

Практическая работа №5: 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.

Постановка задачи

Порядок выполнения работы

Содержание отчёта

From:

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

Permanent link:

http://se.moevm.info/doku.php/courses:data_analysis_and_interpretation:task5

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

