

## **Сравнительный анализ методов факторизации в социологических исследованиях с применением симуляции методом Монте-Карло**

Несмотря на длительное и повсеместное использование факторного анализа и большое количество методов выделения факторной структуры, в социальных науках сложилась традиция подразумевать под факторным анализом сокращение размерности при помощи метода главных компонент. При этом учебники и руководства по применению факторного анализа обычно ограничиваются рассмотрением оснований, на которых этот метод выделяет факторы, но не указывают, каким исходным данным соответствует каждый из этих методов и как различаются итоговые результаты факторизации. Некоторые из них описывают исключительно метод главных компонент, объясняя это его распространенностью. В части исследований с применением факторного анализа и руководств по его применению зачастую трудно даже определить, какой из методов факторизации описывается. Исследований или руководств, которые бы систематизировали и эмпирически подтверждали разрозненные рекомендации по применению того или иного метода факторизации в различных исследовательских ситуациях, пока не существует. Таким образом, **проблема** исследования сводится к отсутствию четких рекомендаций относительно выбора метода факторизации в зависимости от исследовательской ситуации. **Цель** исследования – разработать алгоритм выбора адекватного метода факторизации в зависимости от исследовательской ситуации.

В данной работе систематизируются и обобщаются разрозненные рекомендации о выборе одного из семи методов факторизации (**объект** исследования): *метод главных компонент, метод невзвешенных наименьших квадратов, обобщенный метод наименьших квадратов, метод максимального правдоподобия, метод главных осей, альфа-факторный анализ и анализ образов*. Эти методы уже неоднократно сравнивались между собой, однако прежде сравнения ограничивались двумя или тремя методами факторизации из этого списка в достаточно узких или специфических исследовательских ситуациях. Под *исследовательской ситуацией* здесь понимается определенная комбинация следующих характеристик: размер выборки; количество индикаторов, приходящихся на один фактор; размер и разброс общностей; наличие или отсутствие модельной ошибки и форма распределения индикаторов.

Поскольку исследование предполагало экспериментальный дизайн, использование реальных выборочных данных с неизвестной ошибкой оценивания могло повлиять на

качество результатов с точки зрения достоверности и воспроизводимости. Поэтому для того, чтобы результаты исследования можно было генерализовать, в качестве информационной базы был использован набор массивов, сгенерированных методом Монте-Карло.

На основе анализа математического аппарата, заложенного в каждый из методов факторизации, а также предыдущих исследований их поведения в различных исследовательских ситуациях, был составлен теоретический алгоритм выбора метода факторизации в зависимости от исследовательской ситуации. Он был эмпирически протестирован при помощи статистического эксперимента, состоящего из 5 последовательных шагов: 1) спецификация матриц факторных нагрузок; 2) моделирование корреляционных матриц; 3) симуляция массивов; 4) применение методов факторизации к симулированным массивам; 5) сравнение ситуативной адекватности методов факторизации в различных исследовательских ситуациях и выявление наиболее ситуативно адекватного метода для каждой из ситуаций.

В результате эксперимента только для двух исследовательских ситуаций из шести смоделированных теоретически обоснованный алгоритм выбора метода подтвердился. Вероятно, причиной таких результатов стало сочетание большого количества критериев, образующих исследовательскую ситуацию: предыдущие исследования, на которых основывался теоретический алгоритм выбора методов факторизации, сравнивали эти методы в ситуациях, образованных не более чем тремя критериями, в нашем же случае варьируемых параметров было 6. Возможно, именно их характерные сочетания, а не каждый из параметров в отдельности, оказали решающий эффект на ситуативную адекватность методов.

По итогам эксперимента теоретически обоснованный алгоритм был уточнен. В целом, рекомендации по выбору метода можно свести к следующему:

1) **метод главных осей и альфа-факторный анализ** рекомендуется использовать в случае подозрения на наличие модельной ошибки;

2) **метод максимального правдоподобия и обобщенный метод наименьших квадратов** в равной степени рекомендованы к использованию в ситуациях, когда при отсутствии модельной ошибки а) распределение индикаторов является нормальным и объем выборки достаточен, и б) распределение индикаторов отличается от нормального и выборка недостаточно велика;

3) **метод максимального правдоподобия** рекомендуется к использованию тогда, когда при отсутствии модельной ошибки а) распределение индикаторов отличается

от нормального и выборка достаточно велика, и б) распределение индикаторов нормально, выборка недостаточно велика и общности не превышают 0,6;

4) **обобщенный метод наименьших квадратов** рекомендуется к использованию, когда при условии уверенности в отсутствии модельной ошибки распределение индикаторов нормально, объем выборки недостаточен и общности достаточно велики.

Три из семи рассмотренных методов – невзвешенный метод наименьших квадратов, анализ образов и метод главных компонент – в результате эмпирического тестирования алгоритма не оказались наиболее ситуативно адекватными в сформулированных для них исследовательских ситуациях. Этот результат существенен в свете того, что метод главных компонент, наиболее часто используемый в качестве синонима факторного анализа в социологических исследованиях, не оказался наиболее ситуативно адекватным в ситуации, представляющей собой комбинацию общепринятых требований к входным данным для факторного анализа (не менее 6 индикаторов на фактор, нормальное распределение индикаторов,  $N > 300$ , модельная ошибка отсутствует, размер общностей превышает 0,6, а размах не превышает 0,3), однако именно в этой ситуации, по сравнению с остальными пятью рассмотренными в работе, он показал себя наилучшим образом. Таким образом, если исследователь рассматривает РСА как единственный доступный для себя метод факторизации, ему рекомендуется *строго придерживаться всех требований к входным данным для анализа и контролировать полученные модели на размер и размах общностей*. Единственным допустимым отклонением от требований может быть распределение индикаторов, отличное от нормального, которое ухудшает показатели ситуативной адекватности метода главных компонент при прочих равных всего на 1%.