5 Метод главных компонент в факторном анализе.

Мы будем работать с моделью факторного анализа: $\xi = \alpha f + \varepsilon$. Будем предполагать, что обощенные факторы $f^{(1)}, ..., f^{(m)}$ нормированы, а вектор исходных признаков ξ центрирован, то есть $E\xi_i = 0, i = \overline{1,k}$, $Ef^{(i)} = 0, Df^{(i)} = 1$ и ковариационная матрица вектора исходных признаков ξ равна A. Согласно методу главных компонент полагаем, что обощенные факторы выражаются в виде линейных комбинаций компонент вектора исходных признаков:

$$f^{(i)} = (\beta^{(i)T})\xi = \beta_1^{(i)}\xi_1 + \beta_2^{(i)}\xi_2 + ... + \beta_k^{(i)}\xi_k, \quad i = \overline{1,m}, m < k.$$

Отсюда

$$Df^{(i)} = E(\beta^{(i)T}\xi\xi^T\beta^{(i)}) = \beta^{(i)T}A\beta^{(i)}$$

Найдем первый фактор из соображений, чтобы первый фактор вносил наибольший вклад в суммарную дисперсию исходных признаков, то есть будем решать следующую задачу оптимизации:

$$Q = \sum_{j=1}^{k} D(\xi_j - \alpha_j^{(1)} f^{(1)}) = \sum_{j=1}^{k} E(\xi_j - \alpha_j^{(1)} f^{(1)})^2 = \sum_{j=1}^{k} E\xi_j^2 - 2\sum_{j=1}^{k} \alpha_j^{(1)} E(\xi_j f^{(1)}) +$$

$$+ \sum_{j=1}^{k} (\alpha_j^{(1)})^2 E(f^{(1)})^2 = trA - 2\sum_{j=1}^{k} \alpha_j^{(1)} \sum_{s=1}^{k} E(\xi_j \beta_s^{(1)} \xi_s) + \sum_{j=1}^{k} (\alpha_j^{(1)})^2 = trA - 2\alpha^{(1)T}A\beta^{(1)} + \alpha^{(1)T}\alpha^{(1)}.$$

Таким образом, необходимо решить задачу на условный экстремум:

$$Q = trA - 2\alpha^{(1)T}A\beta^{(1)} + \alpha^{(1)T}\alpha^{(1)} \to \min_{\alpha_1}$$

при условии

$$Df^{(1)} = \beta^{(1)T}A\beta^{(1)} = 1.$$

Составляем функцию Лагранжа:

$$L(\alpha_1, \beta_1, \lambda) = trA - 2\alpha^{(1)T}A\beta^{(1)} + \alpha^{(1)T}\alpha^{(1)} + \lambda(\beta^{(1)T}A\beta^{(1)} - 1).$$

Получаем систему :

$$\begin{cases} \frac{\partial L}{\partial \alpha^{(1)}} & = 2\alpha^{(1)} - 2A\beta^{(1)} = 0 \\ \frac{\partial L}{\partial \lambda} & = \beta^{(1)T}A\beta^{(1)} - 1 = 0 \\ \frac{\partial L}{\partial \beta^{(1)}} & = -2A\alpha^{(1)} + 2\lambda A\beta^{(1)}. \end{cases}$$

Так как $det(A) \neq 0$, то получаем:

$$\begin{cases} \alpha^{(1)} = A\beta^{(1)} \\ \beta^{(1)T}A\beta^{(1)} = 1 \\ \alpha^{(1)} = \lambda\beta^{(1)}. \end{cases}$$

Из первого и третьего уравнения получаем, что $A\beta^{(1)} = \lambda\beta^{(1)}$, то есть $\beta^{(1)}$ - собственный вектор матрицы коварицй вектора исходных признаков ξ, соответствующий собственному значению λ. Используя второе уравнение, получим:

$$1 = \beta^{(1)T} A\beta^{(1)} = \beta^{(1)T}\lambda\beta^{(1)} \quad \Rightarrow \quad \lambda|\beta^{(1)}|^2 = 1.$$

Тогда $|\beta^{(1)}|^2 = \dfrac{1}{\lambda}$. Так как $\alpha^{(1)} = \lambda\beta^{(1)}$, то $\alpha^{(1)}$ также является собственным вектором матрицы A. Найдем собственное значение, соответствующее собственным векторам:

$$\alpha^{(1)} = \lambda\beta^{(1)} \quad \Rightarrow \quad |\alpha^{(1)}|^2 = \lambda^2|\beta^{(1)}|^2 \quad \Rightarrow \quad |\alpha^{(1)}|^2 = \lambda^2\frac{1}{\lambda}$$

Отсюда получаем $\lambda_1 = |\alpha^{(1)}|^2$ собственное значение. Тогда

$$Q = trA - 2\alpha^{(1)T}A\beta^{(1)} + \alpha^{(1)T}\alpha^{(1)} = trA - 2\alpha^{(1)T}\alpha^{(1)} + \alpha^{(1)T}\alpha^{(1)} = trA - \alpha^{(1)T}\alpha^{(1)} = trA - \lambda.$$

Таким образом, функция Q достигает своего минимума при наибольшем собственном значении λ для ковариационной матрицы вектора исходных признаков. Теперь по найденному собственному значению λ_1, можно восстановить вектор коэффициентов приведения к главным компонентам $\beta^{(1)}$ и вектор матрицы факторных нагрузок $\alpha^{(1)}$.

Так как

$$D\left(\sum_{j=1}^{k}\alpha_j^{(1)}f^{(1)}\right) = \sum_{j=1}^{k}\left(\alpha_j^{(1)}\right)^2 Df^{(1)} = \sum_{j=1}^{k}\left(\alpha_j^{(1)}\right)^2 = \lambda_1.$$

То есть число λ_1 показывает долю дисперсии вектора исходных признаков ξ, которая обясняется первым обощенным фактором $f^{(1)}$, и

$$Q = trA - \lambda = \sum_{j=1}^{k} D\xi_j - \sum_{j=1}^{k} D(\alpha_j^{(1)}f^{(1)}).$$

Вычислим ковариации между компонентами вектора исходных признаком и первым фактором:

$$cov(\xi_j, f^{(1)}) = E\left(\xi_j\, f^{(1)}\right) = E\left(\xi_j\, \sum_{i=1}^{k}\beta_i^{(1)}\xi_i\right) = \sum_{i=1}^{k}\beta_i^{(1)}E(\xi_j\,\xi_i)$$
$$= \sum_{i=1}^{k}\beta_i^{(1)}a_{ij} = \alpha_j^{(1)}, \quad \text{так как} \quad \alpha^{(1)} = A\beta^{(1)}.$$

Для нахождения второго фактора $f^{(2)}$ необходимо уже минимизировать функцию

$$Q = \sum_{j=1}^{k} D(\xi_j - \alpha_j^{(1)}f^{(1)} - \alpha_j^{(2)}f^{(2)})$$

по коэффициентам $\alpha_j^{(2)}$ при условиях:

$$Df^{(2)} = \beta^{(2)T}A\beta^{(2)} = 1.$$

и условии некоррелированности факторов:

$$cov(f^{(1)}, f^{(2)}) = \beta^{(1)T} A \beta^{(2)} = 0.$$

Решая аналогично первому шагу задачу на условный экстремум , можно показать, что:

$$\begin{cases} A\beta^{(2)} = \lambda\beta^{(2)} \\ \alpha^{(2)} = \lambda\beta^{(2)} \end{cases}$$

Причем $Q = trA - \lambda_1 - \lambda$, то есть минимум будет достигаться $\lambda = \lambda_2$ - наибольшем собственном значении матрицы ковариаций вектора ξ, не учитывая число λ_1. По найденному собственному значению можно также восстановить собственные векторы, соответствующие ему : $\alpha^{(2)}, \beta^{(2)}$ и показать, что

$$cov(\xi_j, f^{(2)}) = \alpha_j^{(2)}.$$

Таким образом, продолжая этот процесс далее , получаем, что

$$cov(\xi_j, f^{(i)}) = \alpha_j^{(i)}$$

и $|\alpha_j^{(1)}|^2 = \lambda_i$ - доля дисперсии вектора исходных признаков, объясняемая i -ым фактором.

Для сокращения количества обобщенных факторов $(m < k)$ используется те же критерии, что и в методе главных компонент : критерий Кайзера, критерий, основанный на доле суммарной дисперсии, приходящейся на первые факторы, критерий Кэттела и др.

Замечание 5.1. *Так как матрица ковариаций на практике неизвестна, то находится ее оценка на основе выборочных данных, и для выборочной ковариационной матрицы находятся собственные векторы и собственные значения, которые будут являться оценками собственных значений и собственных векторов ковариационной матрицы A.*

6 Канонические методы факторного анализа

Пусть $A = M(\xi\xi^T)$ - ковариационная матрица исходных признаков, а $\Sigma = M(\varepsilon\varepsilon^T)$ - диагональная матрица ковариаций характерных факторов, тогда:

$$A = M((\alpha f + \varepsilon)(\alpha f + \varepsilon)^T) = \alpha M(ff^T)\alpha^T + \Sigma = \alpha\alpha^T + \Sigma \qquad (30)$$

или:

$$\begin{cases} cov(\xi_i, \xi_j) = \sum_{s=1}^m \alpha_i^{(s)}\alpha_j^{(s)}, \quad i = \overline{1,k}, \quad j = \overline{1,k}, \quad i \neq j \\ D(\xi_i) = \sum_{s=1}^m \left(\alpha_i^{(s)}\right)^2 + D(\varepsilon_i), \quad i = \overline{1,k} \end{cases} \qquad (31)$$