ФАКТОРНЫЙ АНАЛИЗ

В факторном анализе исходят из предположения о том, что каждый из исходных признаков ξ_j, $j = \overline{1,k}$ может быть представлен в виде суммы линейной комбинации небольшого числа латентных (скрытых) обобщенных факторов $f^{(1)}, f^{(2)}, \ldots, f^{(m)}$ и характерного фактора ε_j. При этом считается, что каждый обобщенный фактор имеет существенное значение для анализа всех исходных признаков. В то же время изменения в характерном факторе ε_j воздействуют на значения только соответствующего признака ξ_j. То есть, характерный фактор ε_j отражает ту специфику признака ξ_j, которая не может быть выражена через обобщенные факторы. Таким образом, в основе данного метода факторного анализа лежит модель вида:

$$\xi_j = \alpha_j^{(1)} f^{(1)} + \alpha_j^{(2)} f^{(2)} + \ldots + \alpha_j^{(m)} f^{(m)} + \varepsilon_j, \ j = \overline{1,k},$$

или в векторной форме:

$$\vec{\xi} = \vec{\alpha}^{(1)} f^{(1)} + \vec{\alpha}^{(2)} f^{(2)} + \ldots + \vec{\alpha}^{(m)} f^{(m)} + \vec{\varepsilon} = \alpha \vec{f} + \vec{\varepsilon} \tag{1}$$

где $\vec{\xi} = (\xi_1, \xi_2, \ldots, \xi_k)$ - вектор исходных признаков (факторов);

$\vec{f} = (f^{(1)}, f^{(2)}, \ldots, f^{(m)})$ - вектор латентных (обобщенных) центрированных и нормированных обобщенных факторов, некоррелированных между собой, $m < k$;

$\vec{\alpha}^{(1)}, \vec{\alpha}^{(2)}, \ldots, \vec{\alpha}^{(m)}$ – векторы факторных нагрузок; $\alpha = (\vec{\alpha}^{(1)}, \vec{\alpha}^{(2)}, \ldots, \vec{\alpha}^{(m)})$ - матрица факторных нагрузок;

$\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$ - вектор характерных факторов, некоррелированных как между собой, так и с обобщенными факторами.

В ходе факторного анализа необходимо оценить минимальное число факторов, определить векторы факторных нагрузок, дисперсии характерных факторов и вычислить значения факторов для каждого наблюдаемого объекта. Поскольку число обобщенных факторов предполагается существенно меньше числа исходных признаков, данная задача не имеет однозначного решения. Заметим, что если представление в виде (1) существует, то оно определяется с точностью до ортогонального преобразования матрицы факторных нагрузок. Действительно, пусть C - ортогональная матрица порядка m, тогда:

$$\alpha f = \alpha C C^T f = (\alpha C)(C^T f) = \tilde{\alpha} \tilde{f},$$

при этом: $\tilde{\alpha}\tilde{\alpha}^T = \alpha\alpha^T$, $A(\tilde{f}) = E$.

В зависимости от того, какие условия накладываются на обобщенные факторы, существуют различные модели и методы факторного анализа. Можно выделить два основных подхода к построению факторных моделей. В первом случае обобщенные факторы должны выделять большую часть суммарной дисперсии исходных факторов, во втором - обобщенные факторы должны наилучшим образом описывать связи между исходными факторами. Основным представителем группы методов, которые реализуют первый подход, является метод главных компонент. Методы же факторного анализа, базирующиеся на втором подходе, принято называть классическими методами факторного анализа (или методами канонического факторного анализа).

МЕТОД ГЛАВНЫХ КОМПОНЕНТ КАК МЕТОД ФАКТОРНОГО АНАЛИЗА

Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_k)$ – центрированная многомерная случайная величина с матрицей ковариаций $A = A(\vec{\xi})$. Положим

$$\xi_j = \alpha_j^{(1)} f^{(1)} + \alpha_j^{(2)} f^{(2)} + \ldots + \alpha_j^{(m)} f^{(m)} + \varepsilon_j,\ j = \overline{1,m},\ (m \le k)$$

где обобщенные факторы $f^{(1)}, f^{(2)}, \ldots, f^{(m)}$ - независимые случайные величины с нулевым математическим ожиданием и единичной дисперсией.

Согласно методу главных компонент факторного анализа, обобщенные факторы $f^{(1)}, f^{(2)}, \ldots, f^{(m)}$ ищутся как линейные комбинации исходных признаков $\xi_1, \xi_2, \ldots, \xi_k$:

$$f^{(j)} = \vec{\beta}^{(j)T}\vec{\xi} = \beta_1^{(j)}\xi_1 + \beta_2^{(j)}\xi_2 + \ldots + \beta_k^{(j)}\xi_k,$$

где $\vec{\beta}^{(j)}$, $j = \overline{1,m}$ – векторы неизвестных коэффициентов преобразования, удовлетворяющие условию:

$$D\left(f^{(j)}\right) = E\left(f^{(j)} f^{(j)T}\right) = \vec{\beta}^{(j)T} E\left(\xi\xi^T\right)\vec{\beta}^{(j)} = \vec{\beta}^{(j)T} A(\vec{\xi})\vec{\beta}^{(j)} = 1.$$

Линейные комбинации выбираются таким образом, чтобы среди всех возможных линейных комбинаций исходных признаков первый обобщенный фактор объяснял бы большую часть дисперсии исходных признаков. Второй обобщенный фактор выбирается из условия не коррелированности с первым и так, чтобы он объяснял большую часть оставшейся дисперсии и т.д.

Задача отыскания обобщенных факторов, удовлетворяющих условиям описанным выше, сводится к задаче на собственные значения и собственные векторы матрицы ковариаций A. Векторы коэффициентов преобразования $\vec{\beta}^{(i)}$, $\mathrm{i} = \overline{1,\mathrm{m}}$ удовлетворяющие условиям задачи поиска обобщенных факторов (или главных компонент), являются собственными векторами матрицы ковариаций A, соответствующими собственным значениям $\lambda_1, \lambda_2, \ldots, \lambda_\mathrm{m}$, упорядоченным по убыванию значений, и удовлетворяющие условию $\vec{\beta}^{(i)T}\vec{\beta}^{(i)} = 1/\lambda_i$. Векторы факторных нагрузок также являются собственными векторами матрицы ковариаций и связаны с векторами $\vec{\beta}^{(i)}$ соотношениями: $\vec{\alpha}^{(i)} = \lambda_i \vec{\beta}^{(i)}$. При этом дисперсия, объясняемая каждым фактором равна: $\sum_{i=1}^{k} D(\alpha_i^{(j)} f^{(j)}) = \left|\vec{\alpha}^{(j)}\right|^2 = \lambda_j,\ j = \overline{1,m}$. Характерные факторы при этом рассматриваются, как малозначимые величины, не оказывающие существенного влияния на исходные признаки. Такой подход оправдан в случае, когда обобщенные факторы объясняют большую часть дисперсии исходных признаков.

Оценка главных компонент на основе выборочных данных строится на основе выборочной матрицы ковариаций. Оценки собственных значений, являющиеся собственными числами выборочной матрицы ковариаций, в случае нормального распределения генеральной совокупности являются оценками максимального правдоподобия. Если единицы измерения исходных признаков различаются или их значения сильно различаются, то лучше использовать при нахождении оценок главных компонент вместо выборочной матрицы ковариаций выборочную корреляционную матрицу.

Координаты векторов факторных нагрузок $\alpha_i^{(j)}, i = \overline{1,k},\ j = \overline{1,m}$, являются ковариациями между исходными признаками ξ_i и нормированными факторами $f^{(j)}$. Поэтому по значениям координат векторов факторных нагрузок можно судить о степени влияния факторов на признаки (или, наоборот, о степени вклада каждого признака в фактор). Соответственно, можно попытаться провести классификацию факторов на основе признаков, вносящих наибольший вклад в факторы. Если факторные нагрузки имеют более или менее равномерное распределение, то задача интерпретации фактора усложняется. В этом случае целесообразно прибегать к вращению факторов. Так как обобщенные факторы определяются с точностью до ортогонального преобразования, то мы можем осуществлять ортогональное вращение факторного пространства, добиваясь такого расположения осей, при котором факторы допускают наиболее содержательную интерпретацию.

Пусть $X = \{(X_1^{(1)}, X_2^{(1)}, \ldots, X_k^{(1)}), (X_1^{(2)}, X_2^{(2)}, \ldots, X_k^{(2)}), \ldots, (X_1^{(n)}, X_2^{(n)}, \ldots, X_k^{(n)})\}$ - выборка объема n из k-мерной совокупности (матрица размеров $n \times k$).

1) Находим выборочную матрицу ковариаций $\overline{A} = \frac{1}{n}\widehat{X}^T \widehat{X}$, где $\widehat{X}$ - центрированная выборочная матрица (из элементов каждого столбца матрицы X вычитаем среднее

для этого столбца). Если исходные данные сильно различаются, следует использовать нормированную центрированную матрицу, то есть поделить элементы каждого столбца матрицы $\hat{X}$ на величину выборочного среднеквадратичного отклонения $\sqrt{D}$ для данного столбца.

2) Находим собственные значения λ_i и собственные векторы $\vec{\beta}^{(i)}$ матрицы $\overline{A}$, упорядоченные по убыванию собственных значений. Каждый из собственных векторов должен удовлетворять условию нормировки: $\vec{\beta}^{(i)T}\vec{\beta}^{(i)} = 1/\lambda_i$. Определяем также векторы факторных нагрузок $\vec{\alpha}^{(i)} = \lambda_i \vec{\beta}^{(i)}$.

3) Определяем количество главных факторов *m*, которые следует оставить в факторной модели. Критерии отбора главных факторов аналогичны критериям отбора главных компонент.

4) Находим значения обобщенных факторов $f^{(i)} = \vec{\beta}^{(i)T}\vec{\xi} = \beta_1^{(i)}\xi_1 + \beta_2^{(i)}\xi_2 + \ldots + \beta_k^{(i)}\xi_k$, $i = \overline{1,m}$ (точнее оценки значений) для каждого из n наблюдений. Матрицу значений Y главных компонент можно получить как: $Y = \hat{X}B$, где B - матрица столбцами которой являются векторы $\vec{\beta}^{(i)}$, $i = \overline{1,m}$.

Проводим интерпретацию обобщенных факторов.

Пусть $\alpha_{s_1}^{(p)}, \alpha_{s_2}^{(p)}, \ldots, \alpha_{s_q}^{(p)}$ факторные нагрузки фактора $f^{(p)}$, соответствующие признакам $\xi_{(s_1)}, \ldots, \xi_{(s_q)}$. Тогда, число $K_u = \dfrac{\sum_{j=1}^{q}\left(\alpha_{s_j}^{(p)}\right)^2}{\left\|\alpha^{(p)}\right\|^2}$ называется коэффициентом информативности признаков $\xi_{(s_1)}, \ldots, \xi_{(s_q)}$. Данное число определяет, какой вклад в признаки $\xi_{(s_1)}, \ldots, \xi_{(s_q)}$ вносит фактор $f^{(p)}$. Если можно выделить какую либо группу признаков, коэффициент информативности которой для фактора $f^{(p)}$ существенно выше коэффициента информативности оставшихся признаков, то первая группа признаков и будет определять данный фактор (возможно, например, что эти признаки можно классифицировать единым способом).

Если факторные нагрузки имеют более или менее равномерное распределение, то задача интерпретации фактора усложняется. В этом случае целесообразно прибегать к вращению факторов. Так как обобщенные факторы определяются с точностью до ортогонального преобразования, то мы можем осуществлять ортогональное вращение факторного пространства, добиваясь такого расположения осей, при котором факторы допускают наиболее содержательную интерпретацию. Обычно используют метод варимакс, при этом при вращении максимизируют величину: $V=\sum_{j=1}^{m}\frac{1}{k}\sum_{i=1}^{k}\left(\left(\alpha_i^{(j)}\right)^2-\frac{1}{k}\sum_{s=1}^{k}\left(\alpha_s^{(j)}\right)^2\right)^2$, которая характеризует различие столбцов матрицы факторных нагрузок (по сути V - это суммарная по всем векторам $\vec{\alpha}^{(j)}$ выборочная дисперсия квадратов координат этих векторов).

Если C - матрица поворота, то в новой системе координат матрица факторных нагрузок определится как: $\alpha'=\alpha C$. Обычно в процедуре вращения осуществляют последовательное попарное ортогональное вращение осей. В этом случае полная матрица преобразования будет равна произведению матриц попарных поворотов. Матрицу поворота для двухфакторной модели матрицу поворота можно записать, например, как функцию от угла поворота: $C=C(\varphi)=\begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}$ (в случае вращения против часовой стрелки). Таким образом, для двухфакторной модели, мы должны найти φ, при котором величина $V(\alpha')=V(\alpha C(\varphi))$ достигает максимума. Аналогично, для m-факторной модели, мы должны найти последовательно C_m^2 углов попарных поворотов факторных осей.