{\gamma_1 + \dots + \gamma_p}.$$

Величина $m_q \cdot 100\%$ показывает ту долю разброса наблюдений, которая может быть объяснена факторами, определяющими первые q главных компонент. Если $m_q \cdot 100\%$ достаточно велико, то оставшиеся $p - q$ главных компонент могут быть исключены из дальнейшего рассмотрения. Точное значение понятия „достаточно велико“ конкретизируется, исходя непосредственно из рассматриваемой исследовательской задачи. Можно предложить следующую градацию для описания информативности главных компонент.

Если $m_q \cdot 100\% > 75\%$, тогда говорят, что первые q главных компонент содержат *основную массу* информации о разбросе наблюдаемых признаков, если $m_q \cdot 100\% > 90\%$, то *почти всю* информацию, и наконец, если $m_q \cdot 100\% > 95\%$, то *всю* информацию.

Относительная величина потери информации

Обратимся теперь к вопросу о потере информации при переходе к части главных компонент. С этой целью исследуем возможность восстановления исходных значений компонент вектора \vec{X} по значениям первых q главных компонент.

Теорема II.2. 1) Наилучшая линейная аппроксимация вектора \vec{X} по значениям первых q главных компонент $\vec{\zeta}^{(q)}$ достигается при линейном преобразовании $\vec{X}^* = C^{(q)} \vec{\zeta}^{(q)}$ с матрицей $C^{(q)}$ размера $p \cdot q$, построенной на первых q столбцах матрицы C .

2) Относительная (к сумме дисперсий всех показателей \vec{X}) суммарная ошибка этого прогноза равна

$$1 - m_q.$$

§ 4. Интерпретация главных компонент

Другой аспект применения главных компонент связан с выделением некоторых обобщающих показателей для измеряемых характеристик. Для этого можно, например, проанализировать коэффициенты векторов \vec{c}_j , связанных с первыми главными компонентами. Если среди этих коэффициентов найдутся несколько (1-3) существенно больших по абсолютной величине остальных коэффициентов, то можно, объединив соответствующие им показатели в одну группу и придумав этой группе некое звучное название, считать её (эту группу) ответственной за раброс наблюдаемых характеристик. Следующий пример иллюстрирует эту схему.

Пример исследования линейных размеров цветка

Измерялись 4 характеристики цветка *Iris Versicolor*: X_1 – длина чашелистика, X_2 – ширина чашелистика, X_3 – длина

лепестка, X_4 – ширина лепестка. Была получена следующая матрица ковариаций

$$\begin{pmatrix} 0.266 & 0.085 & 0.183 & 0.056 \\ 0.085 & 0.098 & 0.083 & 0.041 \\ 0.183 & 0.083 & 0.221 & 0.073 \\ 0.056 & 0.041 & 0.073 & 0.039 \end{pmatrix}.$$

Собственные числа и матрица собственных векторов равны

$$(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (0.49, 0.09, 0.05, 0.01),$$

$$C = \begin{pmatrix} 0.687 & -0.669 & -0.265 & 0.102 \\ 0.305 & 0.568 & -0.730 & -0.229 \\ 0.627 & 0.343 & -0.627 & -0.316 \\ 0.215 & 0.335 & 0.064 & 0.915 \end{pmatrix}$$

Таким образом, первые две главные компоненты несут почти всю информацию о разбросе данных ($> 90\%$), причем 1-ая компонента является обобщающей характеристикой для измерений длины элементов растений, а 2-ая компонента для измерений чашелистика. С другой стороны, видно, что 4-ая главная компонента (почти совпадающая с шириной лепестка - X_4) может быть отброшена без существенного ущерба для полноты картины и не участвовать в дальнейших изысканиях, связанных с *Iris Versicolor*. Здесь, конечно, трудно специалисту в ботанике дать разумное объяснение 3-ей главной компоненте, однако именно это затруднение может стать началом нового этапа исследований.

Доказательства

Доказательство Теоремы II.1.

Так как нас интересуют только дисперсии линейных комбинаций $\vec{X} - \vec{\mu}$, то можно ограничиться случаем $\vec{\mu} = \vec{0}$. Начнем с определения ζ_1 . Пусть $Y = \vec{c}' \vec{X} = c_1 X_1 + \dots + c_p X_p$ — произвольная линейная комбинация \vec{X} . Дисперсия Y равна

$$\mathbf{D}Y = \mathbf{Cov}(Y) = \mathbf{Cov}(\vec{c}' \vec{X}) = \vec{c}' \mathbf{Cov}(\vec{X}) \vec{c} = \vec{c}' \Lambda \vec{c}.$$

Необходимо найти максимум $\vec{c}' \Lambda \vec{c}$ по векторам \vec{c} , удовлетворяющим условию $c_1^2 + \dots + c_p^2 = 1$. Рассмотрим функцию

$$\begin{aligned} H(c_1, \dots, c_p) &= \vec{c}' \Lambda \vec{c} - \gamma \vec{c}' \vec{c} = \sum_{j,m=1}^p \lambda_{jm} c_j c_m - \gamma (c_1^2 + \dots + c_p^2) = \\ &= \lambda_{11} c_1^2 + \sum_{j=2}^p \lambda_{j1} c_j c_1 + \sum_{m=2}^p \lambda_{1m} c_m c_1 + \sum_{j,m \neq 1} \lambda_{jm} c_j c_m - \gamma (c_1^2 + \dots + c_p^2), \end{aligned}$$

где γ — неопределенный коэффициент Лагранжа. Производная функции H по c_1 равна

$$\frac{\partial H}{\partial c_1} = 2c_1 \lambda_{11} + \sum_{j=2}^p \lambda_{j1} c_j + \sum_{m=2}^p \lambda_{1m} c_m - 2\gamma c_1 = 2 \sum_{j=1}^p \lambda_{1j} c_j - 2\gamma c_1.$$

Аналогично, для всех остальных $k = 2, \dots, p$

$$\frac{\partial H}{\partial c_k} = 2 \sum_{j=1}^p \lambda_{kj} c_j - 2\gamma c_k.$$

Следовательно, система уравнений для нахождения экстремума функции H в матричной форме имеет вид:

$$\Lambda \vec{c} = \gamma \vec{c}.$$

Это уравнение есть не что иное, как характеристическое уравнение для определения собственных значений и собственных

векторов матрицы Λ . Известно, что для любой невырожденной положительно определенной p -мерной матрицы существует ровно p собственных значений (для простоты предположим, что они все различны) и p соответствующих им собственных векторов, которые будут ортогональны между собой. Обозначим собственные числа Λ через $\gamma_1, \dots, \gamma_p$, причем выберем их так, чтобы $\gamma_1 > \gamma_2 > \dots > \gamma_p$, и пусть $\vec{c}_1, \dots, \vec{c}_p$ – соответствующие им нормированные собственные векторы.

Заметим, что если \vec{c}_j – нормированный собственный вектор Λ с собственным значением γ_j , то дисперсия

$$\mathbf{D} \vec{c}'_j \vec{X} = \vec{c}'_j \Lambda \vec{c}_j = \gamma_j \vec{c}'_j \vec{c}_j = \gamma_j.$$

Таким образом, если максимум дисперсии $\vec{c}' \vec{X}$ существует, то он достигается при $\vec{c} = \vec{c}_1$. Покажем, что второй дифференциал функции H отрицателен в точке $\vec{c} = \vec{c}_1$. Легко видеть, что матрица вторых производных H равна $\Lambda - \gamma \mathbf{I}$. Поэтому второй дифференциал H равен

$$d^2 H = d\vec{c}' (\Lambda - \gamma \mathbf{I}) d\vec{c} = d\vec{c}' \Lambda d\vec{c} - \gamma d\vec{c}' d\vec{c},$$

где $d\vec{c} = (dc_1, \dots, dc_p)$. Из условия $c_1^2 + \dots + c_p^2 = 1$ следует, что дифференциал $d(c_1^2 + \dots + c_p^2) = 0$, то есть $2c_1 dc_1 + \dots + 2c_p dc_p = 0$ или

$$\vec{c}' d\vec{c} = 0.$$

Если положить $\vec{c} = \vec{c}_1$, то из этого равенства следует, что вектор $d\vec{c}$ ортогонален \vec{c}_1 . Так как все собственные векторы $\vec{c}_1, \dots, \vec{c}_p$ ортогональны и нормированы, то они могут быть выбраны в качестве новой системы координат. Следовательно, вектор $d\vec{c}$, ортогональный \vec{c}_1 , можно представить в виде линейной комбинации векторов $\vec{c}_2, \dots, \vec{c}_p$: $d\vec{c} =$

$b_2 \vec{c}_2 + \dots + b_p \vec{c}_p$. Таким образом, если только не все $b_j = 0$ (что выполняется для $d\vec{c} \neq \vec{0}$), то при $\vec{c} = \vec{c}_1$ и $\gamma = \gamma_1$

$$\begin{aligned}
d^2H &= \sum_{j=2}^p b_j \vec{c}'_j \Lambda \sum_{m=2}^p b_m \vec{c}_m - \gamma_1 \sum_{j=2}^p b_j \vec{c}'_j \sum_{m=2}^p b_m \vec{c}_m = \\
&= \sum_{j=2}^p b_j \vec{c}'_j \sum_{m=2}^p b_m \gamma_m \vec{c}_m - \gamma_1 \sum_{j=2}^p b_j \vec{c}'_j \sum_{m=2}^p b_m \vec{c}_m = \\
&\quad (\text{т.к. } \vec{c}'_j \vec{c}_m = 0) \\
&= \sum_{j=2}^p \gamma_j b_j^2 \|\vec{c}_j\|^2 - \gamma_1 \sum_{j=2}^p b_j^2 \|\vec{c}_j\|^2 = \\
&= \sum_{j=2}^p (\gamma_j - \gamma_1) b_j^2 < 0,
\end{aligned}$$

так как все $\gamma_j < \gamma_1$, $j = \overline{2, p}$, что доказывает утверждение теоремы в части определения ζ_1 .

Перейдем теперь к определению ζ_2 . Пусть $\zeta_2 = \vec{e}' \vec{X}$, тогда

$$\text{Cov}(\zeta_1, \zeta_2) = \mathbf{E} \vec{e}' \vec{X} \vec{c}'_1 \vec{X} = \mathbf{E} \vec{e}' \vec{X} \vec{X}' \vec{c}_1 = \vec{e}' \Lambda \vec{c}_1 = \gamma_1 \vec{e}' \vec{c}_1$$

и условие некоррелированности ζ_1, ζ_2 выполняется только в том случае, если векторы \vec{e} и \vec{c}_1 ортогональны и $\vec{e}' \vec{c}_1 = 0$.

Рассмотрим функцию

$$Q = \vec{e}' \Lambda \vec{e} - \gamma(e_1^1 + \dots + e_p^2) - \alpha(e_1 c_1 + \dots + e_p c_p),$$

где γ, α – множители Лагранжа. Так же, как и при рассмотрении функции H , легко показать, что уравнение, которому удовлетворяют экстремальные точки функции Q , имеет вид

$$\Lambda \vec{e} = \gamma \vec{e} + \alpha \vec{c}_1.$$

Если это уравнение умножить слева на \vec{c}'_1 , то с учетом равенств

$$\vec{c}'_1 \Lambda = \gamma_1 \vec{c}'_1, \quad \vec{c}'_1 \vec{e} = 0 \quad \text{и} \quad \vec{c}'_1 \vec{c}_1 = 1$$

получаем, что $\alpha = 0$. То есть снова для определения вектора \vec{e} , максимизирующего дисперсию $\vec{e}' \vec{X}$, мы приходим к характеристическому уравнению $\Lambda \vec{e} = \gamma \vec{e}$.

Второй дифференциал Q в этом случае равен

$$d^2Q = d\vec{e}' \Lambda d\vec{e} - \gamma d\vec{e}' d\vec{e}.$$

Необходимо определить знак d^2Q при условиях $e_1^2 + \dots + e_p^2 = 1$ и $e_1 c_{11} + \dots + e_p c_{1p} = 0$, которые приводят к следующим условиям на дифференциал $d\vec{e}$:

$$\begin{aligned} \vec{e}' d\vec{e} &= 0 \quad \text{или} \quad \vec{e} \perp d\vec{e}, \\ \vec{c}'_1 d\vec{e} &= 0 \quad \text{или} \quad \vec{c}_1 \perp d\vec{e}. \end{aligned}$$

Если выбрать $\vec{e} = \vec{c}_2$, то отсюда следует, что дифференциал $d\vec{e}$ представляется линейной комбинацией только векторов $\vec{c}_3, \dots, \vec{c}_p$,

$$d\vec{e} = \sum_{j=3}^p b_j \vec{c}_j.$$

$$\text{Поэтому} \quad d^2Q = \sum_{j=3}^p (\gamma_j - \gamma_2) b_j^2 < 0.$$

Дальнейшее доказательство легко продолжается по аналогии.

Доказательство теоремы несколько усложнится, если некоторые из собственных чисел будут совпадать. Мы опустим этот момент.

⊠

Доказательство Свойств главных компонент.

Свойства 1) – 3) были установлены при доказательстве теоремы II.1. Свойство 4) выводится из свойств 1) и 3) следующим образом:

$$\mathfrak{E}^2(\vec{\zeta}) = |\mathbf{Cov}(\vec{\zeta})| = |C' \Lambda C| = |C'| |C| |\Lambda| = |\Lambda|,$$

так как C – ортогональная матрица и $|C| = |C'| = \pm 1$.

Для доказательства 5-го свойства заметим, что так как средние значения $\mathbf{E} X_j = 0$, то

$$\sum_{j=1}^p \mathbf{D} X_j = \sum_{j=1}^p \mathbf{E} X_j^2 = \mathbf{E} \vec{X}' \vec{X}.$$

В силу свойства 1) $\vec{X} = C \vec{\zeta}$ и $C' C = \mathbf{I}$, следовательно,

$$\mathbf{E} \vec{X}' \vec{X} = \mathbf{E} \vec{\zeta}' C' C \vec{\zeta} = \mathbf{E} \vec{\zeta}' \vec{\zeta} = \sum_{j=1}^p \mathbf{E} \zeta_j^2 = \sum_{j=1}^p \mathbf{D} \zeta_j .$$

⊠

Доказательство Теоремы II.2.

Пусть $\vec{\zeta}^{(q)} = (\zeta_1, \dots, \zeta_q)'$ – q первых главных компонент вектора \vec{X} . Найдем линейное преобразование $\vec{b}' \vec{\zeta}^{(q)}$, наилучшим образом аппроксимирующее значения j -ой характеристики X_j вектора \vec{X} . Средне-квадратическая ошибка такой аппроксимации равна

$$\mathbf{E} (X_j - \sum_{l=1}^q b_l \zeta_l)^2.$$

В силу свойства 2) главных компонент производная этой функции по b_m равна

$$-2 \mathbf{E} \zeta_m (X_j - \sum_{l=1}^q b_l \zeta_l) = -2(\mathbf{E} \zeta_m X_j - b_m \gamma_m).$$

Так как $\vec{\zeta} = C' \vec{X}$, то $\vec{X} = C \vec{\zeta}$ и $X_j = \sum_{i=1}^p c_{ij} \zeta_i$. Следовательно, ковариация $\mathbf{E} \zeta_m X_j = c_{mj} \gamma_m$. Приравнивая к нулю первую производную, находим, что коэффициенты наилучшего прогноза X_j равны

$$b_m = c_{mj}, \quad m = \overline{1, q}.$$

Следовательно, чтобы получить наилучший прогноз вектора \vec{X} по первым q главным компонентам, мы должны в матрице C отбросить последние $p - q$ столбцов: $\vec{X} \approx C^{(q)} \vec{\zeta}^{(q)}$, где $C^{(q)} = (\vec{c}_1, \dots, \vec{c}_q)$.

Найдем значение суммарной среднеквадратической ошибки такой аппроксимации. Она равна

$$\begin{aligned} \sum_{j=1}^p \mathbf{E} (X_j - \sum_{l=1}^g c_{lj} \zeta_l)^2 &= \sum_{j=1}^p \mathbf{E} (\sum_{l=1}^p c_{lj} \zeta_l - \sum_{l=1}^g c_{lj} \zeta_l)^2 = \\ &= \sum_{p=1}^p \mathbf{E} (\sum_{l=q+1}^p c_{lj} \zeta_l)^2 = (\text{в силу свойств 2) и 3)}) \\ &= \sum_{j=1}^p \sum_{l=q+1}^p c_{lj}^2 \gamma_l = \sum_{l=q+1}^p (\sum_{j=1}^p c_{lj}^2) \gamma_l = \sum_{l=q+1}^p \gamma_l. \end{aligned}$$

⊠

Глава III. ФАКТОРНЫЙ АНАЛИЗ

Не в совокупности ищи единства,
но более – в единообразии
разделения.

Введение

Суть рассматриваемых в данной главе методов факторного анализа опирается на идею представления p -мерного вектора наблюдаемых характеристик в виде линейной комбинации меньшего числа факторов так, чтобы структура связей исходного вектора осталась неизменной. Например, в эксперименте по изучению успеваемости учащихся по школьным предметам можно попытаться представить результаты тестирования по каждому предмету с помощью всего трех составляющих: общей одаренности, одаренности в области гуманитарных наук и в области технических наук. В некотором смысле такое представление помогает выявить причинно-следственную связь, если интерпретировать ненаблюдаемые факторы как причины, а наблюдаемые признаки как следствия.

§ 1. Модель факторного анализа

Рассмотрим p -мерный случайный вектор \vec{X} с невырожденным распределением (то есть с матрицей ковариаций

$\Lambda = \mathbf{Cov}(\vec{X}) > 0$) и нулевым вектором средних $\vec{\mu} = 0$. Кроме того, как и в методе главных компонент, здесь нужно предположить, что все наблюдаемые признаки измеряются в одной шкале. Последние два требования могут быть достигнуты простым центрированием и нормированием каждого признака — $X_j^* = (X_j - \mu_j)/\sigma_j$. В этом случае матрица ковариаций наблюдаемого вектора будет совпадать с матрицей корреляций $\Lambda = \mathcal{P} = \mathbf{Corr}(\vec{X})$.

Предположим, что каждый признак X_j может быть представлен как линейная комбинация $k (< p)$ случайных величин f_1, \dots, f_k :

$$X_j = \sum_{i=1}^k q_{ji} f_i + \psi_j e_j, \quad j = \overline{1, p}, \quad k < p. \quad (\text{III.1})$$

Сл. величины f_1, \dots, f_k называются объясняющими переменными или общими факторами. Поскольку они ненаблюдаемы, то их иногда называют латентными признаками. Основное предположение относительно этих признаков состоит в том, что они центрированы, нормированы и не коррелированы между собой. Другими словами, имеет место равенство

$$\mathbf{Cov}(\vec{f}) = \mathbf{E} \vec{f} \vec{f}' = \mathbf{I}.$$

Коэффициенты q_{ji} , отражающие степень влияния общих факторов \vec{f} на наблюдаемые признаки \vec{X} , называются факторными нагрузками (нагрузка i -го фактора на j -ый признак).

Случайная ошибка e_j , связанная только с X_j , представляет собой ту часть наблюдаемого признака, которая не может быть объяснена общими факторами. Поэтому эта случайная составляющая называется характерным фактором. Все

характерные факторы также предполагаются центрированными и нормированными и, кроме того, не коррелирующими как между собой, так и с общими факторами. Коэффициент ψ_j называется характерной нагрузкой j -ого признака.

В матричной форме последнее представление имеет вид

$$\vec{X} = Q \vec{f} + \vec{\psi}' \vec{e}, \quad (\text{III.2})$$

где $Q = (q_{ji})^{(p \cdot k)}$ – матрица нагрузок. Отсюда легко получаем представление матрицы корреляций \mathcal{P} вектора \vec{X} :

$$\mathcal{P} = \mathcal{P}_h + \Psi, \quad (\text{III.3})$$

где

$\Psi = \text{diag}(\psi_1^2, \dots, \psi_p^2)$ – диагональная матрица $(p \cdot p)$, содержащая дисперсии характерных факторов, называемые характерностями,

$\mathcal{P}_h = QQ'$ – редуцированная матрица корреляций, которая отличается от матрицы корреляций \mathcal{P} лишь тем, что на её главной диагонали стоят не единицы, а некоторые положительные (и меньше 1) числа h_1^2, \dots, h_p^2 , называемые общностями.

З а м е ч а н и е 1. Из представления (III.3) видно, что сумма общности и характерности любого признака равна его дисперсии: $h_j^2 + \psi_j^2 = 1$ ($= \mathbf{D} X_j$). Таким образом, общность есть не что иное, как суммарный вклад ненаблюдаемых общих факторов в дисперсию наблюдаемого признака. Среднее значение всех общностей (умноженное на 100%)

$$h^2 / p \cdot 100\% = \frac{1}{p} \sum_{j=1}^p h_j^2 \cdot 100\%$$

представляет собой долю суммарной изменчивости наблюдаемых признаков, которая объясняется влиянием на них ненаблюдаемых факторов.

Поскольку матрица \mathcal{P}_h равна произведению матриц факторных нагрузок

$$\mathcal{P}_h = QQ', \quad (\text{III.4})$$

то общность j -ого признака X_j и его характерность равны

общность X_j	характерность X_j
$h_j^2 = \sum_{i=1}^k q_{ji}^2$	$\psi_j^2 = 1 - h_j^2$

Представление (III.4) называется **фундаментальной теоремой факторного анализа**.

З а м е ч а н и е 2. Факторная модель не единственна. Пусть B – любая ортогональная матрица порядка k и \vec{f} – вектор некоторых факторов, удовлетворяющих (III.2). Тогда, поскольку $B'B = BB' = \mathbf{I}$, то вектор $\vec{f}^* = B'\vec{f}$ также имеет единичную ковариационную матрицу:

$$\text{Cov}(\vec{f}^*) = B' \text{Cov}(\vec{f})B = B'\mathbf{I}B = \mathbf{I}.$$

Полагая $Q^* = QB$, получаем при той же размерности k представление

$$\vec{X} = Q^* \vec{f}^* + \vec{\psi}' \vec{e} \quad (= QB B' \vec{f} + \vec{\psi}' \vec{e} = Q \vec{f} + \vec{\psi}' \vec{e}),$$

отличное от начального (III.2). Это означает, что имеется бесконечное число факторных нагрузок, удовлетворяющих исходным предположениям модели. Подобное затруднение может быть преодолено введением дополнительных ограничений на матрицу Q . Здесь существует несколько вариантов

Ограничения такого типа возникают в ситуациях, когда исследователь располагает некоторой априорной информацией о реальном „смысле“ общих факторов и может априори предполагать отсутствие зависимости исходных признаков от некоторых из общих факторов.

З а м е ч а н и е 3. Так как любое ортогональное преобразование представляет собой вращение системы координат, то из неоднозначности факторного решения можно извлечь некоторую пользу. После того как найдено какое-либо решение, исследователь волен выбрать такой поворот осей f_1, \dots, f_k , при котором получаемые новые факторы $\tilde{f}_1, \dots, \tilde{f}_k$ допускают наиболее естественную и содержательную интерпретацию. С другой стороны, рассматривая расположение исходных признаков в плоскости первых двух общих факторов, естественно попытаться повернуть координатную систему таким образом, чтобы координатные оси прошли через наиболее четко выраженные сгущения точек-признаков. При этом иногда бывает полезно отказаться от ортогональности общих факторов, переходя к косоугольной системе координат. Один из методов вращения, известный как варимакс, реализован в большинстве популярных пакетов статистических программ для РС.

§ 2. Методы построения факторной модели

Факторный анализ по существу представляет собой поиск разложения вида (III.3) с возможно наименьшим числом объясняющих факторов k . Поиск факторного решения осуществляют обычно в два этапа. На первом этапе матрица корреляций \mathcal{P} заменяется редуцированной матрицей \mathcal{P}_h , то есть

тем или иным способом находятся общности всех наблюдаемых признаков. На втором этапе ищется матрица нагрузок Q , удовлетворяющая фундаментальной теореме.

Оценка общностей.

I. По коэффициенту множественной корреляции. Известно, что общность j -го признака ограничена снизу квадратом множественного коэффициента корреляции (I.3)

$$\rho_{j*}^2 \leq h_j^2 .$$

Кроме того, при увеличении числа наблюдаемых признаков X_1, \dots, X_p эта нижняя граница сходится к истинному значению общности. Поэтому метод оценивания общностей по множественному коэффициенту корреляции

$$\hat{h}_j^2 = \rho_{j*}^2$$

наиболее привлекателен не только в силу удобства вычисления по формуле (I.3), стр. 11, но и в силу некой „теоретической обоснованности“ .

II. По наибольшему коэффициенту корреляции в строке.

В этом методе, обоснованном соображениями исключительно эмпирического характера, в качестве оценки общности h_j^2 выбирается максимальное значение модуля коэффициента корреляции в соответствующей j -ой строке матрицы корреляций \mathcal{P} :

$$\hat{h}_j^2 = \max_{\substack{k=1,p \\ k \neq j}} |\rho_{jk}| .$$

III. По средней корреляции в строке. Аналогично предыдущему методу, здесь рассматриваются коэффициенты кор-

реляции в j -ой строке \mathcal{P} , и общность полагается равной среднему значению модулей этих коэффициентов:

$$\hat{h}_j^2 = \frac{1}{p-1} \sum_{\substack{k=1 \\ k \neq j}}^p |\rho_{jk}|.$$

IV. По методу триад. Снова рассматриваются все коэффициенты корреляции j -го признака $\{\rho_{jk}\}, k = \overline{1, p}, k \neq j$. Среди этих коэффициентов выбираются два наибольших по модулю ρ_{jm}, ρ_{jl} и коэффициент корреляции ρ_{ml} , соответствующий признакам X_m, X_l . Оценка общности полагается равной

$$\hat{h}_j^2 = \left| \frac{\rho_{jm} \rho_{jl}}{\rho_{ml}} \right|.$$

Рассмотрим на примере все эти методы. Пусть матрица корреляций 6-мерного случайного вектора равна

$$\mathcal{P} = \begin{pmatrix} 1 & 0.321 & 0.319 & -0.080 & -0.116 & -0.097 \\ 0.321 & 1 & 0.378 & -0.007 & -0.047 & -0.039 \\ 0.319 & 0.378 & 1 & 0.026 & -0.014 & -0.010 \\ -0.080 & -0.007 & 0.026 & 1 & 0.348 & 0.294 \\ -0.116 & -0.047 & -0.014 & 0.348 & 1 & 0.307 \\ -0.097 & -0.039 & -0.010 & 0.294 & 0.307 & 1 \end{pmatrix}.$$

I. Чтобы найти оценку общностей по методу I, необходимо обратить матрицу корреляций \mathcal{P} . Не вдаваясь в подробности вычислений, приведем здесь диагональные элементы матрицы \mathcal{P}^{-1} :

$$1.196 \quad 1.231 \quad 1.232 \quad 1.194 \quad 1.208 \quad 1.159.$$

Таким образом, оценка $\hat{h}_1^2 = 0.164 \approx 1 - 1/1.196$.

II. Максимальный элемент в 1-ой строке равен 0.321. Поэтому оценка общности 1-ого признака по методу II равна

$$\hat{h}_1^2 = 0.321.$$

III. Сумма модулей всех элементов 1-ой строки равна 0.935. Оценка общности 1-ого признака по методу III равна $\hat{h}_1^2 = 0.935 / 5 \approx 0.187$.

IV. Два наибольших элемента 1-ой строки равны 0.321 (второй столбец) и 0.319 (третий столбец). Коэффициент корреляции между вторым и третьим признаками равен 0.378 (пересечение 2-й строки и 3-го столбца), поэтому по методу IV оценка общности первого признака равна $\hat{h}_1^2 = (0.321 \cdot 0.319) / 0.378 \approx 0.271$.

Аналогичным образом находятся оценки всех остальных общностей. Результаты вычислений соберем в следующую таблицу.

Метод	Оценка общностей						Сумма общностей, %
	\hat{h}_1^2	\hat{h}_2^2	\hat{h}_3^2	\hat{h}_4^2	\hat{h}_5^2	\hat{h}_6^2	
I	0.164	0.188	0.189	0.162	0.172	0.137	16.9%
II	0.321	0.378	0.378	0.348	0.348	0.307	34.7%
III	0.187	0.158	0.149	0.151	0.166	0.150	16.0%
IV	0.271	0.380	0.376	0.333	0.362	0.260	33.0%

Выбор того или иного метода зависит от предпочтений исследователя. Чаще всего все эти методы перебирают с целью получения наиболее предпочтительного факторного решения. Практика применения факторного анализа показала, что при большом числе исходных признаков ($p > 10$) ошибки в определении h_j^2 незначительно влияют на результаты содержательной интерпретации.

Построение матрицы нагрузок.

Проблема состоит в том, что не для всякой редуцированной матрицы, то есть не для всякого выбранного набора общностей можно найти соответствующую матрицу нагрузок, удовлетворяющую соотношению (III.4). Поэтому основное требование, которое накладывают на факторное решение, состоит в минимизации отклонения матрицы корреляций QQ' от матрицы корреляций \mathcal{P} . Построенное таким образом факторное решение позволяет наилучшим образом воспроизвести связи наблюдаемых признаков X_1, \dots, X_p .

Метод главных факторов.

Этот метод по сути своей есть метод главных компонент, примененный к редуцированной матрице корреляций \mathcal{P}_h . Предположим, что матрица \mathcal{P}_h неотрицательно определена.

Сначала находятся факторные нагрузки, при которых достигается максимум вклада первого фактора в суммарную общность:

$$\gamma_1 = \vec{q}'_1 \vec{q}_1 = \sum_{j=1}^p a_{j1}^2 \longrightarrow \max$$

при условии, что $QQ' = R_1 = \mathcal{P}_h$.

На каждом k -ом из последующих шагов ищется матрица нагрузок, обеспечивающая достижение максимума вклада k -го фактора в суммарную общность

$$\gamma_k = \vec{q}'_k \vec{q}_k = \sum_{j=1}^p a_{jk}^2 \longrightarrow \max$$

при условии, что $QQ' = R_k = (R_{k-1} - \vec{q}_{k-1} \vec{q}'_{k-1})$, где матрица R_k представляет собой матрицу остаточных коэффициентов

корреляции, полученную после исключения влияния факторов, найденных на предыдущих $k - 1$ шагах.

Можно показать, что, как и в методе главных компонент (см. предыдущую главу), значения $\gamma_1, \gamma_2, \dots$ представляют собой упорядоченные по убыванию собственные числа редуцированной матрицы \mathcal{P}_h . Известно, что сумма собственных чисел равна сумме диагональных элементов матрицы. Поэтому процесс построения факторного решения остановится, когда сумма собственных чисел матрицы \mathcal{P}_h станет равна следу $\text{tr } \mathcal{P}_h = \sum_{j=1}^p \hat{h}_j^2$ – сумме оценок всех общностей. Соответствующие нормированные (деленные на $\sqrt{\gamma_j}$) собственные векторы будут образовывать матрицу нагрузок Q .

З а м е ч а н и е 4. Так как общности вычисляются приближенно, то на практике часто матрица \mathcal{P}_h теряет свойство положительной определенности и, следовательно, часть её действительных собственных чисел будет отрицательной. Поэтому сумма всех положительных собственных чисел будет больше суммы оценок общностей. Однако и в этом случае принято останавливать процесс построения факторной модели на том шаге k , когда впервые сумма действительных собственных чисел $\gamma_1 + \dots + \gamma_k$ будет равна или превысит сумму всех общностей.

Итерационный метод построения факторной модели.

Приведем здесь общую схему выявления структуры модели факторного анализа. Пусть \mathcal{P}_h – редуцированная матрица корреляций наблюдаемого вектора признаков.

1. Задаемся „нулевым“ приближением $\Psi_{(0)}$ матрицы Ψ .
2. Используя основное соотношение факторной модели (III.2), вычисляем „нулевое“ приближение для матрицы

$$V = QQ' :$$

$$V_{(0)} := \mathcal{P}_h - \Psi_{(0)}.$$

3. Находим „нулевые“ приближения столбцов матрицы Q .
4. Определяем следующее „нулевое“ приближение

$$\Psi_{(0)} = Q_{(0)}Q'_{(0)}$$

и переходим ко второму шагу.

Конкретная реализация третьего пункта этой схемы зависит от условий, накладываемых на матрицу Q с целью идентификации факторного решения. Так, например, в случае условия а) (стр.60) столбцы q_1, \dots, q_k матрицы Q являются решениями обобщенного характеристического уравнения

$$V_{(0)}q_i = \gamma_i\Psi_{(0)}q_i \quad (i = 1, \dots, k),$$

где γ_i – i -ый по величине корень уравнения $\|V_{(0)} - \gamma\Psi_{(0)}\| = 0$.

Для второго условия идентификации б) (стр.60) обобщенное характеристическое уравнение заменяется на обычное характеристическое уравнение относительно матрицы $V_{(0)}$.

Для условий типа в) (стр.60) описанный итерационный метод приводит к центроидному методу, имеющему интересную геометрическую интерпретацию. Каждому исходному признаку X_1, \dots, X_p поставим в соответствие некоторый единичный вектор так, чтобы косинусы углов между этими векторами равнялись соответствующим парным коэффициентам корреляции. Тогда 1-й фактор определяется как сумма вспомогательных векторов. Если направления этих векторов изменить так, чтобы большинство коэффициентов корреляции

были положительны, тогда построенный таким образом фактор будет проходить через середину пучка векторов, сгруппированных какбы в одном направлении, то есть будет служить обобщающей характеристикой этого пучка. Следующие факторы находятся аналогично по остаточной матрице корреляций.

Можно показать, что центроидный метод эквивалентен методу главных факторов с дополнительным условием, что собственные векторы матрицы \mathcal{P}_h имеют компоненты, равные ± 1 .

§ 3. Проблема вращения

Конечной целью факторного анализа является получение факторов, поддающихся наибольшей содержательной интерпретации. Это достигается с помощью вращения системы координат (см. Замечание 3 на стр.61). В этой ситуации возникает потребность в выработке определенного критерия качества полученного факторного решения. Таких критериев предложено достаточно много. Большинство из них опирается на некое „эвристическое“ понятие простой структуры. Такая структура должна удовлетворять ряду условий, помогающих упростить „литературное“ оформление полученных факторов. Так, например, одно из условий простой структуры требует, чтобы в каждой строке факторных нагрузок присутствовал хотя бы один нулевой элемент. В этом случае признаки зависят не от всех сразу латентных факторов.

Пример применения вращения.

Рассмотрим пример, показывающий возможности вращения факторного решения. Для матрицы корреляций, приведен-

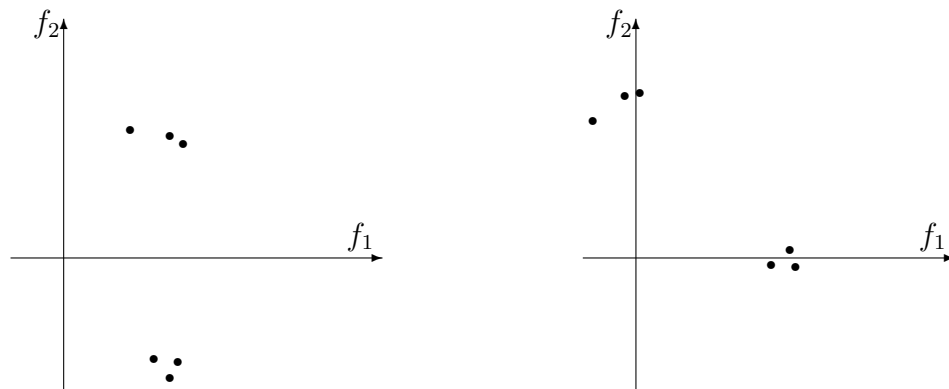
ной на стр.63, с общностями, построенными по методу триад (метод IV), при двух латентных факторах получается следующая матрица факторных нагрузок:

$$Q' = \begin{pmatrix} 0.25 & 0.40 & 0.45 & 0.43 & 0.40 & 0.34 \\ 0.48 & 0.46 & 0.43 & -0.39 & -0.45 & -0.38 \end{pmatrix}.$$

Информативность этого факторного решения близка к нулю. Однако, если произвести поворот системы координат на 45° по ходу часовой стрелки, то получим новую матрицу нагрузок

$$Q' = \begin{pmatrix} -0.16 & -0.04 & 0.01 & 0.58 & 0.60 & 0.51 \\ 0.52 & 0.61 & 0.62 & 0.03 & -0.04 & -0.03 \end{pmatrix}.$$

Таким образом, второй фактор „нагружает“ только 1-й, 2-й и 3-й признаки, а первый фактор – соответственно, 4-й, 5-й и 6-й признаки. Если первые три признака суть время пробегания одним и тем же спортсменом дистанций в 400м., 600м. и 800м., а вторые три признака суть количество подтягиваний этого спортсмена с отягощениями в 0кг., 5кг. и 10кг. соответственно, то отсюда можно сделать вывод, что общая выносливость определяется двумя составляющими: силовой выносливостью и скоростной выносливостью.



На этих рисунках показано расположение шести двумерных векторов факторных нагрузок до и после поворота.

Принцип варимакса.

В основе одного из самых популярных методов вращения, позволяющего достигнуть некоторых оптимальных свойств, лежит принцип наиболее экономного описания точки в двумерной системе координат при прохождении через неё одной из осей координат. Идея метода восходит к связи между общностями и факторными нагрузками

$$\sum_{m=1}^k q_{jm}^2 = h_j^2, \quad j = \overline{1, p},$$

или в преобразованном виде:

$$\sum_{m=1}^k q_{jm}^2 / h_j^2 = 1, \quad j = \overline{1, p}.$$

Просуммировав последние равенства по $j = \overline{1, p}$, получим соотношение

$$\sum_{m=1}^k \sum_{j=1}^p q_{jm}^2 / h_j^2 = p,$$

которое остается верным при любом ортогональном вращении. Таким образом, „среднее значение“ множества величин

$$\left\{ Z_{jm} = q_{jm}^2 / h_j^2, \quad j = \overline{1, p}, \quad m = \overline{1, k} \right\}$$

остается неизменным для всех факторных решений. Принцип варимакса состоит в отыскании (путем вращения) таких нагрузок, для которых суммарная „дисперсия“ величин Z_{jm} по каждому из факторов максимальна:

$$\sum_{m=1}^k \left[\frac{1}{p} \sum_{j=1}^p \left(q_{jm}^2 / h_j^2 \right)^2 - \left(\frac{1}{p} \sum_{j=1}^p q_{jm}^2 / h_j^2 \right)^2 \right] \longrightarrow \max .$$

Глава IV. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Взирая на солнце, прищурь глаза
свои, и ты смело разглядишь в
нем пятна.

На практике чаще всего исходят из предположения нормальности распределения наблюдаемого вектора. Широкая применимость этой модели обусловлена двумя факторами. Во-первых, нормальная распределенность измерений может быть объяснена, исходя из центральной предельной теоремы и предположения, что разброс экспериментальных значений вызван суммарным влиянием большого числа малых воздействий. Во-вторых, для нормальной модели получены наиболее содержательные результаты многомерного статистического анализа. В этой главе мы приведем определение многомерного нормального распределения и дадим его основные свойства.

§ 1. Плотность, характеристическая функция, моменты

О п р е д е л е н и е 1. Говорят, что случайный вектор $\vec{X} = (X_1, \dots, X_p)'$ имеет *p*-мерное нормальное распределение с параметрами $\vec{\mu} = (\mu_1, \dots, \mu_p)'$ и $\Lambda = (\lambda_{jl})_{j,l=1}^p$ (сокращенно

$\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \Lambda)$, если его функция плотности равна

$$f(\vec{x}) = \frac{1}{\sqrt{(2\pi)^p |\Lambda|}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{\mu})' \Lambda^{-1} (\vec{x} - \vec{\mu}) \right\}.$$

З а м е ч а н и е 1. Матрица Λ должна быть невырожденной и положительно определенной. В этом случае функция f действительно будет плотностью вероятностей, так как она положительна и интеграл (после очевидной замены)

$$\int_{\mathcal{R}^n} f(\vec{x}) d\vec{x} = \frac{1}{\sqrt{(2\pi)^p |\Lambda|}} \int_{\mathcal{R}^n} \exp \left\{ -\frac{1}{2} \vec{y}' \Lambda^{-1} \vec{y} \right\} d\vec{y} = 1$$

в силу интегрального равенства В.1, стр. 95, справедливого, в частности, при $\vec{t} = 0$.

Можно определить нормальное распределение и при вырожденной матрице ковариаций. В этом случае распределение задается через характеристическую функцию (см. (IV.1), стр. 73), и это распределение будет сосредоточено в пространстве меньшей, чем p размерности.

Найдем моменты нормального случайного вектора.

Т е о р е м а IV.1. Если $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \Lambda)$, то

$$E \vec{X} = \vec{\mu}, \quad \text{Cov}(\vec{X}) = \Lambda.$$

В дальнейшем всегда будем говорить, что случайный вектор \vec{X} имеет нормальное распределение с вектором средних $\vec{\mu}$ и матрицей ковариаций Λ .

В двумерном случае плотность имеет следующий вид.

Теорема IV.2. Плотность двумерного нормального распределения со средними $\vec{\mu} = (\mu_1, \mu_2)'$ и матрицей вторых моментов

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

где $\rho = \lambda_{12}/\sqrt{\lambda_{11}\lambda_{22}} = \lambda_{12}/\sigma_1\sigma_2$ – коэффициент корреляции между X_1 и X_2 , равна

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp \left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right) \right\}.$$

Вывод всех свойств нормального распределения намного упрощается, если иметь дело не с функцией распределения, а с характеристической функцией. Следующее утверждение непосредственно следует из интегрального равенства В.1, стр. 95.

Теорема IV.3. Если $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \Lambda)$, то характеристическая функция

$$\varphi(\vec{t}) = \varphi_{\vec{X}}(\vec{t}) = \mathbf{E} e^{i\vec{t}'\vec{X}} = \exp \left\{ i\vec{t}'\vec{\mu} - \frac{1}{2}\vec{t}'\Lambda\vec{t} \right\}. \quad (\text{IV.1})$$

Таким образом, с.в. \vec{X} имеет нормальное распределение тогда и только тогда, когда его характеристическая функция задается формулой (IV.1).

§ 2. Линейные преобразования нормального вектора

Важным свойством нормальной модели является ее инвариантность относительно линейных преобразований.

Теорема IV.4. Если $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \Lambda)$ и матрица C размера $k \cdot p$ имеет ранг k , то вектор $\vec{Y} = C \vec{X} \sim \mathcal{N}_k(C \vec{\mu}, C \Lambda C')$.

Из этой теоремы вытекают несколько важных следствий.

Следствие 1. Любая линейная комбинация $c_1 X_1 + \dots + c_p X_p$ нормального случайного вектора \vec{X} имеет одномерное нормальное распределение.

Для доказательства этого утверждения достаточно взять матрицу $C = (c_1, \dots, c_p)$.

Следствие 2. Частное распределение любого случайного подвектора $(X_{j_1}, \dots, X_{j_k})$, полученного из элементов нормального вектора, также нормально с соответствующими параметрами.

Вместо доказательства рассмотрим две матрицы

$$C_1 = (1 \ 0 \ 0 \ \dots \ 0) \text{ и } C_2 = \begin{pmatrix} 0 & 0 & 1 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \end{pmatrix}.$$

Для этих матриц $C_1 \vec{X} = X_1$, а $C_2 \vec{X} = (X_3, X_2)'$. Любые другие наборы компонент получаются аналогичным образом.

З а м е ч а н и е 2. В силу утверждения следствия 2, каждая компонента нормального случайного вектора имеет (одномерное) нормальное распределение. Обратное утверждение, однако, неверно. Рассмотрим функцию плотности

$$f(x_1, x_2, x_3) = \frac{1}{\sqrt{(2\pi)^3}} \exp \left\{ -\frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \right\} \times \quad (\text{IV.2}) \\ \times \left[1 + x_1 x_2 x_3 \exp \left\{ -\frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \right\} \right].$$

Любое частное распределение (как одномерное, так и двумерное) этого трехмерного распределения будет нормальным, хотя само распределение, очевидно, не является нормальным.

Т е о р е м а IV.5. Случайный вектор $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \Lambda)$ тогда и только тогда, когда для любого ненулевого вектора \vec{a} одномерная случайная величина $\vec{a}' \vec{X}$ имеет нормальное распределение с параметрами $(\vec{a}' \vec{\mu}, \vec{a}' \Lambda \vec{a})$.

Это утверждение можно положить в основу определения многомерного нормального распределения, что мгновенно повлечет за собой справедливость всех предыдущих утверждений.

Независимость нормальных случайных величин

Хорошо известно, что из независимости случайных величин вытекает их некоррелированность. Оказывается, что для нормального случайного вектора справедливо и обратное утверждение.

Теорема IV.6. Пусть $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \Lambda)$, тогда

i) с.в. X_i, X_j , $i \neq j$, независимы тогда и только тогда, когда парный коэффициент корреляции $\rho_{ij} = \mathbf{Corr}(X_i, X_j) = 0$;

ii) компоненты X_1, \dots, X_p вектора \vec{X} независимы в совокупности тогда и только тогда, когда все парные коэффициенты корреляции равны нулю.

З а м е ч а н и е 3. Функция плотности (IV.2) дает пример непрерывного распределения, у которого любая пара компонент независима, в то же время весь вектор не независим в совокупности.

Для двумерных случайных векторов имеет место следующий интересный факт.

Теорема IV.7. Случайный вектор (X_1, X_2) имеет двумерное нормальное распределение с одинаковыми дисперсиями компонент тогда и только тогда, когда случайные величины $X_1 + X_2$ и $X_1 - X_2$ независимы и нормально распределены.

Доказательство этой теоремы существенным образом опирается на утверждение теоремы IV.5.

§ 3. Квадратичные формы от нормальных случайных векторов

В статистике важную роль играют случайные величины, образованные как квадратичные формы от некоторого (чаще всего, нормального) случайного вектора. Имеет место следующая

Теорема IV.8. Если $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \Lambda)$, то квадратичная форма $Y^2 = (\vec{X} - \vec{\mu})' \Lambda^{-1} (\vec{X} - \vec{\mu})$ имеет хи-квадрат распределение с p степенями свободы.

Замечание 4. Если все средние значения вектора \vec{X} равны нулю ($\vec{\mu} = \vec{0}$), все дисперсии равны 1 и все компоненты попарно не коррелируют ($\mathbf{Cov}(\vec{X}) = \Lambda = \mathbf{I}$), то квадратичная форма в последней теореме представляет собой не что иное, как сумму квадратов

$$X_1^2 + X_2^2 + \dots + X_p^2,$$

которая, как известно, определяет хи-квадрат распределение с p степенями свободы.

§ 4. Условные распределения

Следующие две теоремы касаются условных распределений нормальных случайных векторов. Эти теоремы играют ключевую роль в регрессионном анализе. В первой теореме рассматривается условное распределение одной компоненты нормального вектора при фиксированных всех остальных.

Теорема IV.9. Если $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \Lambda)$, то условное распределение X_1 при фиксированных значениях $X_2 = x_2, \dots, X_p = x_p$ нормально со средним, равным линейной регрессии X_1 на (X_2, \dots, X_p) ,

$$\mu_{1*(2\dots p)} = \mu_{1*(x_2, \dots, x_p)} = \mu_1 + \sum_{j=2}^p (-1)^j \frac{|\Lambda_{1j}|}{|\Lambda_{11}|} (x_j - \mu_j),$$

и дисперсией, равной остаточной дисперсии,

$$\sigma_{1*(2\dots p)}^2 = \frac{|\Lambda|}{|\Lambda_{11}|}.$$

Таким образом, условное среднее линейно зависит от зафиксированных значений компонент сл.вектора \vec{X} . Другими словами,

среднеквадратическая регрессия в нормальной модели совпадает с линейной среднеквадратической регрессией.

Еще раз подчеркнем, что этот факт служит одним из оправданий того, что на практике чаще всего строится именно линейная регрессия.

В следующей теореме рассматривается условное распределение двух компонент нормального вектора при фиксированных всех остальных.

Теорема IV.10. Если $\vec{X} \sim \mathcal{N}_p(\vec{\mu}, \Lambda)$, то условное распределение (X_1, X_2) при фиксированных значениях $X_3 = x_3, \dots, X_p = x_p$ нормально $\mathcal{N}_2(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \rho_{12*})$ со средними, равными соответствующим линейным регрессиям X_1 и X_2 на (X_3, \dots, X_p) ,

$$\hat{\mu}_1 = \mu_{1*}(x_3, \dots, x_p), \quad \hat{\mu}_2 = \mu_{2*}(x_3, \dots, x_p),$$

с дисперсиями, равными остаточным дисперсиям,

$$\hat{\sigma}_1^2 = \sigma_{1*(3\dots p)}^2, \quad \hat{\sigma}_2^2 = \sigma_{2*(3\dots p)}^2,$$

и коэффициентом корреляции, равным частному коэффициенту корреляции между X_1, X_2 за вычетом влияния (X_3, \dots, X_p) ,

$$\rho_{12*} = \rho_{12*(3\dots p)} = \frac{|\Lambda_{12}|}{\sqrt{|\Lambda_{11}| |\Lambda_{22}|}}.$$

Замечание 5. Таким образом, в нормальной модели частный коэффициент корреляции равен условному коэффициенту корреляции при фиксированных значениях мешающих переменных. На практике этот факт переносят обычно на модели общего вида и интерпретируют частный коэффициент корреляции как условный коэффициент корреляции.

Доказательства

Доказательство Теоремы IV.4.

Найдем характеристическую функцию вектора \vec{Y} . Пусть \vec{t} – произвольный числовой k -мерный вектор и $\vec{s}' = \vec{t}'C$. Тогда

$$\varphi_{\vec{Y}}(\vec{t}) = \mathbf{E} e^{i\vec{t}'\vec{Y}} = \mathbf{E} e^{i\vec{t}'C\vec{X}} = \mathbf{E} e^{i\vec{s}'\vec{X}}.$$

Последнее выражение представляет собой значение характеристической функции случайного вектора \vec{X} в точке \vec{s} , следовательно,

$$\varphi_{\vec{Y}}(\vec{t}) = \exp \left\{ \vec{s}'\vec{\mu} - \frac{1}{2} \vec{s}'\Lambda\vec{s} \right\} = \exp \left\{ \vec{t}'(C\vec{\mu}) - \frac{1}{2} \vec{t}'(C\Lambda C')\vec{t} \right\},$$

что совпадает с характеристической функцией k -мерного нормального закона $\mathcal{N}_k(C\vec{\mu}, C\Lambda C')$.

⊠

Доказательство Замечания 2.

Покажем, что эта функция есть плотность. Во-первых, она всегда положительна, поскольку $|x \exp\{-x^2/2\}| \leq 1/\sqrt{e} < 0.7$, и следовательно, выражение в квадратных скобках > 0 . Далее, так как второе слагаемое в квадратных скобках является нечетной функцией x_3 , то соответствующий этому слагаемому интеграл при интегрировании $f(x_1, x_2, x_3)$ по x_3 в пределах от $-\infty$ до ∞ равен нулю, а интеграл от $\exp\{-x_3^2/2\}/\sqrt{2\pi}$ в этой же области равен 1. Таким образом,

$$\int_{-\infty}^{\infty} f(x_1, x_2, x_3) dx_3 = \frac{1}{2\pi} \exp \left\{ -\frac{1}{2}(x_1^2 + x_2^2) \right\},$$

что совпадает с функцией плотности двумерного нормального закона $\mathcal{N}_2(\vec{0}, \mathbf{I})$. Отсюда заключаем, что, во-первых, функ-

ция $f(x_1, x_2, x_3)$ является плотностью некоторого случайного вектора \vec{X} , во-вторых, любая двумерная маргинальная плотность есть плотность двумерного нормального закона, а следовательно, каждая компонента \vec{X} нормальна.

⊠

Доказательство Теоремы IV.5.

Пусть \vec{t} – произвольный числовой вектор, \vec{m} – вектор средних значений, L – матрица ковариаций \vec{X} . Тогда характеристическая функция \vec{X}

$$\varphi_{\vec{X}}(\vec{t}) = \mathbf{E} e^{i\vec{t}'\vec{X}} = \mathbf{E} e^{iY},$$

где $Y = \vec{t}'\vec{X}$. Последнее выражение есть характеристическая функция случайной величины Y , вычисленная в точке 1. Среднее Y равно

$$\mathbf{E} Y = \mathbf{E} \vec{t}'\vec{X} = \vec{t}'\vec{m},$$

а дисперсия $\mathbf{D}(Y) = \mathbf{Cov}(Y) = \vec{t}'L\vec{t}$.

По предположению теоремы, распределение Y нормально, поэтому характеристическая функция (при аргументе, равном 1) равна

$$\varphi_Y(1) = \exp\{i \cdot 1 \cdot \mathbf{E} Y - \frac{1}{2} \mathbf{D}(Y) \cdot 1^2\} = \exp\{i \vec{t}'\vec{m} - \frac{1}{2} \vec{t}'L\vec{t}\},$$

что совпадает с характеристической функцией p -мерного нормального случайного вектора с параметрами $\vec{\mu} = \vec{m}$, $\Lambda = L$. Что, в частности, доказывает утверждение теоремы IV.1 о моментах.

⊠

Доказательство Теоремы IV.6.

і) Необходимым и достаточным условием независимости случайных величин является представление их совместной функции плотности в виде произведения плотностей компонент.

Для двумерного случайного вектора (X_i, X_j) такое представление возможно тогда и только тогда, когда коэффициент корреляции $\rho = \rho_{ij} = 0$, что весьма просто вытекает из вида двумерной плотности (теорема IV.2).

ii) В условиях теоремы ковариационная матрица \vec{X} диагональна – $\Lambda = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$. Поэтому совместная плотность

$$\begin{aligned} f(x_1, \dots, x_p) &= \frac{1}{\sqrt{(2\pi)^p \sigma_1^2 \cdots \sigma_p^2}} e^{-\frac{1}{2} \sum_{i=1}^p \frac{(x_i - \mu_i)^2}{\sigma_i^2}} = \\ &= \prod_{i=1}^p \frac{1}{\sqrt{2\pi \sigma_i^2}} e^{-\frac{1}{2} \frac{(x_i - \mu_i)^2}{\sigma_i^2}}, \end{aligned}$$

т.е. равна произведению частных плотностей компонент.

⊠

Доказательство Теоремы IV.7.

Предположим сначала, что случайный вектор $(X_1, X_2) \sim \mathcal{N}_2(\vec{\mu}, \Lambda)$ с $\lambda_{11} = \lambda_{22}$ (т.е. с одинаковыми дисперсиями). Тогда, по теореме IV.4, случайный вектор $(X_1 + X_2, X_1 - X_2) \sim \mathcal{N}_2$ как линейное преобразование нормального вектора. Ковариация

$$\begin{aligned} \text{Cov}(X_1 + X_2, X_1 - X_2) &= \\ &= \mathbf{E}(X_1 + X_2 - (\mu_1 + \mu_2))(X_1 - X_2 - (\mu_1 - \mu_2)) = \quad (\text{IV.3}) \\ &= \mathbf{E}(X_1 - \mu_1)^2 - \mathbf{E}(X_2 - \mu_2)^2 = \lambda_{11} - \lambda_{22} = 0. \end{aligned}$$

Следовательно, в силу утверждения теоремы IV.6, компоненты вектора $(X_1 + X_2, X_1 - X_2)$ независимы.

Для доказательства обратного утверждения рассмотрим произвольное линейное преобразование

$$c_1 X_1 + c_2 X_2 = \frac{1}{2}(c_1 + c_2)(X_1 + X_2) + \frac{1}{2}(c_1 - c_2)(X_1 - X_2).$$

Так как по условию теоремы случайные величины $X_1 + X_2$ и $X_1 - X_2$ независимы и имеют нормальное распределение, то вектор $(X_1 + X_2, X_1 - X_2) \sim \mathcal{N}_2$, и следовательно, линейная комбинация $c_1 X_1 + c_2 X_2 \sim \mathcal{N}_1$. Отсюда и из теоремы IV.5 получаем, что вектор (X_1, X_2) нормально распределен. Совпадение дисперсий X_1 и X_2 следует из условия независимости $(X_1 + X_2), (X_1 - X_2)$ и соотношения IV.3.

⊠

Доказательство Теоремы IV.8.

Известно, что для любой положительно определенной матрицы Λ^{-1} существует такая невырожденная матрица Q , что $Q'\Lambda^{-1}Q = I$. Рассмотрим вектор \vec{Z} , связанный с \vec{X} линейным преобразованием $\vec{Z} = Q^{-1}(\vec{X} - \vec{\mu})$. По теореме IV.4 вектор \vec{Z} распределен нормально с вектором средних $\mathbf{E} \vec{Z} = Q^{-1} \mathbf{E}(\vec{X} - \vec{\mu}) = \vec{0}$ и матрицей ковариаций $\mathbf{Cov}(\vec{Z}) = Q^{-1}\Lambda(Q^{-1})' = I$. Последнее равенство следует из соотношения $Q'\Lambda^{-1}Q = I$, если в нем произвести операцию обращения с обеих сторон знака равенства. Таким образом, вектор \vec{Z} состоит из независимых одинаковых нормальных $(0, 1)$ случайных величин. Квадратичная форма

$$Y^2 = (\vec{X} - \vec{\mu})'\Lambda^{-1}(\vec{X} - \vec{\mu}) = \vec{Z}'Q'\Lambda^{-1}Q\vec{Z} = \vec{Z}'\vec{Z} = \sum_{i=1}^p Z_i^2,$$

то есть равна сумме квадратов p независимых стандартных нормальных случайных величин, и, следовательно, имеет хи-квадрат распределение с p степенями свободы.

⊠

Доказательство Теоремы IV.9.

Рассмотрим функцию плотности

$$f(\vec{x}) = \frac{1}{\sqrt{(2\pi)^p |\Lambda|}} \exp\left\{-\frac{1}{2} Q\right\}$$

с квадратичной формой

$$Q = Q(x_1, \dots, x_p) = \vec{x}' \Lambda^{-1} \vec{x} = \sum_{ij} (-1)^{i+j} \frac{|\Lambda_{ij}|}{|\Lambda|} x_i x_j,$$

где в целях упрощения записи произведена замена $\vec{x} - \vec{\mu} \mapsto \vec{x}$. Распишем теперь Q , выделив слагаемые, содержащие x_1 :

$$\begin{aligned} Q &= \frac{|\Lambda_{11}|}{|\Lambda|} x_1^2 - 2x_1 \sum_{j=2}^p (-1)^j \frac{|\Lambda_{1j}|}{|\Lambda|} x_j + C_0 = \\ &= \frac{|\Lambda_{11}|}{|\Lambda|} \left(x_1 - \sum_{j=2}^p (-1)^j \frac{|\Lambda_{1j}|}{|\Lambda_{11}|} x_j \right)^2 + C_1, \end{aligned}$$

с константами C_0 и C_1 , зависящими только от x_2, \dots, x_p . Поэтому, если обозначить плотность вектора X_2, \dots, X_p через $f_*(\vec{x}^{(1)})$, то условная плотность X_1 может быть записана в виде

$$\begin{aligned} f_{X_1}(x_1 | x_2, \dots, x_p) &= \frac{f(\vec{x})}{f_*(\vec{x}^{(1)})} = \\ &= C \exp \left\{ -\frac{1}{2} \frac{|\Lambda_{11}|}{|\Lambda|} (x_1 - \mu_{1*(2..p)})^2 \right\}, \end{aligned}$$

где снова константа C не зависит от x_1 . Полученное выражение есть плотность нормального закона с указанными в теореме параметрами.

⊠

Доказательство Теоремы IV.10.

Утверждения о средних и дисперсиях легко выводятся из предыдущей теоремы переходом к частным распределениям. Вывод формулы для коэффициента корреляции, основанное на сложных матричных соотношениях, мы опускаем.

⊠

Приложение А. СЛУЧАЙНЫЕ ВЕКТОРЫ

Многие вещи нам непонятны не потому, что наши понятия слабы; но потому, что сии вещи не входят в круг наших понятий.

Элементы векторной алгебры

✓ *Вектор (вектор-столбец)* – упорядоченный набор чисел

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix}.$$

✓ *Вектор-строка* – транспонированный вектор-столбец

$$\vec{x}' = (x_1, x_2, \dots, x_k).$$

✓ Векторы $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ *линейно независимы*, если уравнение

$$a_1\vec{x}_1 + a_2\vec{x}_2 + \dots + a_n\vec{x}_n = 0$$

имеет только нулевое решение: $a_i = 0, i = \overline{1, n}$. Число линейно независимых векторов не может быть больше размерности: $n \leq k$.

✓ *Скалярное произведение* векторов одинаковой размерности

$$\vec{x}'\vec{y} = x_1y_1 + x_2y_2 + \dots + x_ky_k.$$

- ✓ Длина вектора $|\vec{x}| = \sqrt{\vec{x}'\vec{x}} = \sqrt{x_1^2 + \dots + x_k^2}$.
- ✓ Набор n вектор-строк размера k (или k вектор-столбцов размера n) образует *матрицу* размера $n \cdot k$ (n строк, k столбцов)

$$A = A^{n \cdot k} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{pmatrix} = \begin{pmatrix} \vec{a}'_1 \\ \vec{a}'_2 \\ \vdots \\ \vec{a}'_n \end{pmatrix}.$$

Матрицу, получающуюся из матрицы A путем вычеркивания i -ой строки и j -го столбца, будем обозначать A_{ij} .

- ✓ Матрицу, у которой все элементы, находящиеся не на главной диагонали, равны нулю — $a_{ij} = 0$, $i \neq j$, будем называть *диагональной* и обозначать

$$\text{diag}(a_1, \dots, a_k),$$

с указанием соответствующих элементов главной диагонали.

- ✓ *Сумма* двух матриц одинаковой размерности определяется как матрица, элементы которой равны сумме элементов слагаемых матриц.

- ✓ *Произведение* двух матриц AB определяется только в том случае, если количество столбцов первого сомножителя A совпадает с количеством строк второго сомножителя B . Пусть $A = A^{n \cdot k}$, $B = B^{k \cdot m}$, тогда произведение AB будет иметь размерность $n \cdot m$, и элемент этого произведения, стоящий на пересечении i -ой строки и j -ого столбца, равен скалярному произведению i -ой строки матрицы A на j -ый столбец матрицы B :

$$AB = \left(\vec{a}'_i \vec{b}_j \right)_{i,j=1}^k.$$

- ✓ *Определителем* квадратной матрицы $A^{k \cdot k}$ называется

число $|A|$, равное сумме всевозможных произведений k элементов матрицы, взятых по одному из каждой строки и каждого столбца, со знаком, определяемым четностью перестановки индексов этих элементов.

(\mathfrak{F}) *Разложение* определителя квадратной матрицы по строке:

$$\sum_{j=1}^k a_{ij} (-1)^{j+m} |A_{mj}| = \begin{cases} |A|, & \text{если } m = i, \\ 0, & \text{если } m \neq i. \end{cases}$$

✓ *Следом* квадратной матрицы $A^{k \cdot k}$ называется сумма элементов ее главной диагонали:

$$\text{tr } A = \sum_{i=1}^k a_{ii}.$$

(\mathfrak{F}) Если определены произведения матриц AB и BA , то

$$\text{tr}(AB) = \text{tr}(BA).$$

✓ Для любой невырожденной ($|A| \neq 0$) квадратной матрицы существует *обратная матрица* A^{-1} , определяемая соотношениями

$$AA^{-1} = A^{-1}A = \mathbf{I}.$$

(\mathfrak{F}) Элементы обратной матрицы \tilde{a}_{ij} вычисляются по формуле

$$\tilde{a}_{ij} = \frac{(-1)^{i+j} |A_{ji}|}{|A|}.$$

✓ Матрица A *симметрична*, если она совпадает со своей транспонированной: $A = A'$.

✓ Симметричная матрица A *положительно определена* ($A > 0$), если для всех векторов $\vec{x} \neq 0$ *квадратичная форма*

$$\vec{x}' A \vec{x} = \sum_{i,j} a_{ij} x_i x_j > 0.$$

(\mathfrak{F}) Для положительно определенной матрицы A

$$|A| \leq a_{11} |A_{11}| \leq \dots \leq \prod_{j=1}^k a_{jj},$$

причем знак равенства в каждом из неравенств достигается только в том случае, если все элементы кроме диагонального в соответствующей строке равны нулю.

✓ Матрица A ортогональна, если $A'A = AA' = \mathbf{I}$.

(\mathfrak{F}) Для любой симметричной матрицы A существует такая ортогональная матрица Q , что

$$Q'AQ = \text{diag}(\lambda_1, \dots, \lambda_k).$$

✓ Здесь числа λ_i суть так называемые *собственные числа* матрицы A , определяемые как решение *характеристического уравнения*

$$|A - \lambda \mathbf{I}| = 0.$$

(\mathfrak{F}) Все собственные числа симметричной матрицы действительны.

(\mathfrak{F}) Если матрица A положительно определена, то все ее собственные числа строго положительны.

✓ Каждому собственному числу λ матрицы A соответствует *собственный вектор* \vec{x} , удовлетворяющий уравнению

$$A\vec{x} = \lambda\vec{x}.$$

(\mathfrak{F}) Собственные векторы, соответствующие различным ненулевым собственным числам симметричной матрицы, ортогональны.

(\mathfrak{F}) Если собственное число имеет кратность m , то этому числу соответствуют m взаимно ортогональных собственных векторов.

✓ Ранг матрицы $\text{rang} A$, определяется как число линейно независимых столбцов (или строк) A .

(\mathfrak{Z}) Ранг матрицы равен количеству ненулевых собственных чисел A .

Случайные векторы

✓ Случайный вектор (сл.вектор) $\vec{X} = (X_1, \dots, X_k)' = \vec{X}^{(k)}$ задается своей функцией распределения

$$F(x_1, \dots, x_k) = \mathbf{P}\{X_1 < x_1, \dots, X_k < x_k\}.$$

✓ Функция плотности $f(x_1, \dots, x_k)$, если она существует, определяется как любая функция, удовлетворяющая соотношению

$$F(x_1, \dots, x_k) = \int_{-\infty}^{x_1} dt_1 \cdots \int_{-\infty}^{x_k} dt_k f(t_1, \dots, t_k).$$

(\mathfrak{Z}) Функция плотности может быть вычислена как k -ая смешанная производная

$$f(x_1, \dots, x_k) = \frac{\partial^k F(x_1, \dots, x_k)}{\partial x_1 \cdots \partial x_k}.$$

✓ Вектор средних значений

$$\vec{\mu} = \mathbf{E} \vec{X} = (\mathbf{E} X_1, \dots, \mathbf{E} X_k)'$$

✓ Величина

$$\lambda_{ij} = \mathbf{E}(X_i - \mu_i)(X_j - \mu_j) = \mathbf{E} X_i X_j - \mu_i \mu_j$$

называется *ковариацией* с.в. X_i, X_j .

✓ Ковариация X_i с X_i есть дисперсия с.в. X_i :

$$\lambda_{ii} = \mathbf{D} X_i = \sigma_i^2.$$

✓ Матрица ковариаций

$$\Lambda = (\lambda_{ij})_{i,j=1}^k = \mathbf{Cov}(\vec{X}) = \mathbf{E} \vec{X} \vec{X}' - \vec{\mu} \vec{\mu}'.$$

(\mathfrak{F}) Если случайный вектор $\vec{Y} = A \vec{X}$, где A – некоторая матрица подходящего размера, то матрица ковариаций

$$\mathbf{Cov}(\vec{Y}) = A \mathbf{Cov}(\vec{X}) A'.$$

(\mathfrak{F}) Дисперсия линейной комбинации компонент сл.вектора \vec{X} равна

$$\mathbf{D}(\vec{t}' \vec{X}) = \vec{t}' \Lambda \vec{t}.$$

✓ Коэффициент корреляции между с.в. X_i и X_j равен

$$\rho_{ij} = \frac{\lambda_{ij}}{\sigma_i \sigma_j}.$$

(\mathfrak{F}) Матрица корреляций

$$\mathcal{P} = \mathbf{Corr}(\vec{X}) = (\rho_{ij})_{i,j=1}^k = \Sigma^{-1} \Lambda \Sigma^{-1},$$

где $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k)$ – диагональная матрица стандартных отклонений.

Теорема А.1. Матрица ковариаций $\Lambda = \text{Cov}(\vec{X}) \geq 0$ – неотрицательно определена, причем она вырождена ($|\Lambda| = 0$) тогда и только тогда, когда компоненты вектора \vec{X} стохастически линейно зависимы:

$$c_1 X_1 + \dots + c_k X_k = c_0 \quad (\mathbf{P} - \text{п.н.})$$

для некоторого ненулевого вектора детерминированных (не случайных) констант (c_0, c_1, \dots, c_k) . Другими словами, случайный вектор \vec{X} сосредоточен в пространстве меньшей размерности.

✓ Другой способ задания распределения сл.вектора - выбор соответствующей *характеристической функции*

$$\varphi(\vec{t}) = \varphi_{\vec{X}}(\vec{t}) = \mathbf{E} e^{i\vec{t}'\vec{X}}, \quad \vec{t} \in \mathcal{R}^k.$$

Свойства характеристических функций.

(\mathfrak{A}) Если вектор \vec{Y} состоит из первых m компонент вектора \vec{X} , то

$$\varphi_{\vec{Y}}(t_1, \dots, t_m) = \varphi_{\vec{X}}(t_1, \dots, t_m, 0, \dots, 0).$$

(\mathfrak{B}) Сл.вектор $\vec{Y} = \vec{Y}^{(m)}$ независит от сл.вектора $\vec{Z} = \vec{Z}^{(l)}$ только тогда, когда совместная характеристическая функция сл.вектора $\vec{X} = (\vec{Y}', \vec{Z}')$ может быть представлена в виде произведения характеристических функций \vec{Y} и \vec{Z} :

$$\varphi_{\vec{X}}(t_1, \dots, t_{m+l}) = \varphi_{\vec{Y}}(t_1, \dots, t_m) \varphi_{\vec{Z}}(t_{m+1}, \dots, t_{m+l}).$$

(\mathfrak{C}) Последовательность сл.векторов \vec{X}_n слабо (по распределению) сходится к сл.вектору \vec{X} только тогда, когда предел

последовательности характеристических функций

$$\lim_{n \rightarrow \infty} \varphi_{\vec{X}_n}(\vec{t}) = \varphi_{\vec{X}}(\vec{t}), \quad \forall \vec{t} \in \mathcal{R}^k.$$

(\mathfrak{Z}) Если сл.вектор \vec{X} нормален с параметрами $(\vec{\mu}, \Lambda)$, то

$$\varphi_{\vec{X}}(\vec{t}) = \exp \left(i \vec{t}' \vec{\mu} - \frac{1}{2} \vec{t}' \Lambda \vec{t} \right).$$

Следующая теорема содержит утверждения, называемые обычно *Законом Больших Чисел* и *Центральной Предельной Теоремой* и доказываемые с помощью аппарата характеристических функций.

Теорема А.2. Если $\vec{X}_1, \dots, \vec{X}_n$ – выборка объема n из p -мерного распределения с вектором средних $\vec{\mu}$, то

(\mathfrak{Z}) выборочное среднее $\vec{X}_\cdot = \sum_{i=1}^n \vec{X}_i / n$ сходится (при $n \rightarrow \infty$) по вероятности к $\vec{\mu}$:

$$\vec{X}_\cdot \xrightarrow{\mathbf{P}} \vec{\mu} \quad (n \rightarrow \infty);$$

(\mathfrak{Z}) если, вдобавок, существует матрица ковариаций Λ генеральной совокупности, то выборочное среднее асимптотически нормально с вектором средних $\vec{\mu}$ и матрицей ковариаций Λ/n :

$$\sqrt{n} (\vec{X}_\cdot - \vec{\mu}) \rightsquigarrow \mathcal{N}_p(\vec{0}, \Lambda).$$

В следующей теореме устанавливается асимптотическое распределение функции от асимптотически нормального случайного вектора.

Теорема А.3. Пусть

- i) p -мерный вектор \vec{U}_n асимптотически нормален $\mathcal{N}_p(\vec{b}, \frac{1}{n}Q)$;
- ii) функция $f(\vec{u})$ дважды дифференцируема в окрестности точки \vec{b} ;
- iii) $\vec{\phi}$ – вектор первых производных f в точке \vec{b} .

Тогда случайная величина $f(\vec{U}_n)$ асимптотически нормальна со средним $\mu = f(\vec{b})$ и дисперсией $\sigma_n^2 = \frac{1}{n} \vec{\phi}' Q \vec{\phi}$:

$$\frac{f(\vec{U}_n) - \mu}{\sigma_n} \rightsquigarrow \mathcal{N}_1(0, 1).$$

Доказательство Теоремы А.1.

Первая часть теоремы следует из того, что для любого вектора \vec{t} квадратичная форма $\vec{t}' \Lambda \vec{t} = \mathbf{D}(\vec{t}' \vec{X}) \geq 0$. Если матрица $\Lambda = \mathbf{Cov}(\vec{X})$ вырождена, тогда строки матрицы Λ линейно зависимы, и найдется такой ненулевой вектор \vec{c} , что $\vec{c}' \Lambda = \vec{0}'$. Поэтому дисперсия случайной величины $\vec{c}' \vec{X}$ равна

$$\mathbf{D} \vec{c}' \vec{X} = \mathbf{Cov}(\vec{c}' \vec{X}) = \vec{c}' \Lambda \vec{c} = 0.$$

Следовательно, линейная комбинация $\vec{c}' \vec{X}$ почти наверное равна константе, что доказывает утверждение теоремы.

Обратно, если линейная комбинация $\vec{c}' \vec{X}$ почти наверное равна константе, то одна из случайных величин, например, X_1 линейно выражается через все остальные: $X_1 = b_2 X_2 + \dots + b_k X_k + b_0$. Найдём отсюда ковариации X_1 со всеми остальными компонентами.

$$\begin{aligned} \mathbf{Cov}(X_1, X_i) &= \lambda_{1i} = b_2 \mathbf{Cov}(X_2, X_i) + \dots + b_k \mathbf{Cov}(X_k, X_i) = \\ &= b_2 \lambda_{2i} + \dots + b_k \lambda_{ki}. \end{aligned}$$

Таким образом, первая строка матрицы Λ линейно зависит от остальных строк, то есть матрица Λ вырождена.

⊠

Доказательство Теоремы А.3.

Рассмотрим разложение функции f при $\vec{u} \rightarrow \vec{b}$:

$$f(\vec{u}) = f(\vec{b}) + \vec{\phi}'(\vec{u} - \vec{b}) + O(|\vec{u} - \vec{b}|^2).$$

Отсюда при $n \rightarrow \infty$

$$\frac{f(\vec{U}_n) - \mu}{\sigma_n} = \frac{\vec{\phi}'(\vec{U}_n - \vec{b})}{\sigma_n} + O\left(\frac{|\vec{U}_n - \vec{b}|^2}{\sigma_n}\right).$$

С помощью известного неравенства Чебышева легко устанавливается, что по вероятности при $n \rightarrow \infty$ последовательность

$$|\vec{U}_n - \vec{b}|^2 / \sigma_n \longrightarrow 0.$$

С другой стороны, предельное значение последовательности характеристических функций случайных величин $\sqrt{n} \vec{\phi}'(\vec{U}_n - \vec{b})$ равно

$$\begin{aligned} \lim_{n \rightarrow \infty} \varphi_n(t) &= \lim_{n \rightarrow \infty} \mathbf{E} \exp\{i t \sqrt{n} \vec{\phi}'(\vec{U}_n - \vec{b})\} = \\ &= \exp\left\{-\frac{1}{2}(t \vec{\phi})' Q (t \vec{\phi})\right\} = \exp\left\{-\frac{1}{2} \vec{\phi}' Q \vec{\phi} t^2\right\}, \end{aligned}$$

что совпадает с характеристической функцией одномерного нормального закона с нулевым средним и дисперсией $n\sigma_n^2$. Следовательно, $\vec{\phi}'(\vec{U}_n - \vec{b})/\sigma_n \rightsquigarrow \mathcal{N}_1(0, 1)$.

Окончательное доказательство теоремы следует теперь из известной теоремы Слуцкого, утверждающей, что сумма асимптотически нормальной $(0,1)$ последовательности с.в. и асимптотически „нулевой“ последовательности также будет асимптотически нормальной $(0,1)$.

⊠

Приложение В. НЕКОТОРЫЕ
ИНТЕГРАЛЬНЫЕ СООТНОШЕНИЯ

✓ Гамма-функция: $\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx, \quad p > 0.$

Свойства

- (ℑ) $\Gamma(p+1) = p\Gamma(p); \quad \Gamma(n) = (n-1)!;$
- (ℑ) формула Стирлинга $\Gamma(p) \sim \left(\frac{p}{e}\right)^p \frac{\sqrt{2\pi}}{\sqrt{p}}, \quad \text{при } p \rightarrow \infty;$
- (ℑ) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$

✓ Бета-функция: $B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx, \quad p, q > 0.$

Свойства

- (ℑ) $B(p, q) = \Gamma(p)\Gamma(q) / \Gamma(p+q);$
- (ℑ) $B(p, p) = B\left(p, \frac{1}{2}\right) / 2^{2p-1}$

(ℑ) Для любой положительно определенной матрицы A и любого k -мерного вектора \vec{t}

$$\int_{\mathcal{R}^k} e^{(i\vec{t}'\vec{x} - \frac{1}{2}\vec{x}'A\vec{x})} d\vec{x} = \frac{(\sqrt{2\pi})^k}{\sqrt{|A|}} e^{-\frac{1}{2}\vec{t}'A^{-1}\vec{t}}. \quad (\text{B.1})$$

(ℑ) Объем k -мерного эллипсоида $\vec{x}'A\vec{x} \leq c^2$ равен

$$V_{\text{элл}} = \int_{\vec{x}'A\vec{x} \leq c^2} d\vec{x} = \frac{(\sqrt{\pi})^k}{\Gamma\left(\frac{k}{2} + 1\right)} \frac{c^k}{\sqrt{|A|}}. \quad (\text{B.2})$$

(\Im) Смешанные моменты

$$\int_{\vec{x}'A\vec{x} \leq c^2} \vec{x} \vec{x}' d\vec{x} = \frac{c^2 V_{\mathcal{E}\mathcal{L}\mathcal{L}}}{k+2} A^{-1}. \quad (\text{B.3})$$

Доказательство соотношения В.1.

Для матрицы A существует такая ортогональная матрица Q , что $Q'AQ = \Lambda$ – диагональная матрица со строго положительными элементами $\lambda_1, \dots, \lambda_k$. Положим $\vec{y} = Q'\vec{x}$. Тогда $\vec{x} = Q\vec{y}$, и поскольку якобиан ортогонального преобразования равен 1, то искомый интеграл равен

$$\begin{aligned} I &= \int_{\mathcal{R}^k} e^{i(\vec{t}'Q\vec{y} - \frac{1}{2}\vec{y}'Q'AQ\vec{y})} d\vec{y} = \int_{\mathcal{R}^k} e^{i(\vec{u}'\vec{y} - \frac{1}{2}\vec{y}'\Lambda\vec{y})} d\vec{y} = \\ &= \int_{\mathcal{R}^k} e^{i(u_1 y_1 + \dots + u_k y_k) - \frac{1}{2}(\lambda_1 y_1^2 + \dots + \lambda_k^2 y_k^2)} d\vec{y} = \prod_{j=1}^k \int_{-\infty}^{\infty} e^{i u_j y_j - \frac{1}{2} \frac{y_j^2}{\lambda_j}} dy_j, \end{aligned}$$

где произведена замена $\vec{u} = Q'\vec{t}$. Интеграл под знаком произведения представляет собой умноженную на $\sqrt{2\pi/\lambda_j}$ характеристическую функцию нормального $\mathcal{N}_1(0, \frac{1}{\lambda_j})$ распределения. Эта характеристическая функция, как известно, равна $\exp(-u_j^2/2\lambda_j)$. Поэтому, так как $\prod_1^k \lambda_j = |\Lambda| = |A|$, то

$$I = \frac{(\sqrt{2\pi})^k}{\sqrt{\prod_1^k \lambda_j}} e^{-\frac{1}{2}(\frac{u_1^2}{\lambda_1} + \dots + \frac{u_k^2}{\lambda_k})} = \frac{(\sqrt{2\pi})^k}{\sqrt{|A|}} e^{-\frac{1}{2}(\vec{u}'\Lambda^{-1}\vec{u})}.$$

Окончательный результат получается, если заметить, что обратная матрица $\Lambda^{-1} = Q'A^{-1}Q$, и следовательно, квадратичная форма $\vec{u}'\Lambda^{-1}\vec{u} = \vec{t}'QQ'A^{-1}QQ'\vec{t} = \vec{t}'A^{-1}\vec{t}$.

⊠

ЛИТЕРАТУРА

Усердный в службе не должен
бояться своего незнания; ибо
каждое новое дело он прочтет.

- [1] Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерности. – М.: „Финансы и статистика“, 1989. – 607 с.
- [2] Андерсон Т. Введение в многомерный статистический анализ. – М.: „Физматгиз“, 1963. – 500 с.
- [3] Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы. – М.: „Финансы и статистика“, 1998. – 350 с.
- [4] Кендалл М.Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. – М.: „Наука“, 1976. – 736 с.
- [5] Крамер Г. Математические методы статистики. – М.: „Мир“, 1975. – 648 с.
- [6] Окунь Я. Факторный анализ. – М.: „Статистика“, 1974. – 200 с.
- [7] Себер Дж. Линейный регрессионный анализ. – М.: „Мир“, 1980. – 456 с.
- [8] Симушкин С.В. Дисперсионный анализ. Ч.1, Ч.2. – Казань.: „Издательство КГУ“, 1998.
- [9] Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. – М.: „Финансы и статистика“, 1995. – 384 с.
- [10] Шеффе Г. Дисперсионный анализ. – М.: „Наука“, 1980. – 512 с.

Предметный указатель

- Асимптотическое распределение функций 93
- Варимакса принцип 70
- Вращение факторов 59, 69
- Главные компоненты 44
 - информативность 47, 48
- Детерминации коэффициент
 - линейный 11
 - нелинейный 26
- Дисперсия обобщенная 22
- Ковариаций матрица 90
 - вырожденная 91
- Корреляции коэффициент
 - канонический 17–19
 - ложный 16
 - множественный 10, 11
 - полный 90
 - скрытый 16
 - частный 12, 13
- Корреляционное отношение 26
- Латентный признак 57
- Модель факторная 57, 58
- Нагрузка
 - общая 57
 - характерная 57
- Нормальное распределение
 - квадратичные формы 77
 - моменты 72
 - независимость 76
 - плотность 72
 - условные распределения 78, 79
 - характеристическая функция 73
- Общность 58
- Остаточная дисперсия 8, 9
- Ошибка линейной регрессии 8
- Регрессия среднеквадратическая
 - линейная
 - — определение 5
 - — формула
 - — — для 1-й компоненты 6
 - — — для всех компонент 7
 - — — для коэффициентов 7
 - нелинейная
 - — определение 25
 - — формула 25
 - ортогональная 24
- Фактор
 - общий 57
 - характерный 57
- Фундаментальная теорема факторного анализа 59
- Характерность 58
- Эллипсоид рассеяния
 - определение 21
 - формула 21
 - объем 95