

Практическая работа №6: Исследование методов факторного анализа

Цель работы

Ознакомиться с методами факторного анализа на основе языка R.

Основные теоретические положения

Главными целями факторного анализа являются сокращение числа переменных (редукция данных) и определение структуры взаимосвязей между переменными, т.е. классификация переменных. Поэтому факторный анализ используется или как метод сокращения данных или как метод классификации.

Предположим, что вы проводите (до некоторой степени «глупое») исследование, в котором измеряете рост ста людей в дюймах и сантиметрах. Таким образом, у вас имеются две переменные. Если далее вы захотите исследовать, например, влияние различных пищевых добавок на рост, будете ли вы продолжать использовать обе переменные? Вероятно, нет, т.к. рост является одной характеристикой человека, независимо от того, в каких единицах он измеряется.

Теперь предположим, вы хотите измерить удовлетворенность людей жизнью, для чего составляете вопросник с различными пунктами; среди других вопросов задаете следующие: удовлетворены ли люди своим хобби (пункт 1) и как интенсивно они им занимаются (пункт 2). Результаты преобразуются так, что средние ответы (например, для удовлетворенности) соответствуют значению 100, в то время как ниже и выше средних ответов расположены меньшие и большие значения, соответственно. Две переменные (ответы на два разных пункта) коррелированы между собой. Из высокой коррелированности двух этих переменных можно сделать вывод об избыточности двух пунктов опросника.

Зависимость между переменными можно обнаружить с помощью диаграммы рассеяния. Полученная путем подгонки линия регрессии дает графическое представление зависимости. Если определить новую переменную на основе линии регрессии, изображенной на этой диаграмме, то такая переменная будет включать в себя наиболее существенные черты обеих переменных. Итак, фактически, вы сократили число переменных и заменили две одной. Отметим, что новый фактор (переменная) в действительности является линейной комбинацией двух исходных переменных.

Пример, в котором две коррелированные переменные объединены в один фактор, показывает главную идею факторного анализа или, более точно, анализа главных компонент (это различие будет обсуждаться позднее). Если пример с двумя переменными распространить на большее число переменных, то вычисления становятся сложнее, однако основной принцип представления двух или более зависимых переменных одним фактором остается в силе.

В основном процедура выделения главных компонент подобна вращению, максимизирующему дисперсию (варимакс) исходного пространства переменных. Например, на диаграмме

рассеяния вы можете рассматривать линию регрессии как ось X, повернув ее так, что она совпадает с прямой регрессии. Этот тип вращения называется вращением, максимизирующим дисперсию, так как критерий (цель) вращения заключается в максимизации дисперсии (изменчивости) «новой» переменной (фактора) и минимизации разброса вокруг нее.

Напомним, что анализ главных компонент является методом сокращения или редукции данных, т.е. методом сокращения числа переменных. Возникает естественный вопрос: сколько факторов следует выделять? Отметим, что в процессе последовательного выделения факторов они включают в себя все меньше и меньше изменчивости. Решение о том, когда следует остановить процедуру выделения факторов, главным образом зависит от точки зрения на то, что считать малой «случайной» изменчивостью. Это решение достаточно произвольно, однако имеются некоторые рекомендации, позволяющие рационально выбрать число факторов.

Общая формулировка задачи

- Ознакомиться с основами факторного анализа.
- Ознакомиться с работой в пакете Statistica с разделом Факторный анализ.
- Провести исследование по построению факторных моделей для исходных данных.
- Подготовить отчет по исследованию.

Порядок выполнения работы

1. Исследование по построению факторных моделей

1. В системе STATISTICA в панели Statistics→Multivariate Exploratory Techniques (Методы исследования многомерных данных) нужно выбрать пункт Factor (Дискриминантный анализ).
2. Кликнув по Values выбрать переменные для факторного анализа: номер класса не включаем. Кликнуть ОК.
3. На вкладке «Quick» ввести максимальное количество факторов равным 4, минимальное собственное число – 4. Кликнуть ОК.
4. В открывшемся окне Factor Analysis Results во вкладке Quick кликнуть по «Eigenvalues», таким образом выведем главные характеристики - собственные числа матрицы корреляций.
5. Вывести полученную таблицу в отчет.
6. Перейти во вкладку Explained Variance. Кликнуть Scree plot. Привести полученный график собственных чисел в отчете, а также матрицу корреляций (Вкладка Descriptives → Review Correlations...→ Correlations).
7. Привести в отчете Факторная нагрузка – корреляция между старыми переменными (признаками) и новыми факторами, выделенными из признаков. Для этого в окне Factor Analysis Results выбрать Factor Loadings.
8. Отобразить графики зависимостей между факторами (Вкладка Quick, кнопка Plot of factor loadings, 2D), а также все факторы в трехмерном пространстве (Вкладка Loadings, кнопка Plot of loadings, 3D)
9. Вывести коэффициенты факторной оценки: для анализа главных компонент, точные коэффициенты могут быть вычислены из (вращаемых или не вращаемых) факторных

нагрузок. Для этого перейдите во вкладку Scores и кликните по кнопке «Factor Scores Coefficients».

10. Также приведите таблицу с факторным множеством, основанным на факторных коэффициентах счета.

2. Выполнение исследования с поворотом факторных нагрузок

1. В окне Factor Analysis Results во вкладке Quick установите опцию Factor Rotation в значение Varimax raw.
2. Эта опция выполнит вращение (циклический сдвиг) факторных нагрузок. Это вращение (циклический сдвиг) нацелено, на максимизацию дисперсии необработанных факторных нагрузок поперек переменных для каждого фактора; это - эквивалентное максимизирование разницы в столбцах матрицы необработанных факторных нагрузок.
3. Выполните исследование и приведите в отчете следующие данные:
 - таблица факторных нагрузок;
 - графики зависимостей между факторами и график всех факторов в трехмерном пространстве;
 - коэффициенты оценки факторов.
4. Выполните аналогичную последовательность с другим видом поворота нагрузок. Приведите в отчете вышеперечисленные данные. Объясните полученные результаты.
5. Краткая справка о видах поворота:
 - Unordered List Item Quartimax raw: Эта опция выполнит quartimax вращение факторных нагрузок. Это вращение максимизирует дисперсии факторных нагрузок поперек факторов для каждой переменной; это эквивалентно максимизированию дисперсий в строках матрицы необработанных факторных нагрузок.
 - Biquartimax raw: Эта опция выполнит biquartimax вращение (циклический сдвиг) необработанных факторных нагрузок. Это вращение (циклический сдвиг) может рассматриваться, как бы “четная смесь” varimax и quartimax вращения. Это нацелено, одновременно на максимизацию сумм дисперсий необработанных факторных нагрузок поперек факторов и максимизацию сумм дисперсий необработанных факторных нагрузок поперек переменных; это эквивалентно одновременно максимизированию дисперсий в строках и столбцах матрицы необработанных факторных нагрузок.
 - Equamax raw: Эта опция выполнит equamax вращение необработанных факторных нагрузок. Это вращение может рассматриваться, как бы “тяжелой смесью” varimax и quartimax вращения. Это одновременно максимизирует сумму дисперсий необработанных факторных нагрузок поперек факторов и максимизирует сумму дисперсий необработанных факторных нагрузок поперек переменных; это эквивалентно одновременно максимизированию дисперсий в строках и столбцах матрицы необработанных факторных нагрузок. Однако, в отличие от biquartimax вращения, относительный вес, назначенный к varimax критерию во вращении равен числу факторов разделенных на 2.

3. Исследование метода главной оси

В этом методе, в каждой итерации, собственные числа вычислены из текущих корреляций; затем корреляции повторно вычислены на основании полученных собственных чисел. Новые корреляции затем помещены в диагональ матрицы корреляций, и следующая итерация начинается. Итерации продолжатся до:

- Максимальное число итераций превышено, или
 - Минимум
1. Для использования данного метода в окне «Define Method of Factor Extraction» выберите «Principal axis method».
 2. Приведите в отчете результаты исследования:
 - главные характеристики;
 - график собственных чисел;
 - матрица корреляций;
 - факторная нагрузка;
 - графики зависимостей между факторами;
 - коэффициенты факторной оценки.
 3. Сделайте вывод о проделанном исследовании.

From:

<http://se.moevm.info/> - **se.moevm.info**

Permanent link:

http://se.moevm.info/doku.php/courses:data_analysis_and_interpretation:task6?rev=1562853282



Last update: **2022/12/10 09:08**