

МЕТОДИЧЕСКИЕ ВОПРОСЫ ИСПОЛЬЗОВАНИЯ ФАКТОРНОГО АНАЛИЗА НА ПРИМЕРЕ СПЕКТРАЛЬНЫХ ПОКАЗАТЕЛЕЙ СЕРДЕЧНОГО РИТМА

*МАШИН В. А., Нововоронежский учебно-тренировочный центр подготовки специалистов для АЭС,
Нововоронеж*

Настоящая статья посвящена рассмотрению ряда методических вопросов, касающихся использования факторного анализа для оценки спектральных показателей сердечного ритма, а именно: допущения многомерности нормального распределения исследуемых признаков, целесообразности применения факторной модели анализа к исходным данным, процедуры отбора общих факторов, значимости полученного факторного решения. В проведенном исследовании были использованы переменные, характеризовавшиеся нарушением нормального распределения, переменные, преобразованные к нормальному распределению, а также матрицы корреляции Пирсона и Спирмена, модельные данные. В процессе проведенного факторного анализа (метод главных компонент, Varimax-вращение) нами проанализированы как возможности, так и ограничения различных статистических критериев и процедур; рассмотрена природа выделенных общих факторов и возможность использования корреляций Спирмена при нарушениях нормальности распределения.

Ключевые слова: факторный анализ, спектр сердечного ритма, нормальное распределение, статистические критерии согласия, корреляция Пирсона, корреляция Спирмена.

Факторный анализ относится к методам многомерного статистического анализа и широко используется для снижения размерности анализируемого пространства признаков, отбора наиболее информативных показателей и классификации объектов. В основе факторного анализа лежит общая базовая идея, согласно которой структура связей между многочисленными анализируемыми переменными может быть объяснена их зависимостью от меньшего числа непосредственно неизмеряемых («скрытых», «латентных», «гипотетических») общих факторов (Айвазян и др., 1989). Основной предпосылкой для итогового снижения размерности пространства анализируемых признаков служит взаимная коррелированность исходных переменных, а исходными данными для анализа выступает матрица их парных коэффициентов корреляции (или ковариации).

Факторный анализ можно использовать не только как метод сжатия информации, но и как основу классификации, заменяя исходные признаки факторами; в данном случае классификация объектов заключается в получении факторных оценок для каждого из них и последующей визуальной или аналитической группировке этих объектов в факторном пространстве, что позволяет перейти от типологизации по большому числу исходных признаков к построению классификации на основе значительно меньшего числа факторов, с более ясным аналитическим описанием и графическим представлением данных.

Идея использования факторного анализа для классификации функциональных состояний человека на основе спектра сердечного ритма была реализована в ряде психофизиологических исследований Н. Н. Даниловой (Данилова, 1995, 1998; Данилова, Аста-

фьев, 1999). В этих исследованиях основу факторной модели составили три показателя спектра мощности сердечного ритма, которые должны были характеризовать дыхательные (0.12÷0.16 Гц), сосудистые (0.06÷0.10 Гц) и метаболические (0.02÷0.04 Гц) воздействия на сердечный ритм (Данилова, Астафьев, 1999). Поскольку границы частотных диапазонов указанных показателей могут сильно колебаться, для классификации функциональных состояний было предложено использовать метод факторного анализа. Для этого были рассчитаны показатели мощности спектра сердечного ритма в 25 частотных диапазонах, которые были сгруппированы в 10 переменных: 0.005÷0.02 и 0.02÷0.04 Гц (метаболическая модуляция); 0.04÷0.06, 0.06÷0.08 и 0.06÷0.1 Гц (сосудистая модуляция); 0.1÷0.12, 0.12÷0.14, 0.14÷0.16, 0.16÷0.3 и 0.3÷0.5 Гц (дыхательная модуляция) (Данилова, 1995). В соответствии с полученной матрицей взаимных корреляций 10 переменных далее было выделено три общих фактора, которые интерпретировались как метаболические, сосудистые и дыхательные модуляции сердечного ритма. Оценки по общим факторам служили основой для типологизации функциональных состояний обследуемых в факторном пространстве вегетативных влияний на сердечный ритм.

В своем исследовании мы решили рассмотреть различные методические вопросы приложения факторного анализа к спектральным показателям сердечного ритма, включая допущение о многомерности нормального распределения исследуемых показателей, проверку целесообразности применения факторной модели анализа к исходным данным, процедуры отбора общих факторов, оценку значимости полученного факторного решения. Отдельного внимания, с нашей точки зрения, заслуживает вопрос о возможности содержательной (физиологической) интерпретации природы полученных общих факторов. Отметим, что спектральные показатели сердечного ритма находят широкое применение в современных психофизиологических исследованиях.

Методика

В эксперименте приняли участие 734 оператора АЭС, которые проходили обследования в Лаборатории психофизиологического обеспечения Нововоронежского учебно-тренировочного центра подготовки специалистов для АЭС (ЛПФО НВУТЦ). У всех обследуемых отсутствовали нарушения здоровья (мужчины, средний возраст 29.87 лет, стандартное отклонение 7.17 лет). Регистрация сердечного ритма выполнялась в спокойном состоянии в положении сидя в течение 10 минут.

Для записи QRS-комплексов нормальных синусовых кардиоциклов ЭКГ и последующего выделения из них R-R интервалов (в миллисекундах) использовался 3-канальный программно-аппаратурный комплекс «Варикард-1.51» (частота дискретизации 500 Гц). Хранение, редактирование R-R интервалов (коррекция артефактов, экстрасистол на ритмограмме) и расчет показателей спектра мощности сердечного ритма выполнялись с применением компьютерной программы «MABP.DBase-HRV», разработанной в ЛПФО НВУТЦ («Delphi-5», «Access-2000»). Перед расчетом спектральных показателей из временного ряда сердечного ритма удалялся линейный тренд, рассчитанный с помощью метода наименьших квадратов. Для соблюдения требования об эквидистантности временного ряда (равные промежутки времени измерения переменной) границы частотных диапазонов спектра кардиоинтервалов корректировались по следующей формуле:

$FD_{cor} = FD \times M_{NN}$, где FD – исходная граница частотного диапазона, FD_{cor} – скорректированная, M_{NN} – среднее значение продолжительности R-R интервалов в выборке (в се-

кундах) (Машин, 2002). Мощность спектральной плотности сердечного ритма (в мсек²) рассчитывалась с помощью быстрого преобразования Фурье со спектральным окном Хэмминга по стационарным «скользящим» выборкам объемом 256 и шагом 10 R-R интервалов. Для проверки выборок на стационарность применялся непараметрический метод Вьялда-Вольфовитца (Зырянов и др., 1990). Весь спектр (0.0÷0.5 Гц) был поделен на 25 частотных областей $D_1 \div D_{25}$ (длительностью 0.02 Гц), по которым были получены усредненные по каждому обследуемому показатели мощности спектральной плотности сердечного ритма (объем выборки $n = 734$), которые послужили исходными данными для проведения факторного анализа.

Для оценки нормальности распределения данных, расчета матриц корреляций и выполнения факторного анализа использовалась статистическая программа «Statistica 6.0». Расчет статистических критериев для проверки целесообразности применения факторной модели анализа, значимости полученного факторного решения и конгруэнтности различных факторных структур был выполнен с помощью специально разработанной программы «МABP-FA».

Результаты исследования и их обсуждение

Перед началом выполнения факторного анализа рассмотрим допущение о многомерном нормальном распределении исходных переменных. Ряд авторов утверждают, что для построения линейных факторных моделей переменные должны быть многомерно нормально распределены (Харман, 1972; Rencher, 2002). Из этого как минимум следует, что каждый из исходных показателей мощности спектра сердечного ритма должен быть распределен по нормальному закону и между исследуемыми показателями должны существовать линейные связи. Существует также точка зрения, что предположение о многомерной нормальности может играть существенную роль лишь на стадии проверки адекватности факторных моделей, поскольку статистические критерии согласия чувствительны к этому допущению (Справочник по прикладной статистике, 1990). Наконец, ряд исследователей полагают, что факторная модель не требует выполнения предположения о многомерной нормальности закона распределения параметров, хотя отмечают, что последствия нарушения этого допущения до сих пор остаются не вполне ясными (Ким и др., 1989).

Визуальный анализ гистограмм полученных нами данных указывал на выраженные нарушения нормального распределения показателей спектра в 25 частотных областях. Статистический анализ дополнительно позволил установить высокие положительные значения коэффициентов асимметрии (2.28÷10.54) и эксцесса (7.42÷165.69). Напомним, что для нормального распределения коэффициент асимметрии равен 0, а коэффициент эксцесса 3. Для преобразования распределений показателей к нормальному виду были использованы формулы, которые приведены в табл. 1. Гипотеза о нормальности распределений показателей проверялась с помощью критерия Шапиро-Уилка. Поскольку для преобразованных распределений частотных диапазонов $D_{16} \div D_{25}$ нарушения нормальности сохранялись ($p < 0.004$), они были дополнительно сгруппированы в две переменные: $T_1 = D_{16} + D_{17} + D_{18} + D_{19} + D_{20}$ (0.30÷0.40 Гц) и $T_2 = D_{21} + D_{22} + D_{23} + D_{24} + D_{25}$ (0.40÷0.50 Гц). После этого мы получили 17 переменных ($V_1 \div V_{17}$), для которых, согласно критерию Шапиро-Уилка (табл. 1), принималась гипотеза о нормальности распределения ($p = 0.062 \div 0.552$).

Таблица 1. Формулы приведения переменных к нормальному виду (\ln – натуральный логарифм) и значения уровня значимости критерия нормальности распределения Шапиро-Уилка (p).

Переменная	Частотный диапазон (Гц)	p
$V_1 = \ln(D_1)$	0.00÷0.02	0.383
$V_2 = \ln(D_2)$	0.02÷0.04	0.261
$V_3 = \ln(D_3)$	0.04÷0.06	0.476
$V_4 = \ln(D_4)$	0.06÷0.08	0.285
$V_5 = \ln(D_5+1)$	0.08÷0.10	0.105
$V_6 = \ln(D_6+2)$	0.10÷0.12	0.106
$V_7 = \ln(D_7+1)$	0.12÷0.14	0.140
$V_8 = \ln(D_8)$	0.14÷0.16	0.552
$V_9 = \ln(D_9)$	0.16÷0.18	0.243
$V_{10} = \ln(D_{10})$	0.18÷0.20	0.175
$V_{11} = \ln(D_{11})$	0.20÷0.22	0.253
$V_{12} = \ln(D_{12}+0.5)$	0.22÷0.24	0.167
$V_{13} = \ln(D_{13}+0.48)$	0.24÷0.26	0.073
$V_{14} = \ln(D_{14}+0.5)$	0.26÷0.28	0.273
$V_{15} = \ln(D_{15}+0.3)$	0.28÷0.30	0.110
$V_{16} = \ln(T_1+0.5)^*$	0.30÷0.40	0.173
$V_{17} = \ln(T_2-0.5)^{**}$	0.40÷0.50	0.062

Примеч.: $T_1 = D_{16}+D_{17}+D_{18}+D_{19}+D_{20}$,
 $T_2 = D_{21}+D_{22}+D_{23}+D_{24}+D_{25}$.

Для проверки линейной связи между переменными использовались диаграммы рассеивания. Поскольку распределения переменных отвечали нормальному закону (Кендалл, Стьюарт, 1973), мы могли оценить значимость парных коэффициентов корреляции Пирсона, которые служат мерой линейной зависимости между переменными (Хальд, 1956). Для этого рассчитывался критерий:

$$t = \frac{r}{\sqrt{1-r^2}} \times \sqrt{n-2}, \text{ где } r - \text{ коэффициент корреляции, } n - \text{ объем выборки (Справочник по прикладной статистике, 1990).}$$

Величина t -критерия (распределенного по закону Стьюдента с $n-2$ степенями свободы, df) изменялась в диапазоне 8.044÷63.468, что позволяло принять гипотезу о значимости коэффициентов корреляции на уровне $p = 0.000$. Но поскольку значимость t -критерия могла быть достигнута также за счет большой выборки данных, то для подтверждения вывода о линейной связи между переменными мы дополнительно использовали визуальный анализ диаграмм рассеивания.

Для проверки последствий нарушения допущения о многомерном нормальном распределении исходных данных мы также включили в анализ не преобразованные к нормаль-

ному виду переменные $D_1 \div D_{15}$, T_1 и T_2 . Это позволило сформировать для факторного анализа две матрицы параметрических коэффициентов корреляции Пирсона: матрицу **A** (парные корреляции преобразованных переменных $V_1 \div V_{17}$) и матрицу **B** (парные корреляции непреобразованных переменных $D_1 \div D_{15}$, T_1 , T_2). Кроме того, была рассчитана матрица непараметрических коэффициентов корреляции Спирмена для непреобразованных переменных $D_1 \div D_{15}$, T_1 , T_2 (матрица **C**). Для всех трех матриц корреляций (порядок матриц 17×17) был выполнен факторный анализ.

Перед выполнением факторного анализа по корреляционным матрицам необходимо провести проверку целесообразности использования факторной модели для исследования показателей спектра сердечного ритма. Итак, поскольку факторный анализ относится к той ветви многомерного статистического анализа, которая исследует внутреннюю структуру матриц корреляций (Лоули, Максвелл, 1967), в самом начале необходимо убедиться в наличии значимых корреляционных зависимостей между переменными с помощью критерия сферичности Бартлетта-Уилкса и критерия адекватности выборки Кайзера (*measure of sampling adequacy – MSA*). Ранее мы проверили значимость каждого парного коэффициента корреляции матрицы **A** с помощью t -критерия. Для оценки значимости всей корреляционной матрицы Бартлетт предложил критерий, усовершенствованный Уилксом (Rencher, 2002; Иберла, 1980):

$$\chi^2 = - \left[n - \frac{1}{6}(2m + 5) \right] \times \ln |R|, \text{ где } n - \text{объем выборки, } m - \text{число переменных, } \ln - \text{натуральный логарифм, } |R| - \text{определитель матрицы, } df = m(m-1)/2.$$

С помощью критерия Бартлетта-Уилкса проверяется гипотеза о единичной матрице, в которой все элементы главной диагонали равны 1, а все остальные – 0. Данный критерий основан на преобразовании определителя матрицы в статистику χ^2 , при этом если корреляция между переменными будет слабой, то значения $|R|$ будут близки к 1 (снижение χ^2), а если 2 или более переменных будут значительно коррелировать друг с другом, то значения $|R|$ будут стремиться к 0 (рост χ^2).

Для наших данных были получены следующие значения критерия Бартлетта-Уилкса:

$\chi_A^2 = 18162.10$ ($|R| = 1.390e-11$), $\chi_A^2 = 10757.91$ ($|R| = 3.707e-07$), $\chi_N^2 = 17452.58$ ($|R| = 3.690e-11$), $p = 0.000$; следовательно, подтверждается гипотеза о значимости всей корреляционной матрицы (переменные зависимы). Данный вывод для матрицы **A** вытекал уже из результатов использования t -критерия (высокая значимость коэффициентов корреляции по данному критерию), поэтому ряд авторов рекомендуют использовать критерий Бартлетта-Уилкса лишь в случае слабой корреляционной зависимости между переменными (Иберла, 1980). Если мы разделим все значения парных корреляций матрицы **A** на 100 (все $r < 0.01$), то значения $\chi^2 = 3.32$ ($|R| = 0.995$), $p = 1.0$ и гипотеза о зависимости переменных отвергается. Дополнительно добавим в матрицу один значимый коэффициент корреляции: $r_{ij} = r_{ji} = 0.465$: в данном случае $\chi^2 = 180.33$ ($|R| = 0.780$), и гипотеза о зависимости переменных принимается на уровне значимости $p = 0.007$. Также отметим зависимость критерия Бартлетта-Уилкса от объема выборки. Разделим значения парных корреляций матрицы **A** на 10 (все $r < 0.1$). В этом случае $\chi^2 = 237.60$ ($|R| = 0.721$), $p = 0.000$ ($n = 734$), – принимается гипотеза о зависимости переменных. Если объем выборки был бы равен, например, 450, то тогда $\chi^2 = 144.72$, $p = 0.288$, – отвергается гипотеза о зависимости переменных.

Критерий адекватности выборки Кайзера не зависит от объема выборки, но для него отсутствуют статистические таблицы значимости (Rencher, 2002; Ким и др., 1989):

$$MSA = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} q_{ij}^2}, \text{ где } \sum_{i \neq j} r_{ij}^2 \text{ – сумма квадратов элементов матрицы коэффициентов корреляции (исключая главную диагональ), } \sum_{i \neq j} q_{ij}^2 \text{ – сумма квадратов соответствующих частных коэффициентов корреляции.}$$

Критерий адекватности выборки сравнивает значения наблюдаемых коэффициентов корреляции со значениями частных коэффициентов корреляции. Частные корреляции оценивают взаимозависимости между переменными i и j при фиксированных значениях всех остальных переменных. Если корреляция между двумя переменными уменьшается, когда мы фиксируем значения других переменных, то это означает, что их взаимозависимость возникает частично через воздействие этих переменных; если же частная корреляция равна 0 или очень мала, то можно сделать вывод, что их взаимозависимость целиком обусловлена этим воздействием. В том случае, когда частная корреляция больше первоначальной корреляции между двумя переменными, можно заключить, что другие переменные ослабляли связь («маскировали» корреляцию) (Кендалл, Стьюарт, 1973). Небольшие значения критерия MSA указывают на то, что корреляции между парами переменных нельзя объяснить другими переменными и что использование факторного анализа нецелесообразно. Пороговые значения для критерия MSA по Кайзеру следующие (Ким и др., 1989): > 0.9 – безусловная адекватность; > 0.8 – высокая адекватность; > 0.7 – приемлемая адекватность; > 0.6 – удовлетворительная адекватность; > 0.5 – низкая адекватность; $\leq 0,5$ – факторный анализ неприменим к выборке. Для наших данных были получены следующие значения критерия: $MSA_A = 0.898$, $MSA_B = 0.786$, $MSA_C = 0.894$. Таким образом, для матриц **A** и **C** по критерию MSA был достигнут высокий уровень адекватности применения факторного анализа, а для матрицы **B** – приемлемый.

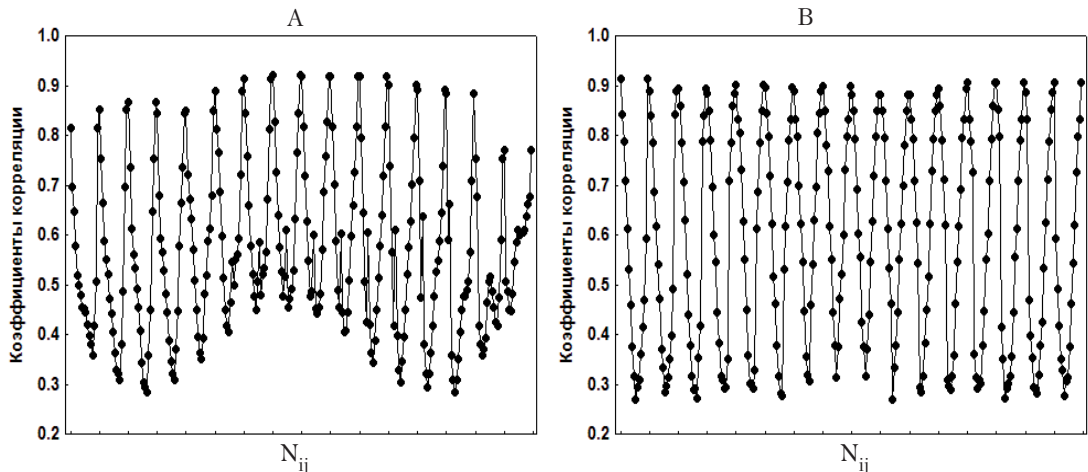


Рис. 1. Графики парных коэффициентов корреляции r_{ij} : (А) – матрица **A**, (В) – матрица **М** (последовательность векторов-строк)

Дополнительно мы решили рассмотреть графики парных коэффициентов корреляции, представив каждую матрицу в виде вектора, состоящего из последовательности строк (значения на главной диагонали исключались). На рис. 1, А (матрица **A**) хорошо

просматривается периодичность в изменениях корреляции: максимумы образованы значениями парных коэффициентов корреляции показателей спектра сердечного ритма соседних частотных диапазонов ($r_{m-1,m}$ и $r_{m,m+1}$), которые постепенно уменьшаются по мере увеличения расстояния между частотными диапазонами. Характерная высокая взаимозависимость между соседними частотными диапазонами становится еще более очевидной при анализе частных коэффициентов корреляции. На рис. 2, А (матрица **A**) отчетливо наблюдаются пики, образованные частными корреляциями показателей спектра соседних частотных диапазонов, при этом частные корреляции показателей спектра более удаленных частотных диапазонов колеблются около 0. (Заметим, что высокие значения 32 частных корреляций соседних частотных диапазонов, когда оставшиеся 104 близки 0, незначительно понижают показатель MSA.) Такая периодичность парных и частных корреляций наблюдается и для матриц **B** и **C**. Это позволило нам смоделировать с помощью программной среды «Delphi-5» матрицу корреляций, используя следующую формулу гармонических колебаний:

$X_i = A_m \times \cos(2 \times \pi \times f) + A_m \times \sin(2 \times \pi \times f) + X_{cp} + \text{RandG}(M, SD)$, где A_m – амплитуда колебаний = 0.22, f – частота колебаний (с периодом = 22), X_{cp} – среднее значение коэффициента корреляций = 0.599, RandG – функция для генерирования случайных чисел, распределенных по нормальному закону Гаусса с математическим ожиданием M и среднеквадратическим отклонением SD ($M = 0$, $SD = 0.12$).

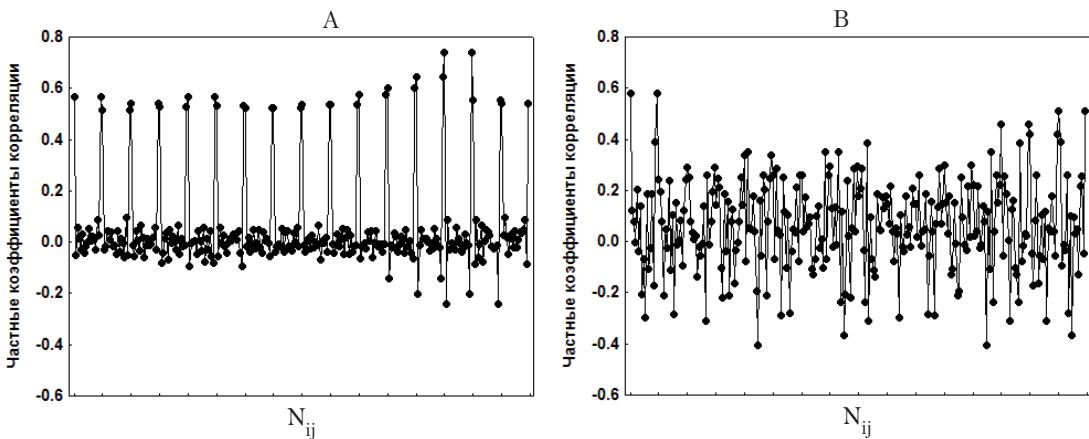


Рис. 2. Графики частных коэффициентов корреляции r_{ij} : (А) – матрица **A**, (В) – матрица **М** (последовательность векторов-строк)

Из полученной последовательности гармонических колебаний отбирались значения векторов-строк ($m = 17$, включая $r_{ij} = 1.0$) для модельной матрицы: максимальная амплитуда колебаний соответствовала значениям корреляций r_{ij} показателей спектра соседних частотных диапазонов. Таким образом, были отобраны элементы для 100 матриц, средние значения по которым стали элементами модельной матрицы **М**. Коэффициент корреляции между элементами матриц **A** и **М** оказался равен 0.882, между соответствующими частными корреляциями $r = 0.433$. На рис. 1, В и 2, В представлены графики парных и частных коэффициентов корреляции матрицы **М** (последовательность векторов-строк). Для матрицы **М** критерий Бартлетта-Уилкса $\chi^2 = 21867.75$ ($|R| = 8.466e-14$), $p = 0.000$, а $MSA = 0.914$. В своем исследовании мы решили проверить, насколько близки будут

факторные модели для матриц **A** и **M**, имеющих близкое распределение парных корреляций и значимые отличия в распределении частных корреляций переменных.

После подтверждения гипотезы о целесообразности применения факторной модели для анализа данных по четырем матрицам (критерии Бартлетта-Уилкса и MSA) на следующем шаге для выделения факторов (компонент) был использован метод главных компонент (*Principal components*). Модель компонентного анализа имеет следующий вид (Харман, 1972):

$z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m$ ($j = 1, 2, \dots, m$), где z_j – наблюдаемый признак (переменная), который линейно зависит от m некоррелированных между собой и упорядоченных по величине вклада в суммарную дисперсию признаков компонент F_1, F_2, \dots, F_m ; m – число анализируемых признаков. При этом первая компонента дает максимально возможный вклад в суммарную дисперсию признаков, а последняя – минимальный. Главные компоненты независимы (парные корреляции равны 0), т. е. в геометрическом плане ортогональны. Модель компонентного анализа позволяет однозначно восстановить значения каждого из наблюдаемых признаков z_j по соответствующим значениям компонент F_1, F_2, \dots, F_m . Коэффициенты при компонентах a_{ji} называют нагрузками (*loadings*), они представляют собой линейную корреляцию между переменными и компонентами и выражают меру влияния компонента на признак.

Обычно суммарная дисперсия признаков раскладывается по главным компонентам таким образом, что первые несколько компонент k уже объясняют почти всю эту дисперсию, а остальные практически ничего не добавляют, поэтому совсем не обязательно выделять все m компонент. На практике ограничиваются несколькими первыми k , так как их оказывается достаточно для удовлетворительного сжатого описания всей исходной информации. С целью определения относительной значимости каждой главной компоненты F_i рассчитываются их собственные значения (λ_i^2); собственное значение (*eigenvalue*) – это часть суммарной по всем переменным дисперсии, которую объясняет компонента. Для вычисления собственного значения нужно суммировать квадраты нагрузок a_{ji} по всем переменным данной компоненты F_i :

$$\lambda_i^2 = \sum_{j=1}^m a_{ji}^2 \quad (i = 1, 2, \dots, m), \text{ где } m \text{ – число переменных.}$$

В табл. 2 приведены собственные значения главных компонент, их доли в общей дисперсии и соответствующие кумуляты для корреляционных матриц **A**, **B**, **C** и **M**.

Таблица 2. Собственные значения главных компонент для корреляционных матриц **A**, **B**, **C** и **M** (N – номер компоненты)

N	Собственные значения				Доля в общей дисперсии (%)				Кумулята собственных значений				Кумулята доли в общей дисперсии (%)			
	A	B	C	M	A	B	C	M	A	B	C	M	A	B	C	M
1	10.2	6.77	9.83	10.5	60.0	39.8	57.8	62.0	10.2	6.77	9.83	10.5	60.0	39.8	57.8	62.0
2	2.38	2.87	2.48	3.25	14.0	16.9	14.6	19.1	12.6	9.64	12.3	13.8	74.0	56.7	72.4	81.1
3	1.39	1.76	1.53	1.84	8.19	10.4	8.98	10.8	14.0	11.4	13.8	15.6	82.2	67.0	81.4	91.9

N	Собственные значения				Доля в общей дисперсии (%)				Кумулята собственных значений				Кумулята доли в общей дисперсии (%)			
	А	В	С	М	А	В	С	М	А	В	С	М	А	В	С	М
4	0.91	1.36	0.93	0.19	5.36	7.98	5.47	1.12	14.9	12.8	14.8	15.8	87.5	75.0	86.9	93.0
5	0.54	0.93	0.55	0.17	3.16	5.50	3.25	1.0	15.4	13.7	15.3	16.0	90.7	80.5	90.1	94.0
6	0.36	0.64	0.37	0.14	2.11	3.79	2.20	0.81	15.8	14.3	15.7	16.1	92.8	84.3	92.3	94.8
7	0.28	0.61	0.29	0.13	1.64	3.56	1.73	0.76	16.1	14.9	16.0	16.3	94.5	87.9	94.1	95.6
8	0.23	0.44	0.25	0.12	1.32	2.59	1.44	0.71	16.3	15.4	16.2	16.4	95.8	90.5	95.5	96.3
9	0.15	0.40	0.16	0.11	0.89	2.35	0.95	0.65	16.4	15.8	16.4	16.5	96.7	92.8	96.4	97.0
10	0.14	0.26	0.14	0.09	0.84	1.56	0.85	0.54	16.6	16.0	16.5	16.6	97.5	94.4	97.3	97.5
11	0.10	0.24	0.11	0.09	0.60	1.40	0.64	0.52	16.7	16.3	16.6	16.7	98.1	95.8	97.9	98.0
12	0.08	0.21	0.09	0.09	0.47	1.22	0.52	0.50	16.8	16.5	16.7	16.8	98.6	97.0	98.5	98.5
13	0.07	0.15	0.08	0.07	0.41	0.86	0.45	0.41	16.8	16.6	16.8	16.8	99.0	97.8	98.9	98.9
14	0.06	0.11	0.06	0.06	0.33	0.67	0.35	0.36	16.9	16.7	16.9	16.9	99.3	98.5	99.3	99.3
15	0.05	0.10	0.05	0.05	0.28	0.58	0.30	0.31	16.9	16.8	16.9	16.9	99.6	99.1	99.6	99.6
16	0.04	0.08	0.04	0.04	0.22	0.47	0.24	0.24	17.0	16.9	17.0	17.0	99.8	99.6	99.8	99.8
17	0.03	0.07	0.03	0.03	0.18	0.43	0.19	0.16	17.0	17.0	17.0	17.0	100	100	100	100

Существует несколько правил отбора необходимого числа первых k главных компонент, которые формулируются в терминах собственных чисел (Rencher, 2002; Ким и др., 1989):

1. Оставить число главных компонент, которые объясняют заданный процент общей дисперсии. Его можно определить, исключив компоненты, которые учитывают менее 5% суммарной дисперсии переменных. Согласно табл. 2, данное условие выполняется при $k = 4$ для матриц **A** и **C** (соответственно, их суммарная доля дисперсии 87.5% и 86.9%), $k = 5$ для матрицы **B** (80.5%) и $k = 3$ для матрицы **M** (91.9%). Это правило основывается на исключении компонент с малым вкладом в суммарную дисперсию, поскольку они не имеют практической ценности (Харман, 1972).

2. Используя график собственных значений главных компонент (график «каменистой осыпи» – «*scree graph*»), определить точку «излома»: переход от резкого уменьшения собственных значений к их постепенному (с небольшим углом наклона линии графика) убыванию с ростом номера компоненты (метод предложен Кэттеллом). Для отбора используют главные компоненты с собственными значениями до данной точки «излома» (Rencher, 2002). Согласно рис. 3, В (матрица **M**), данное условие выполняется для первых трех главных компонент ($k = 3$). На графике, представленном на рис. 3, А (матрица **A**), единственная точка «излома» отсутствует; вместо этого собственные значения плавно изменяются (данная картина наблюдается и для матрицы **C**, для матрицы **B** – в интервале 4÷6 компонент). Такая неоднозначность (включая несколько точек «излома») характерна для многих графиков собственных значений и предлагает исследователю руководствоваться субъективными критериями в содержательных трактовках и описаниях (Ким и др., 1989).

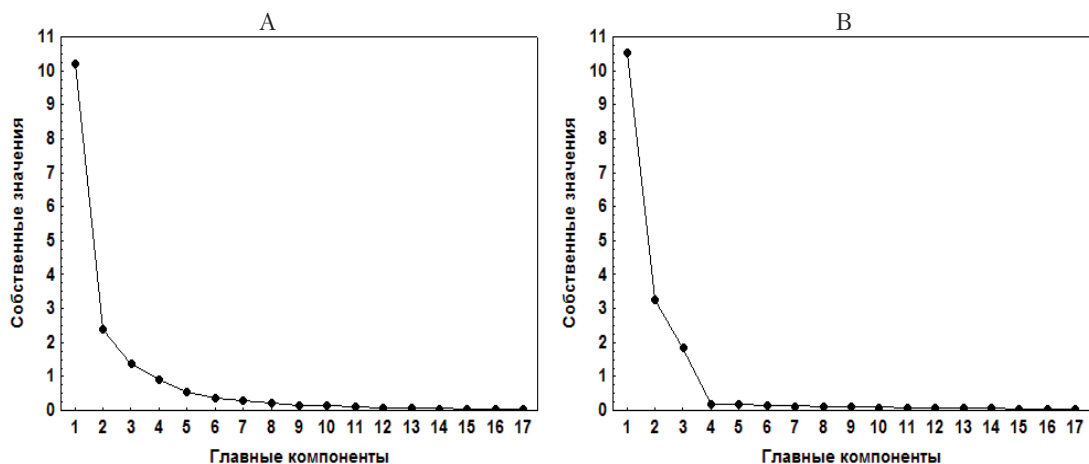


Рис. 3. Графики собственных значений главных компонент: (А) – матрица А, (В) – матрица М

3. Оставить число главных компонент, собственные значения которых превышают 1 (среднее собственное значение для матрицы корреляций). Данная эвристическая процедура была выработана Кайзером после длительных экспериментальных исследований (Ким и др., 1989). Согласно этой процедуре, $k = 3$ для матриц А, С и М и $k = 4$ для матрицы В (табл. 2).

Суммировав полученные результаты отбора первых главных компонент правилами, основанными на собственных числах, получим следующие результаты: для матрицы М по всем трем процедурам $k = 3$, для матриц А и С $k = 3\div 4$ компонентам, для матрицы В – $4\div 5$ компонентам.

Отобранные на основе собственных значений главные компоненты позволяют нам перейти к построению факторной модели. Если метод главных компонент дает возможность проанализировать максимальные вклады в суммарную дисперсию признаков (используя собственные значения), то цель факторного анализа – поиск наилучшей аппроксимации корреляционных связей между признаками, для достижения которой используют вращение факторного пространства. Основную модель факторного анализа можно записать следующим способом (Харман, 1972):

$$z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jk}F_k + u_j \quad (j = 1, 2, \dots, m),$$
 где z_j – наблюдаемый признак (переменная), который линейно зависит от k не коррелированных между собой (независимых, ортогональных) общих факторов F_1, F_2, \dots, F_k и характерной компоненты u_j ; m – число анализируемых признаков. Коэффициенты при общих факторах a_{ji} называются факторными нагрузками (*factor loadings*) и представляют собой линейную корреляцию между переменными и общими факторами, отражая меру влияния общего фактора на признак. В общие факторы объединяются переменные, сильно коррелирующие между собой. Если в модели главных компонент вся дисперсия приписывается m компонентам (характерный компонент u_j всех признаков принимается равным 0, а число главных компонент равно числу исходных признаков), то в факторном анализе дисперсия каждой исходной переменной делится на две части: дисперсию, обусловленную наличием k общих для всех признаков факторов F_1, F_2, \dots, F_k (общность – *communality*), и дисперсию, обусловленную характерным, «специфическим» для признака z_j остаточным случайным компонентом u_j (специфичность или характерность), который включает в себя, как правило, ошибки измерения признака (Айвазян и др., 1989; Rencher, 2002). После построения всех главных компонент остаточная дисперсия

оказывается равной 0, т.е. задача имеет точное математическое решение. В отличие от метода главных компонент, в случае факторного анализа возможность однозначного восстановления значений каждого из наблюдаемых признаков z_j по соответствующим значениям общих факторов F_1, F_2, \dots, F_k отсутствует. Исследователь стремится построить такую факторную модель, которая, с одной стороны, позволяет содержательно интерпретировать минимальное число латентных общих факторов, а с другой, как можно более полно объяснить с их помощью наблюдаемые взаимосвязи между исходными переменными, минимизируя зависимость отдельных признаков от их специфических остаточных случайных компонент. В рамках факторной модели искомые общие факторы могут рассматриваться в качестве причин (детерминантов), а исходные признаки – в качестве их следствий.

Возведенная в квадрат факторная нагрузка a_{ji} (коэффициент корреляции между фактором и переменной) характеризует степень общности (или перекрытия, совпадения) дисперсий данной переменной и данного общего фактора. Если полученные общие факторы не зависят друг от друга («ортогональное» решение), то общность переменной z_j по всем факторам (часть дисперсии переменной, совпадающей с дисперсиями всех факторов – h_j^2) можно оценить через сумму квадратов факторных нагрузок переменной по факторам, включенным в модель (Иберла, 1980):

$$h_j^2 = \sum_{i=1}^k a_{ji}^2 \quad (j = 1, 2, \dots, m), \text{ где } a_{ji} - \text{ факторные нагрузки, } m - \text{ число переменных, } k - \text{ число общих факторов.}$$

Значения общности переменной лежат в диапазоне от 0 до 1, они позволяют понять, какая часть дисперсии переменной объясняется общими факторами. Чем выше значение общности переменной, тем точнее факторная модель объясняет дисперсию анализируемого признака. Если взаимосвязь переменной с другими исследуемыми признаками крайне невысокая, ее общность будет низкой. Это может быть обусловлено тем, что переменная измеряет качественно отличный от других переменных показатель, либо ошибкой измерения.

Факторные нагрузки позволяют также вычислить, какую часть суммарной дисперсии переменных объясняет каждый фактор F_i (λ_i^2). Для этого рассчитываются собственные значения факторов (Иберла, 1980):

$$\lambda_i^2 = \sum_{j=1}^m a_{ji}^2 \quad (i = 1, 2, \dots, k), \text{ где } a_{ji} - \text{ факторные нагрузки, } m - \text{ число переменных, } k - \text{ число общих факторов;}$$

разделив собственные значения факторов на число m , мы получим долю суммарной дисперсии переменных, которую объясняет каждый общий фактор, и сможем оценить их относительную значимость.

Для максимально возможной концентрации дисперсии исходных признаков на координатных осях выделенных общих факторов («контрастирования» факторных нагрузок) и облегчения содержательной интерпретации факторов используют различные методы вращения факторного пространства. Наиболее часто используется вращение по методу Varimax, при котором по каждому фактору минимизируется количество переменных, имеющих высокие нагрузки, что максимально увеличивает дисперсию факторов (максимизируются различия столбцов матрицы нагрузок). Это обеспечивает разделение факторов за счет уменьшения числа исходных переменных, связанных с каждым фактором, и упрощает их содержательную интерпретацию. Метод Varimax предполагает, что общие факторы ортогональны (независимы). В табл. 3 приведены матрицы факторных нагрузок (факторное отображение – factor pattern) для модели с тремя общими факторами: метод выделения факторов – метод главных компонент, метод вращения факторного пространства – Varimax.

Таблица 3. Матрицы факторных нагрузок для модели с тремя общими факторами $F1 \div F3$: N – номер переменной, λ^2 – собственное значение фактора (в скобках указана доля λ^2 в общей дисперсии переменных).

N	Матрица А			Матрица В			Матрица С			Матрица М		
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F1	F2	F3
1	0.159	0.755	0.312	0.076	0.721	0.268	0.146	0.754	0.282	0.011	0.911	0.323
2	0.171	0.871	0.199	0.035	0.828	0.138	0.169	0.871	0.182	0.138	0.926	0.215
3	0.214	0.889	0.125	0.069	0.857	0.025	0.205	0.891	0.122	0.276	0.913	0.119
4	0.295	0.853	0.114	0.183	0.843	-0.001	0.267	0.860	0.119	0.420	0.853	0.073
5	0.478	0.751	0.092	0.379	0.723	-0.054	0.440	0.767	0.092	0.568	0.772	0.049
6	0.661	0.597	0.120	0.603	0.567	0.023	0.639	0.615	0.098	0.702	0.650	0.054
7	0.743	0.487	0.185	0.674	0.454	0.104	0.746	0.481	0.159	0.782	0.532	0.116
8	0.797	0.409	0.233	0.734	0.317	0.108	0.802	0.404	0.195	0.845	0.408	0.200
9	0.860	0.317	0.252	0.837	0.175	0.104	0.872	0.305	0.221	0.862	0.285	0.300
10	0.865	0.230	0.318	0.831	0.051	0.193	0.880	0.217	0.289	0.838	0.177	0.429
11	0.816	0.151	0.413	0.749	0.019	0.386	0.832	0.133	0.389	0.764	0.101	0.559
12	0.724	0.080	0.540	0.561	-0.090	0.524	0.733	0.072	0.528	0.670	0.049	0.685
13	0.567	0.053	0.716	0.418	-0.069	0.723	0.560	0.043	0.715	0.533	0.045	0.791
14	0.375	0.093	0.852	0.183	0.046	0.862	0.365	0.071	0.855	0.392	0.076	0.879
15	0.210	0.160	0.889	0.018	0.118	0.832	0.192	0.153	0.892	0.254	0.126	0.911
16	0.153	0.338	0.737	0.077	0.301	0.594	0.130	0.339	0.730	0.112	0.216	0.929
17	0.337	0.365	0.702	0.388	0.318	0.516	0.302	0.373	0.703	0.002	0.344	0.901
λ^2	5.362	4.641	3.969	4.178	4.055	3.165	5.320	4.668	3.849	5.406	5.031	5.187
	(0.315)	(0.273)	(0.233)	(0.246)	(0.239)	(0.186)	(0.313)	(0.275)	(0.226)	(0.318)	(0.296)	(0.305)

Примеч.: жирным шрифтом выделены факторные нагрузки > 0.5 .

Существует ряд статистических критериев для проверки гипотезы о значимости полученного факторного решения по исходной корреляционной матрице \mathbf{R} . Все они основываются на оценке воспроизведенной с помощью факторной модели корреляционной матрицы \mathbf{R}^+ и матрицы остатков коэффициентов корреляции $\mathbf{R}^- = \mathbf{R} - \mathbf{R}^+$. Во многих из них используются определители матриц \mathbf{R}^+ и \mathbf{R}^- .

1. Критерий Лоули-Бартлетта – проверка гипотезы о достаточности выделенных k общих факторов для воспроизведения корреляционной матрицы \mathbf{R}^+ (Иберла, 1980):

$\chi^2 = (n-1) \times \ln(|\mathbf{R}^+|/|\mathbf{R}|)$, где $|\mathbf{R}^+|$ – определитель матрицы воспроизведенных корреляций, $|\mathbf{R}|$ – определитель исходной матрицы корреляций, n – объем выборки. Табличные значения χ_r^2 рассчитываются для $df = 0.5 [(m-k)^2 - m - k]$, где m – число переменных, k – число общих факторов. Если $\chi^2 > \chi_r^2$, необходимо увеличить число общих факторов.

2. Критерий Бартлетта-Уилкса – проверка гипотезы об окончании процесса выделения факторов по матрице остатков \mathbf{R} (Иберла, 1980). Данный критерий был описан нами выше. Единственное отличие заключается в том, что вместо определителя исходной матрицы корреляций $|\mathbf{R}|$ в формуле используется определитель матрицы $|\mathbf{R}^-|$. Если

$\chi^2 > \chi_r^2$, переменные зависимы, выполняется условие для увеличения числа общих факторов¹.

При расчетах определителя матрицы необходимо учитывать относительную погрешность при решении системы линейных уравнений, обусловленную машинной точностью, с которой совершаются все операции с вещественными числами (Иберла, 1980). Если корреляционная матрица плохо обусловлена, то значения определителя будут очень чувствительны к таким погрешностям. Оценкой погрешности вычислений может служить число обусловленности матрицы (Cond): чем выше величина Cond, тем выше относительная погрешность вычислений. Плохо обусловленными могут считаться матрицы, для которых $\text{Cond} \gg 1$ (Амосов и др., 1994). Например, для исходных матриц корреляций были получены следующие значения числа обусловленности: $\text{Cond}_A = 3.342e+02$, $\text{Cond}_B = 9.341e+01$, $\text{Cond}_C = 3.035e+02$, $\text{Cond}_M = 3.880e+02$. Число обусловленности для соответствующих матриц воспроизведенных корреляций и остатков при различных факторных решениях (количествах общих факторов) колебалось уже в диапазоне $3.074e+04 \div 3.896e+06$. Плохая обусловленность корреляционных матриц ставит под сомнение результаты использования перечисленных критериев. Не случайно во всех случаях применения критерия Лоули-Бартлетта и Бартлетта-Уилкса были получены чрезвычайно высокие значения χ^2 ($p = 0.000$), не позволяющие сделать выбор между различными факторными решениями.

Чтобы обойти проблему, связанную с низкой обусловленностью матриц, были предложены критерии, не связанные с расчетом определителей.

1. Критерий Лоули и Максвелла (аппроксимация критерия Лоули-Бартлетта) (Иберла, 1980):

$$\chi^2 = n' \cdot \sum_{j < i} \frac{r_{ij(ocm)}^2}{u_i^2 \cdot u_j^2}$$
, где $r_{ij(ocm)}$ – коэффициенты корреляций матрицы остатков (наддиагональные элементы), u_i^2 и u_j^2 – значения характеристических (специфических) компонент переменных z_i и z_j (равны соответствующим диагональным элементам матрицы остатков), $n' = n - (2m + 5)/6 - 2k/3$, $df = 0.5 [(m - k)^2 - m - k]$ (n – объем выборки, m – число переменных, k – число общих факторов). Для данного критерия нами также были получены очень высокие значения χ^2 ($p = 0.000$), не позволяющие сделать выбор между различными факторными решениями. В работе (Иберла, 1980) делается предупреждение, что применение данного критерия не всегда приводит к хорошим результатам.

2. Критерий Лоули (аппроксимация критерия Бартлетта-Уилкса) (Иберла, 1980):

$$\chi^2 = n \cdot \sum_{i < j} r_{ij(ocm)}^2$$
, где $r_{ij(ocm)}$ – коэффициенты корреляций матрицы остатков (наддиагональные элементы), n – объем выборки, $df = m(m-1)/2$ (m – число переменных).

Условие независимости остатков корреляции ($\chi^2 < \chi_r^2$ – нецелесообразно дальнейшее увеличение числа общих факторов) для рассматриваемых матриц было достигнуто при следующих значениях числа общих факторов. Матрицы **A** и **C**: $k = 4$, $\chi_A^2 = 115.01$, $p = 0.904$; $\chi_C^2 = 121.41$, $p = 0.810$ ($k = 3$, $\chi_A^2 = 309.21$, $p = 0.000$; $\chi_C^2 = 320.98$, $p = 0.000$). Матрица **B**: $k = 7$, $\chi_B^2 = 117.17$, $p = 0.941$ ($k = 6$, $\chi_B^2 = 176.97$, $p = 0.000$). Матрица **M**: $k = 3$, $\chi_M^2 = 19.26$, $p = 1.0$ ($k = 2$, $\chi_M^2 = 1026.67$, $p = 0.000$). При использовании данного критерия необходимо учитывать зависимость от объема выборки. Например, если бы объем выборки для матрицы **A** был равен 350, то для $k = 3$ критерий Лоули $\chi_A^2 = 147.44$, $p = 0.237$.

¹ В «Statistica 6.0» значения общности переменных h_j^2 приведены в соответствующих элементах главной диагонали матрицы \mathbf{R}^* , а специфичности u_j – матрицы \mathbf{R}^- .

3. Критерий Барта (Иберла, 1980):

$$\chi^2 = (n-3) \cdot \sum_{i < j} (z_{ij} - \bar{z})^2$$
, где z_{ij} – преобразование Фишера = $0.5 \times [(1+r_{ij})/(1-r_{ij})]$ (r_{ij} – наддиагональные элементы коэффициентов корреляции матрицы остатков), \bar{z} – среднее значение преобразований Фишера, n – объем выборки, $df = 0.5 \times [(m-k)^2 - m - k]$ (m – число переменных, k – число общих факторов).

Оценки остатков по критерию Барта приводят к повышению значений числа общих факторов для матриц **A** и **C**: $k = 5$, $\chi_A^2 = 57.94$, $p = 0.587$; $\chi_C^2 = 64.20$, $p = 0.365$ ($k = 4$, $\chi_A^2 = 108.83$, $p = 0.005$; $\chi_C^2 = 114.57$, $p = 0.002$). Для матрицы **M** выводы совпали: $k = 3$, $\chi_M^2 = 16.64$, $p = 1.0$ ($k = 2$, $\chi_M^2 = 1031.74$, $p = 0.000$). Показательно, что для матрицы **B** даже при $k = 10$ нулевая гипотеза отвергалась ($\chi^2 > \chi_{cr}^2$, $p = 0.000$): остатки корреляций статистически значимы, целесообразно продолжить увеличение числа общих факторов. Критерий Барта также зависит от объема выборки. Так, если бы объем матрицы **A** был равен 250, то для критерия Барта ($k = 3$) $\chi_A^2 = 100.38$, $p = 0.173$.

Данные статистических критериев Лоули и Барта в целом подтвердили выводы, сделанные с помощью правил собственных значений главных компонент. Для матрицы **M** наиболее оптимальным является решение с тремя общими факторами ($k=3$). Для матриц **A** и **C**, согласно критерию Лоули, целесообразно рассмотреть содержательную интерпретацию моделей с тремя и четырьмя общими факторами. Критерий Барта дает завышенные оценки числа общих факторов ($k=5$) для матриц **A** и **C**. Для матрицы **B** использование статистических критериев приводит к чрезмерной переоценке числа общих факторов в модели, особенно в случае критерия Барта.

Для определения окончательного числа общих факторов обратимся к научным представлениям о природе спектральных показателей сердечного ритма. Международный стандарт по измерениям и физиологической интерпретации вариабельности сердечного ритма рекомендует рассматривать три частотные области в сердечном спектре, имеющие свои уникальные механизмы регуляции: VLF (0.00÷0.04 Гц), LF (0.04÷0.15 Гц) и HF (0.15÷0.40 Гц) (Heart Rate Variability, 1996). С учетом данных рекомендаций и полученных ранее выводов мы можем остановиться на факторной модели с тремя общими факторами.

Чтобы еще раз проверить правильность факторного решения, воспользуемся принципом «простой структуры», предложенным Тэрстоуном. Достижение простой факторной структуры с помощью вращения факторного пространства облегчает интерпретацию переменных и факторов, правда, концепция «простой структуры» неоднозначна и единых формальных критериев ее определения не существует (Ким и др., 1989). Но с точки зрения простой структуры мы должны стремиться к построению такой матрицы факторных нагрузок, в которой каждая переменная z_j имела бы высокую корреляцию (a_{ji}) только с одним фактором F_i . В табл. 3 мы выделили жирным шрифтом высокие значения факторных нагрузок $a_{ji} > 0.5$. Для всех матриц хорошо видны три группировки переменных, образованные соседними частотными диапазонами, которые имеют высокие нагрузки с отдельными факторами. Например, для матрицы **A** это $V_6 \div V_{13}$ (высокая корреляция с фактором 1), $V_1 \div V_6$ (высокая корреляция с фактором 2) и $V_{12} \div V_{17}$ (высокая корреляция с фактором 3). Как можно заметить, ряд переменных из этих групп имеют высокие нагрузки более чем с одним фактором: V_6 – факторы 1 и 2, V_{12} и V_{13} – факторы 1 и 3. Такие переменные считаются сложными (они отражают влияние более чем одного фактора). Цель вращения состоит в минимизации числа таких переменных, затрудняющих интерпретацию факторов.

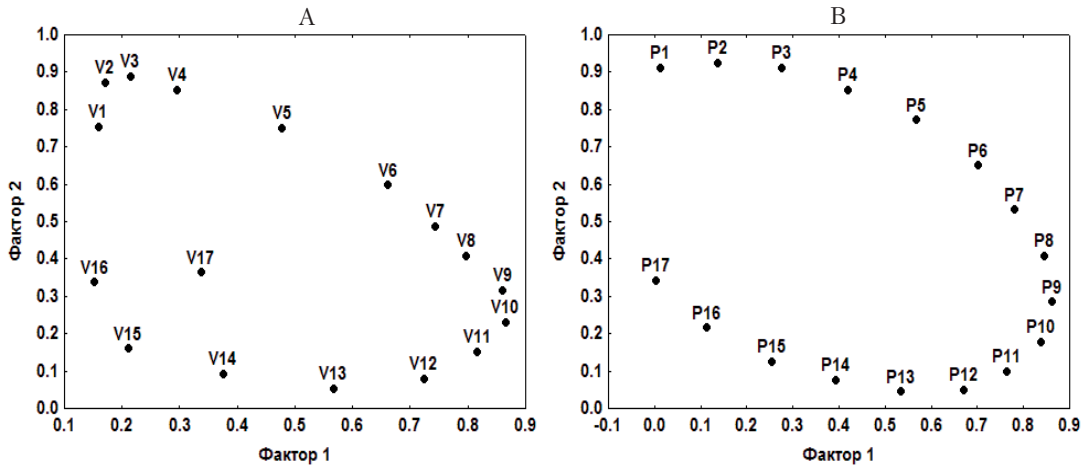


Рис. 4. Графики факторных нагрузок для переменных $V_1 \div V_{17}$ матрицы **A** (А) и переменных $P_1 \div P_{17}$ матрицы **М** (В)

Чтобы понять природу появления этих сложных переменных, обратимся к графической оценке простоты факторной структуры. Для этого построим график факторных нагрузок. На рис. 4 для матриц **A** и **М** изображено двухмерное факторное пространство, образованное ортогональными факторными осями 1 и 2. При наличии простой структуры факторные нагрузки группируются таким образом, чтобы иметь максимальные значения по одной факторной оси и минимальные по другой. В нашем случае факторные нагрузки принимают форму, близкую к эллипсу (отметим, что данная особенность характерна и для факторного решения с четырьмя общими факторами). Нарушение принципа простой структуры отражает характер внутренней взаимосвязи между переменными: высокая корреляция соседних частотных диапазонов и постепенное «затухание» корреляций с увеличением расстояния между диапазонами. Вращение факторного пространства привело к выделению взаимосвязанных между собой групп переменных (по принципу близости частотных диапазонов), но не смогло исключить наличие сложных переменных, «пограничных» для этих групп. При этом объем групп в различных факторах приблизительно равен, как и собственные значения факторов (λ^2). Частотные диапазоны, которые «детерминируют» факторы, совершенно не совпадают с результатами многочисленных физиологических исследований, на которые опираются международные рекомендации (Heart Rate Variability, 1996). Например, для матрицы **A** (табл. 3): первый фактор имеет высокую корреляцию ($a_{ji} > 0.6$) с переменными частотного диапазона $0.10 \div 0.24$ Гц, второй фактор – с переменными частотного диапазона $0.0 \div 0.10$ Гц, третий фактор – с переменными частотного диапазона $0.24 \div 0.50$ Гц. В этом случае правильно говорить о том, что существует корреляция между факторами и переменными, но ошибочно было бы утверждать, что выделенные латентные факторы детерминируют исследуемые переменные. Заметим, что факторная нагрузка для переменной V_{17} ($0.40 \div 0.50$ Гц) несколько выпадает из «ансамбля» (рис. 4, А). Напомним, что международные рекомендации ограничивают высокочастотный диапазон спектра сердечного значением 0.40 Гц (Heart Rate Variability, 1996).

В завершение нашего анализа произведем сравнение факторных структур для четырех матриц. Для этого воспользуемся коэффициентом конгруэнтности (Харман, 1972):

$$\varphi_{pq} = \frac{\sum (a_{ip} \times a_{iq})}{\sqrt{\sum (a_{ip}^2) \times \sum (a_{iq}^2)}} \quad (i=1, 2, \dots, m), \text{ где } a_{ip} - \text{элементы матрицы нагрузок фактора } p, a_{iq} - \text{элементы матрицы нагрузок фактора } q, i - \text{номер переменной,}$$

m – число переменных. Показатель φ_{pq} изменяется в диапазоне от +1 (полное совпадение) или -1 (полное обратное совпадение) до 0 (полное отсутствие связи). Для данного показателя отсутствуют статистические таблицы значимости. Экспериментально было предложено: $\varphi_{pq} = 0.999984 \div 0.93981$ – факторы конгруэнтны, $\varphi_{pq} = 0.459717$ – недостаточно для определения конгруэнтности. Для усиления различий (совпадений) между факторами предложен следующий индекс контрастности:

$$Z = \ln \left(\sqrt{\frac{1 + \varphi_{pq}}{1 - \varphi_{pq}}} \right), \text{ где } \ln - \text{натуральный логарифм. Согласно полученным результатам}$$

(табл. 4), между факторными решениями исследуемых матриц существует высокая конгруэнтность; особенно высоко совпадение факторных нагрузок матриц А и С.

Таблица 4. Коэффициенты конгруэнтности для факторов F1÷F3 исследуемых матриц (в скобках приведены значения индекса контрастности Z)

Матрица А	Матрица В			Матрица С			Матрица М		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
F1	0.985 (2.43)	0.493 (0.54)	0.570 (0.65)	0.999 (4.11)	0.607 (0.70)	0.643 (0.76)	0.983 (2.39)	0.581 (0.66)	0.652 (0.78)
F2	0.546 (0.61)	0.986 (2.50)	0.321 (0.33)	0.600 (0.69)	1.000 (4.55)	0.442 (0.47)	0.616 (0.72)	0.995 (2.98)	0.418 (0.45)
F3	0.577 (0.66)	0.360 (0.38)	0.981 (2.33)	0.654 (0.78)	0.447 (0.48)	0.999 (3.97)	0.600 (0.69)	0.409 (0.44)	0.992 (2.76)

Примеч.: жирным шрифтом выделены коэффициенты конгруэнтности > 0.90.

По аналогии с матрицей А мы построили две дополнительные матрицы А₁ и А₂, разделив исходный объем данных (n=734) на две равные половины (n₁=367, n₂=367). И матрица А₁ и матрица А₂ имели высокие коэффициенты конгруэнтности с факторным решением матрицы А. Также дополнительно мы провели факторный анализ коэффициентов корреляции Спирмена для исходных переменных D₁÷D₂₃ и пришли к схожему факторному решению с тремя общими факторами: суммарная доля дисперсии 91.46%, $\lambda_3^2 > 1$, $h_1^2 = 31.6\%$ (D₁₅÷D₂₃, 0.28÷0.46 Гц), $h_2^2 = 31.4\%$ (D₁÷D₇, 0.0÷0.14 Гц), $h_3^2 = 28.5\%$ (D₈÷D₁₄, 0.14÷0.28 Гц).

Заключение

1. Частотный диапазон спектра сердечного ритма был разбит на 25 равных областей, которые после проверки нормальности распределения показателей мощности спектра были сгруппированы в 17 переменных (см. табл. 1). Факторный анализ был выполнен для четырех корреляционных матриц: А – нормальное распределение переменных (пред-

варительное преобразование к нормальному виду), параметрическая корреляция Пирсона; **В** – нарушение нормального распределения переменных (без преобразования к нормальному виду), параметрическая корреляция Пирсона; **С** – нарушение нормального распределения переменных (без преобразования к нормальному виду), непараметрическая корреляция Спирмена; **М** – моделирование периодичности в изменениях парных коэффициентов корреляции: максимум амплитуды – корреляция соседних частотных диапазонов, уменьшение корреляции по мере увеличения расстояния между частотными диапазонами (рис. 1). Для матриц **А** и **М** было достигнуто сходство в распределении парных корреляций, при этом распределения частных корреляций различались (рис. 2).

2. Для всех четырех матриц критериями Бартлетта-Уилкса и Кайзера была подтверждена целесообразность использования факторной модели. Для матриц **А** и **С** по критерию MSA был достигнут уровень высокой адекватности, для матрицы **М** – уровень безусловной адекватности, для матрицы **В** – уровень приемлемой адекватности факторному анализу.

3. Используя правила собственных чисел для отбора общих факторов (k), выделенных методом главных компонент (табл. 2), были получены следующие факторные решения: для матриц **А** и **С** $k = 3 \div 4$, для матрицы **В** $k = 4 \div 5$, для матрицы **М** $k = 3$.

4. Значимость факторных решений (после вращения факторного пространства методом Varimax) была проанализирована с помощью ряда статистических критериев, из которых наиболее ценными оказались критерии Лоули и Барта. Результаты критерия Лоули в целом подтвердили выводы, сделанные с помощью правил собственных значений главных компонент: для матрицы **М** наиболее оптимальным является решение с тремя общими факторами ($k=3$); для матриц **А** и **С** целесообразно рассмотреть содержательную интерпретацию моделей с тремя и четырьмя общими факторами. Критерий Барта дает идентичные оценки числа общих факторов ($k=3$) для матрицы **М** и завышает для матриц **А** и **С** ($k=5$). Для матрицы **В** использование данных статистических критериев приводит к чрезмерной переоценке числа общих факторов в модели (особенно в случае критерия Барта).

5. С учетом научных представлений о природе спектральных показателей сердечного ритма была выбрана факторная модель с тремя общими факторами. Для заключительной проверки правильности факторного решения был использован принцип «простой структуры». Анализ двухмерного факторного пространства позволил установить нарушение принципа простой структуры, вызванное характером внутренних взаимосвязей между переменными: высокая корреляция соседних частотных диапазонов и постепенное «затухание» корреляций с увеличением расстояния между диапазонами. Вращение факторного пространства привело к выделению взаимосвязанных между собой групп переменных (по принципу близости частотных диапазонов) и появлению сложных переменных, «пограничных» для этих групп (табл. 3). Объем групп в различных факторах был приблизительно равен, как и собственные значения факторов. Частотные диапазоны, с которыми факторы имеют высокую корреляцию, совершенно не совпадают с результатами многочисленных физиологических исследований. В этом случае было бы ошибочно утверждать, что выделенные латентные факторы детерминируют (регулируют) исследуемые переменные, и на этом основании строить психофизиологическую классификацию объектов исследования, как это представлено, например, в работах Н. Н. Даниловой.

6. Для сравнения факторных структур четырех матриц был использован коэффициент конгруэнтности. Согласно полученным результатам (табл. 4), между факторными ре-

шениями исследуемых матриц существует высокая конгруэнтность. Особенно высоко совпадение факторных нагрузок матриц **A** и **C**.

7. Использование в факторном анализе переменных с выраженными нарушениями нормальности распределения (матрица корреляций Пирсона **B**) привело к завышению числа отобранных общих факторов, чему также способствовали статистические процедуры, основанные на собственных значениях главных компонент и чрезмерной переоценке числа общих факторов на основании применения критерия Лоули и особенно критерия Барта (подтверждается чувствительность статистических критериев к допущению о многомерном нормальном распределении исходных переменных; см. Справочник по прикладной статистике, 1990). В то же время факторный анализ для матрицы корреляций Спирмена этих переменных (**C**) продемонстрировал практически полное совпадение с результатами нормально распределенных переменных, представленных матрицей корреляций Пирсона (**A**). Также мы получили высокое согласие факторных решений между матрицей **A** и модельными данными матрицы **M**, имеющими близкое распределение парных корреляций и значимые отличия в распределении частных корреляций переменных.

Таким образом, при выраженных нарушениях нормальности распределения исследуемых показателей в качестве исходных данных для проведения факторного анализа можно рекомендовать:

- а) матрицу парных коэффициентов корреляции Пирсона (параметрическая статистика) с предварительным преобразованием распределений показателей к нормальному виду;
- б) матрицу парных коэффициентов корреляции Спирмена (непараметрическая статистика) без предварительного преобразования распределений показателей к нормальному виду.

В заключение необходимо заметить, что факторный анализ, подобно регрессионному, можно отнести к области математического моделирования: создание модели объекта исследования в форме математического уравнения взаимосвязей исследуемых показателей с факторами. Но любая модель отражает только некоторые характерные черты объекта и никогда не бывает его точной копией. Для одного и того же объекта можно создать множество моделей (Вучков и др., 1987). Поэтому для построения и проверки факторной модели исследователю необходимо использовать как можно более полный набор процедур, критериев и научных представлений об объекте исследования: рассмотреть допущение о многомерном нормальном распределении исходных переменных; выбрать тип матрицы парных коэффициентов корреляции в качестве исходных данных; проверить целесообразность применения факторной модели анализа к исходным данным; выполнить процедуры отбора общих факторов; оценить значимость полученного факторного решения с помощью статистических критериев и принципа «простой структуры».

К сожалению, объем журнальной публикации не позволяет нам проанализировать факторные решения с зависимыми (облическими) факторами, сравнить метод главных компонент с другими методами выделения факторов, а Varimax – с другими методами вращения факторного пространства. Наши исследования показали, что для исходных данных – значений спектра мощности сердечного ритма в заданных частотных диапазонах – метод главных компонент в сочетании с Varimax-вращением является наиболее оптимальным с точки зрения построения факторной модели. Именно на этом мы и постарались детально остановиться в нашей статье.

Литература

- 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.
- Амосов А. Л., Дубинский Ю. Л., Копченова Н. В. Вычислительные методы для инженеров. М.: Высшая школа, 1994. 544 с.
- Вучков И., Бояджијева Л., Солаков Е. Прикладной линейный регрессионный анализ. М.: Финансы и статистика, 1987. 239 с.
- Данилова Н. Н. Сердечный ритм и информационная нагрузка // Вестник МГУ. Серия 14 (Психология). 1995. № 4. С. 14–19.
- Данилова Н. Н. Стрессоустойчивость как индивидуальная особенность // Доклады I-й Международной конференции памяти А. Р. Лурия. М.: МГУ, 1998. С. 177–192.
- Данилова Н. Н., Астафьев С. В. Изменения вариабельности сердечного ритма при информационной нагрузке // Журн. выс. нерв. деят. им. Павлова. 1999. Т. 49. № 1. С. 28–38.
- Зырянов Б. А., Власов С. Н., Костромин Э. В. Методы и алгоритмы обработки случайных и детерминированных периодических процессов. Свердловск: УГУ. 1990. 116 с.
- Иберла К. Факторный анализ. М.: Статистика, 1980. 398 с.
- Кендалл М., Стюарт А. Статистические выводы и связи. М.: Наука. Т. 2. 1973. 899 с.
- Ким Дж.-О., Мьюллер Ч. У., Клекка У. Р., Олдендерфер М. С., Блэшфилд Р. К. Факторный, дискриминантный и кластерный анализ. М.: Финансы и статистика, 1989. 215 с.
- Лоули Д. Н., Максвелл А. Э. Факторный анализ как статистический метод. М.: Мир, 1967. 145 с.
- Машин В. А. Зависимость показателей вариабельности сердечного ритма от средней величины R-R интервалов // Рос. физиол. журн. им. И. М. Сеченова. 2002. Т. 88. № 7. С. 851–855.
- Справочник по прикладной статистике / Под ред. Э. Ллойда, У. Ледермана, С. А. Айвазяна, Ю. Н. Тюрина. Т. 2. М.: Финансы и статистика, 1990. 526 с.
- Хальд А. Математическая статистика с техническими приложениями. М.: Изд. иностранной литературы, 1956. 663 с.
- Харман Г. Современный факторный анализ. М.: Статистика. 1972. 487 с.
- Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology // Circulation. 1996. V. 93. № 5. P. 1043–1065.
- Rencher A. C. Methods of multivariate analysis. Wiley. 2002. 739 p.

METHODOLOGICAL ISSUES OF APPLICATION OF FACTOR ANALYSIS TO PSYCHOPHYSIOLOGICAL STUDIES

MASHIN V.A., *Novovoronezhskiy Training Center for specialists of nuclear power plants, Novovoronezh*

This article considers a number of methodological issues relating to the use of factor analysis for the assessment of the spectral characteristics of heart rate, namely: the assumption of a multivariate of normal distribution of investigated characteristics, appropriateness of applying of the factor analysis model to the original data, of the procedure of the selection of common factors, and of the significance of the factor solution. In our study we used variables characterized by disruption of normal distribution, variables converted to a normal distribution, as well as the correlation matrix of Pearson and Spearman, and the model data. In the process of the factor analysis (the principal components method, Varimax-rotation) we investigate both the possibilities and limitations of various statistical tests and procedures and considered the nature of the marked common factors and capability of the use of Spearman correlations for violations of normality.

Keywords: factor analysis, heart rate spectrum, normal distribution, statistical fitness tests, Pearson's and Spearman's correlation coefficients.

Transliterations of the Russian references

- Ajvazjan S. A., Buhstaber V. M., Enjukov I. S., Meshalkin L. D.* Prikladnaja statistika: Klassifikacija i snizhenie razmernosti. M.: Finansy i statistika, 1989. 607 c.
- Amosov A. L., Dubinskij Ju. L., Kopchenova N. V.* Vychislitel'nye metody dlja inzhenerov. M.: Vysshaja shkola, 1994. 544 s.
- Vuchkov I., Bojadzhieva L., Solakov E.* Prikladnoj linejnij regressionnyj analiz. M.: Finansy i statistika, 1987. 239 s.
- Danilova N. N.* Serdechnyj ritm i informacionnaja nagruzka // Vestnik MGU. Serija 14 (Psihologija). 1995. № 4. S. 14–19.
- Danilova N. N.* Stressoustojchivost' kak individual'naja osobennost' // Doklady I-j Mezhdunarodnoj konferencii pamjati A. R. Lurija. M.: MGU, 1998. S. 177–192.
- Danilova N. N., Astaf'ev S. V.* Izmenenija variabel'nosti serdechnogo ritma pri informacionnoj nagruzke // Zhurn. vys. nerv. dejat. im. Pavlova. 1999. T. 49. № 1. S. 28–38.
- Zyrjanov B. A., Vlasov S. N., Kostromin Je. V.* Metody i algoritmy obrabotki sluchajnyh i determinirovannyh periodicheskikh processov. Sverdlovsk: UGU. 1990. 116 s.
- Iberla K.* Faktornyj analiz. M.: Statistika, 1980. 398 s.
- Kendall M., St'juart A.* Statisticheskie vyvody i svjazi. M.: Nauka. T. 2. 1973. 899 s.
- Kim Dzh.-O., M'juller Ch. U., Klekka U. R., Oldenderfer M. S., Bljeshfild R. K.* Faktornyj, diskriminantnyj i klasternyj analiz. M.: Finansy i statistika, 1989. 215 s.
- Louli D. N., Maksvell A. Je.* Faktornyj analiz kak statisticheskij metod. M.: Mir, 1967. 145 s.
- Mashin V. A.* Zavisimost' pokazatelej variabel'nosti serdechnogo ritma ot srednej velichiny R-R intervalov // Ros. fiziol. zhurn. im. I. M. Sečenova. 2002. T. 88. № 7. S. 851–855.
- Spravochnik po prikladnoj statistike / Pod red. Je. Llojda, U. Ledermana, S. A. Ajvazjana, Ju. N. Tjurina. T. 2. M.: Finansy i statistika, 1990. 526 s.
- Hal'd A.* Matematicheskaja statistika s tehničeskimi prilozhenijami. M.: Izd. inostranoj literatury, 1956. 663 s.
- Harman G.* Sovremennij faktornyj analiz. M.: Statistika. 1972. 487 s.