

# Факторный анализ

- **Главная цель** – уменьшение размерности исходных данных для их экономного описания при условии минимальных потерь исходной информации.
- **Результат** – переход от множества исходных переменных к существенно меньшему числу новых переменных – факторов.

- **Фактор** – причина совместной изменчивости нескольких исходных переменных.
- Основное назначение факторного анализа – анализ корреляций множества признаков.

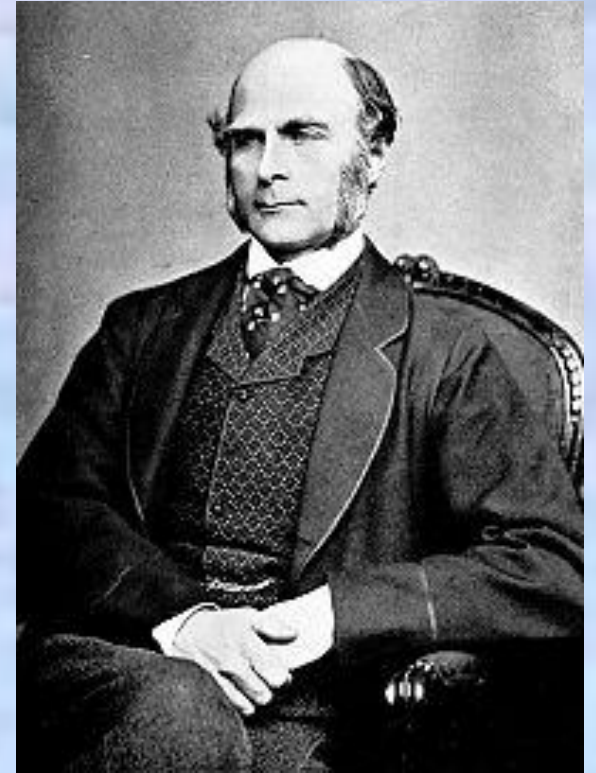
# Основные схемы факторного анализа

- **Эксплораторный факторный анализ** осуществляется при исследовании скрытой факторной структуры без предположения о числе факторов и их нагрузках.
- **Конфирматорный факторный анализ** предназначен для проверки гипотез о числе факторов и их нагрузках.



# История создания метода

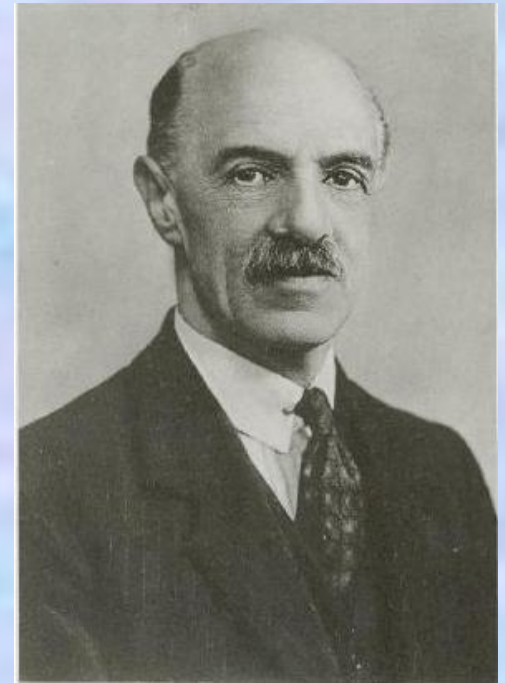
- **Ф. Гальтон:** если несколько признаков, измеренных на группе индивидов, изменяются согласованно, то можно предположить существование одной общей причины этой изменчивости – фактора.



- **К. Пирсон (1901):** создание одого из основных способов уменьшения размерности данных – метода главных компонент.



- **Ч. Спирмен (1904):**  
разработка  
математического аппарата  
для оценки фактора, исходя  
из множества измерений, –  
однофакторного анализа.  
Если ряд признаков попарно  
коррелируют друг с другом,  
можно составить систему  
линейных уравнений,  
связывающих эти признаки.





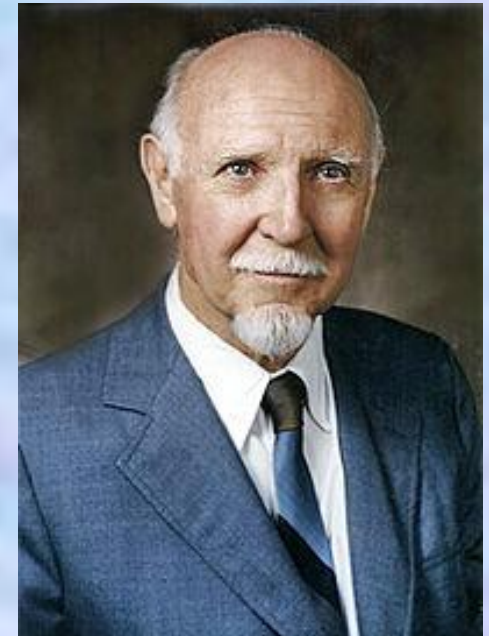
- **Л. Терстоун (1930-е):** многофакторный анализ для описания многочисленных измеренных способностей меньшим числом общих факторов интеллекта – линейной комбинацией исходных способностей.



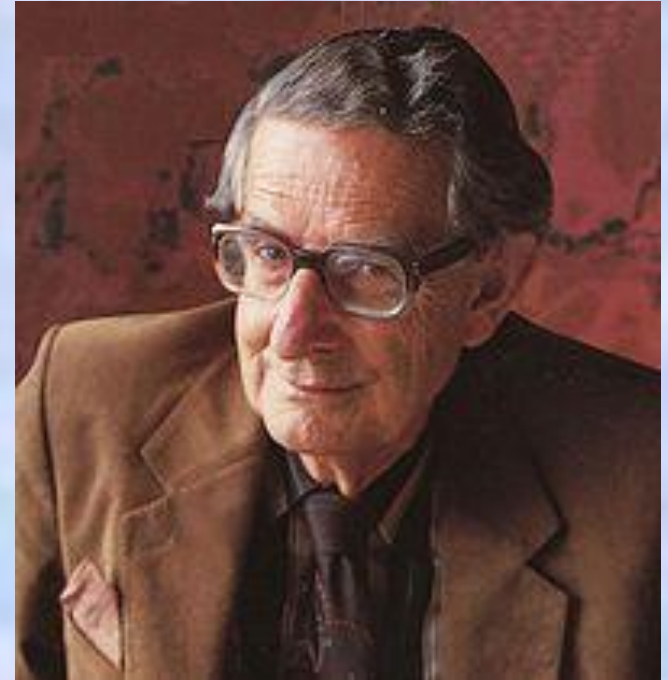


- С 1950-х годов: факторный анализ используется при разработке тестов, обосновании структурных теорий интеллекта и личности.

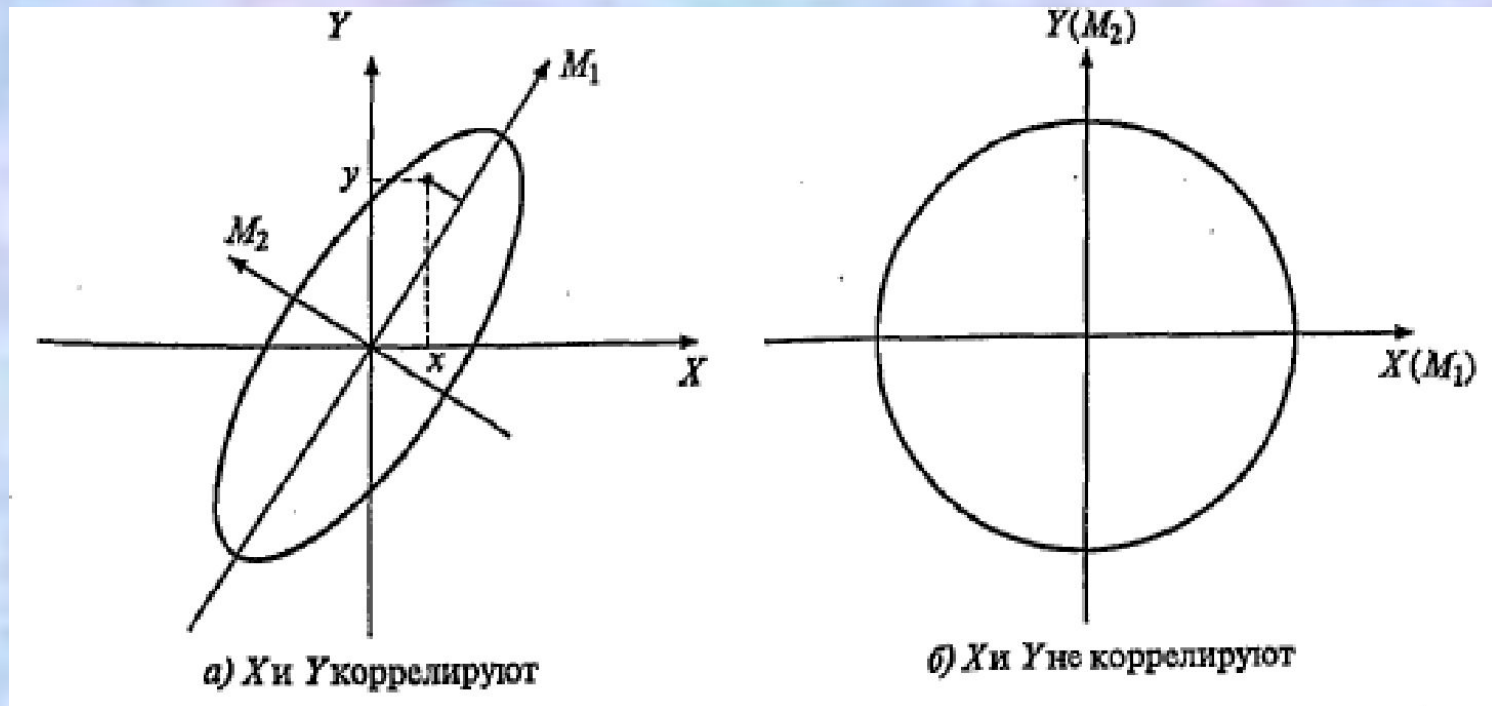
- **Р. Кэттелл.** 4500 наименований личностных особенностей □ 187 вопросов теста □ 16 факторов (свойств личности). Опросник 16PF.



- **Г.Ю. Айзенк:**  
трехфакторная теория личности: психотизм, экстраверсия, нейротизм.
- Опросники: EPI, EPQ.



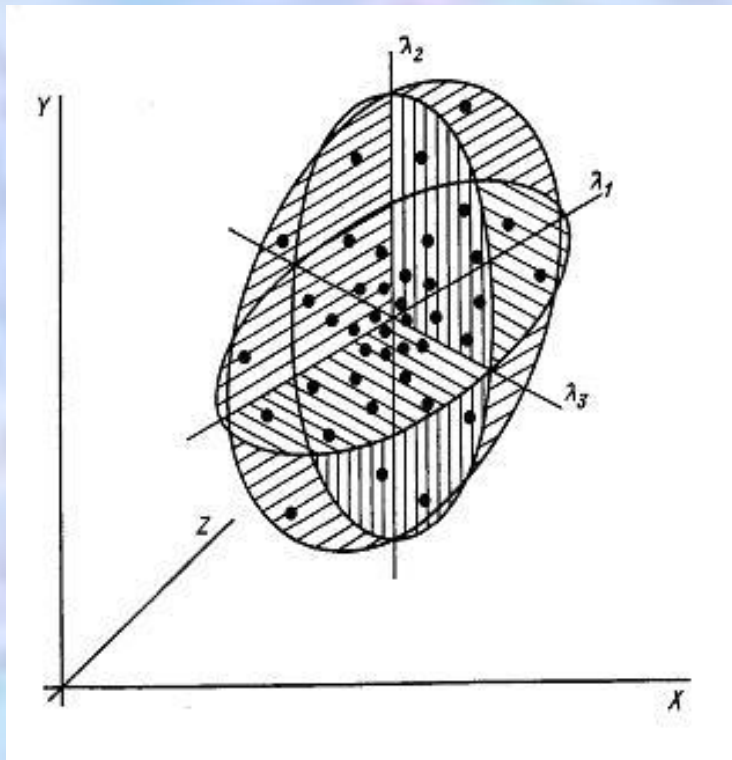
# Метод главных компонент



- Переменные  $X$  и  $Y$  положительно коррелируют. Наибольший разброс данных вдоль оси  $M_1$ . Анализ главных компонент – переход от  $XY$  к проекции их координат на  $M_1$ .

- Анализ главных компонент – преобразование информации, содержащейся в исходных данных.
- Теряется минимум информации об объектах измерения.
- Чем сильнее взаимосвязь переменных, тем меньше исходной информации теряется.
- Если переменные не коррелируют, то компоненты являются равнозначными и невозможно определить главную из них.





- Три переменные:  
X, Y, Z.

- **Система координат главных осей:**
- первой главной осью  $\lambda_1$  является самый длинный диаметр овального тела,
- второй главной осью  $\lambda_2$  является самый длинный диаметр в плоскости, ортогональной к первой главной оси и проходящей через центр тяжести системы (заштрихована вертикально),
- третья главная ось  $\lambda_3$  в трехмерном случае перпендикулярна к первой и второй главным осям и проходит через центр тяжести.

# Стадии факторного анализа:

1. Вычисление корреляционной матрицы для всех переменных, участвующих в анализе.
2. Извлечение факторов.
3. Вращение факторов для создания упрощенной структуры.
4. Интерпретация факторов.

Матрица корреляций пяти показателей интеллекта

№	Показатели	1	2	3	4	5
1	Счет в уме	1,00	0,88	0,33	0,23	0,42
2	Числовые ряды	0,88	1,00	0,32	0,24	0,35
3	Осведомленность	0,33	0,32	1,00	0,58	0,58
4	Словарный запас	0,23	0,24	0,58	1,00	0,54
5	Сходство	0,42	0,35	0,58	0,54	1,00

- 50 испытуемых, 5 показателей интеллекта.
- Все показатели статистически взаимосвязаны, кроме 4 – 1 и 4 – 2.

**Факторная структура пяти показателей способностей  
(метод главных осей, до вращения)**

Переменные	Факторные нагрузки $a_{jk}$		Общность $h_j^2$
	$F_1$	$F_2$	
1	0,807	-0,482	0,88
2	0,774	-0,481	0,83
3	0,661	0,416	0,61
4	0,580	0,470	0,56
5	0,675	0,317	0,56
Собственное значение $\lambda_k$	2,478	0,959	3,44
Доля дисперсии	0,496	0,192	0,69

- Выделены два фактора. Находятся переменные, имеющие наибольшие факторные нагрузки (корреляции между исходной переменной и фактором).



- Чем больше абсолютная величина факторной нагрузки, тем сильнее связь переменной с фактором и тем больше данная переменная обусловлена действием соответствующего фактора.
- Факторы идентифицируются по переменным, с которыми они в наибольшей степени связаны. Фактору присваивается имя, обобщающее по смыслу наименования входящих в него переменных.

# Проблемы факторного анализа

- **1. Проблема общности**
- Общность – часть дисперсии переменной, обусловленная действием общих факторов.
- Вычисляется как сумма квадратов нагрузок по строке:

$$h_i^2 = \sum_{k=1}^M a_{ik}^2$$

- **Полнота факторизации** – сумма квадратов элементов факторной структуры (суммарная дисперсия всех переменных), деленная на количество признаков.

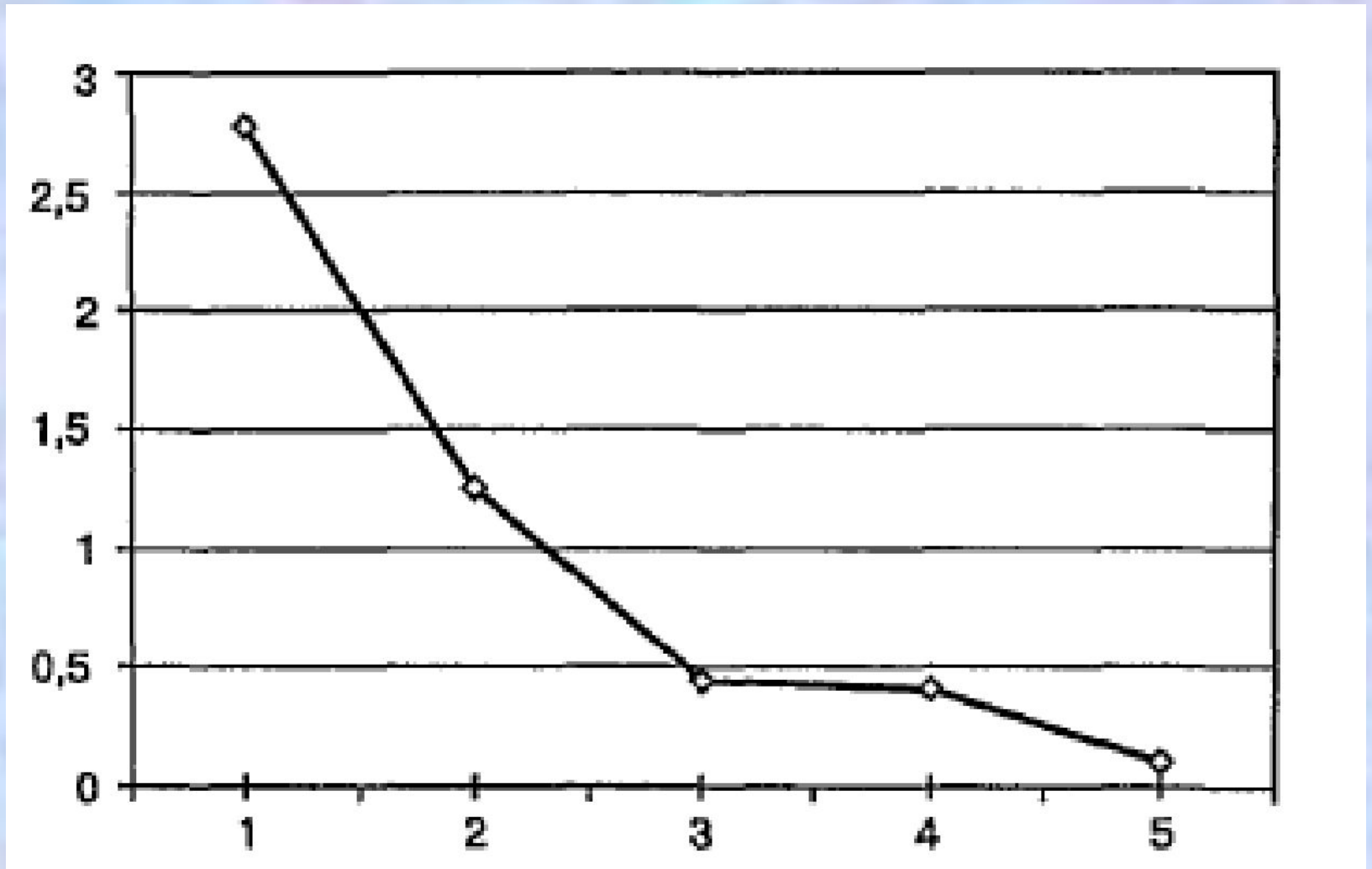
$$V = \sum_{k=1}^M V_k = \frac{1}{P} \sum_{k=1}^M \lambda_k = \frac{1}{P} \sum_{i=1}^P h_i^2 = \frac{1}{P} \sum_{k=1}^M \sum_{i=1}^P a_{ik}^2$$

- Качество факторного анализа тем выше, чем выше полнота факторизации.
- Нижним порогом этого показателя обычно выбирается значение 0,7

- **2. Проблема числа факторов**

- Заранее не известно, сколько факторов достаточно для представления переменных.
- Исследователь заранее определяет их возможное количество:
- **Критерий Кайзера:** число факторов равно числу компонент, собственные значения которых больше 1.
- Критерий отсеивания **Р. Кеттелла** (критерий каменной осыпи): количество факторов определяется по точке перегиба на графике собственных значений ( $K \pm 1$ ).



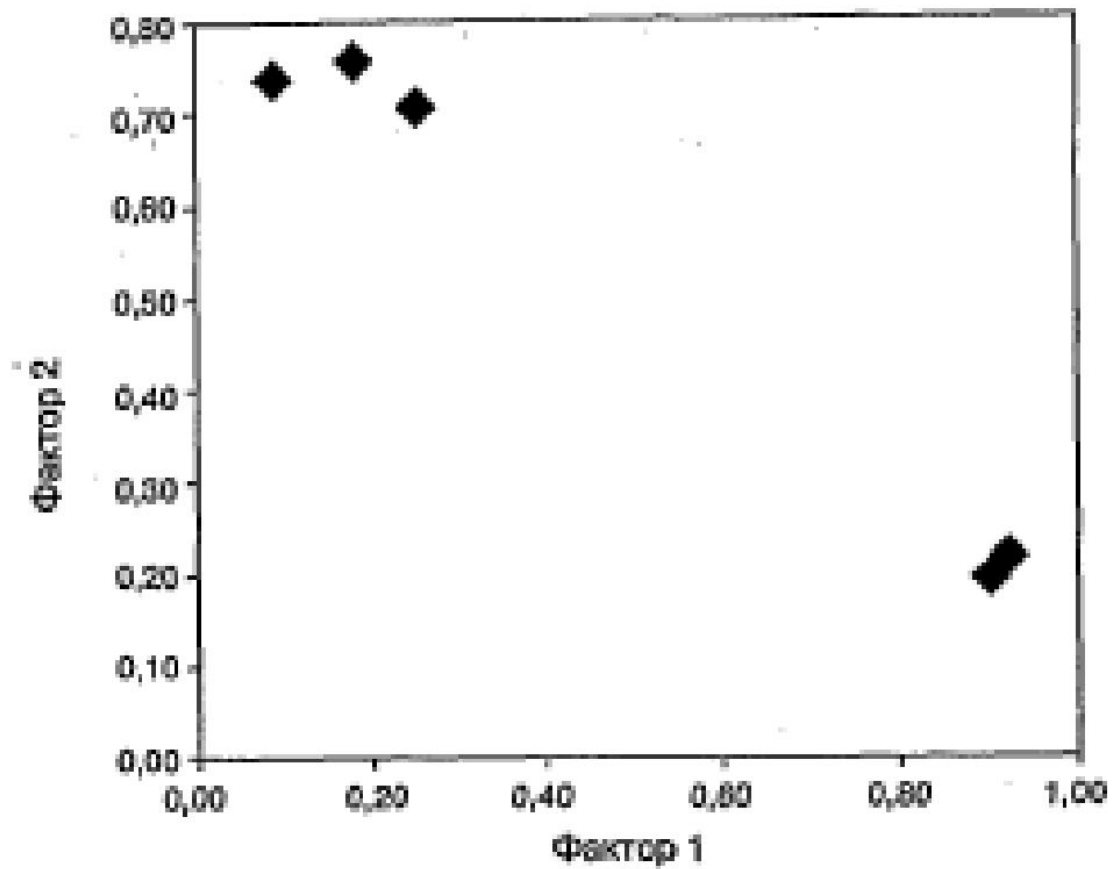


- По критерию Кайзера – 2 фактора, по критерию Кеттелла – 3-4 фактора.

- **3. Проблема вращения**
- Результаты факторного анализа непосредственно не подлежат интерпретации.



- Оси факторов можно повернуть на любой угол относительно переменных, при условии соблюдения взаимной перпендикулярности факторов.
- Желательно, чтобы каждая переменная в результате вращения оказалась вблизи оси фактора, т.е. имела максимальную нагрузку по одному фактору и минимальную – по другим.
- **Варимакс-вращение**



Факторная структура пяти показателей интеллекта после варимакс-вращения



**Факторная структура пяти показателей интеллекта  
после варимакс-вращения**

Переменные	$F_1$	$F_2$	$R^2$
1	0,921	0,217	0,88
2	0,903	0,193	0,83
3	0,183	0,757	0,61
4	0,088	0,739	0,56
5	0,260	0,700	0,56
СКН*	1,773	1,694	3,44
Доля дисперсии	0,355	0,339	0,69

\* СКН — сумма квадратов факторных нагрузок<sup>1</sup>

- **4. Проблема интерпретации факторов**
- Каждый фактор идентифицируется по тем переменным, которые с ним наиболее сильно связаны, то есть которые имеют наиболее высокие факторные нагрузки по данному фактору.