ЛАБОРАТОРНАЯ РАБОТА «ФАКТОРНЫЙ АНАЛИЗ»

ОСНОВНЫЕ ПОЛОЖЕНИЯ ФАКТОРНОГО АНАЛИЗА

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) общих факторов $f^{(1)}, f^{(2)},\ldots, f^{(m)}$ и характерного фактора ε_j. При этом считается, что каждый общий 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 общие 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:

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

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

$f=(f^{(1)}, f^{(2)},\ldots, f^{(m)})$ - вектор латентных (обобщенных) факторов, $m<k$;

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

$\varepsilon=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_k)$ - вектор характерных факторов.

В ходе факторного анализа необходимо оценить минимальное число факторов, определить векторы факторных нагрузок, и вычислить значения факторов для каждого наблюдаемого объекта. Поскольку число обобщенных факторов предполагается существенно меньше числа исходных признаков, данная задача не имеет однозначного решения. 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 ковариацию между исходными факторами. Первый подход в факторном анализе основан на выделении главных компонент, и, соответственно, называется методом главных компонент. Методы же факторного анализа, базирующиеся на втором подходе, собственно и принято называть 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. 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 обобщенные факторы $f^{(1)}, f^{(2)}, \ldots, f^{(m)}$ являются линейными комбинациями исходных признаков $\xi_1, \xi_2, \ldots, \xi_k$:

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

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

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

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 некоррелированности 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. Коэффициенты преобразования $\alpha^{(j)}$, $\beta^{(i)}$,

$j=\overline{1,k}$, удовлетворяющие условиям задачи поиска обобщенных факторов (или главных компонент) являются собственными векторами матрицы ковариаций A, соответствующими собственным значениям $\lambda_1,\lambda_2,...,\lambda_k$ упорядоченным по убыванию значений, и удовлетворяющие условиям нормировки: $\alpha^{(j)T}\alpha^{(j)}=\lambda_j$, $\beta^{(j)T}\beta^{(j)}=1/\lambda_j$. При этом, дисперсия объясняемая каждым фактором равна: $\sum_{i=1}^{k} D(\alpha_i^{(j)} f^{(j)}) = \left|\alpha^{(j)}\right|^2 = \lambda_j,\ j=\overline{1,m}$, а координаты векторов факторных нагрузок есть ковариации между исходными признаками и главными компонентами: $\operatorname{cov}(\xi_j, f^{(s)}) = \alpha_j^{(s)}$.

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.

Пусть $X=\{(X_1^{(1)},X_2^{(1)},\dots,X_k^{(1)}),(X_1^{(2)},X_2^{(2)},\dots,X_k^{(2)}),\dots,(X_1^{(n)},X_2^{(n)},\dots,X_k^{(n)})\}$ - 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 $\widehat{X}$ на величину выборочного среднеквадратичного отклонения $\sqrt{D}$ для данного столбца.
2) Находим собственные значения λ_j матрицы $\overline{A}$, упорядоченные по убыванию собственных значений и соответствующие собственные векторы

$\alpha^{(j)}$, $\beta^{(i)}$, $j=\overline{1,k}$. Собственные векторы должны удовлетворять условиям нормировки: $\alpha^{(j)T}\alpha^{(j)}=\lambda_j$, $\beta^{(j)T}\beta^{(j)}=1/\lambda_j$.

3) Определяем количество главных факторов, которые следует оставить в факторной модели. Критерии отбора главных факторов аналогичны критериям отбора главных компонент.
4) Находим значения обобщенных факторов $\eta_i=\beta^{(i)T}\xi=\beta_1^{(i)}\xi_1+\beta_2^{(i)}\xi_2+\ldots+\beta_k^{(i)}\xi_k$, $i=\overline{1,m}$ (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 $\beta^{(i)}$, $i=\overline{1,m}$.

КАНОНИЧЕСКАЯ МОДЕЛЬ ФАКТОРНОГО АНАЛИЗА

Пусть:

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

где:

$\xi=(\xi^{(1)},\xi^{(2)},\ldots,\xi^{(k)})$ - вектор центрированных исходных признаков;

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

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

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

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

$$A=\alpha\alpha^T+\Sigma \qquad (2)$$

или:

$$\begin{cases} \operatorname{cov}(\xi_i,\xi_j)=\sum_{s=1}^{m}\alpha_i^{(s)}\alpha_j^{(s)}, \quad i=\overline{1,k}, \quad j=\overline{1,k}, \quad i\neq j \\ D(\xi_i)=\sum_{s=1}^{m}\left(\alpha_i^{(s)}\right)^2+D(\varepsilon_i), \quad i=\overline{1,k} \end{cases} \qquad (3)$$

Таким образом, ковариации исходных признаков полностью воспроизводятся матрицей нагрузок, а для воспроизведения их дисперсий нужны также дисперсии характерных факторов. Заметим, что ,

$$\operatorname{cov}(\xi_j, f_s) = \alpha_j^{(s)}, \quad j = \overline{1,k}, \quad s = \overline{1,m}.$$

Модель (1) при условиях (3) носит название канонической модели факторного анализа. Заметим, что если решение системы (3) существует, то оно определяется с точностью до ортогонального преобразования матрицы факторных нагрузок.

ОЦЕНИВАНИЕ ПАРАМЕТРОВ МОДЕЛИ ФАКТОРНОГО АНАЛИЗА МЕТОДОМ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Будем полагать, что обобщенные и характерные факторы имеют нормальное распределение (это соответственно означает, что исходные признаки имеют многомерное нормальное распределение). Для того, чтобы матрица факторных нагрузок определялась однозначно потребуем, чтобы она удовлетворяла условию: $\alpha^T \Sigma^{-1} \alpha = J$, где J - диагональная матрица с упорядоченными по убыванию диагональными элементами. Заметим, что количество условий, наложенных на параметры равно $k(k+1)/2 + m(m-1)/2$, а число неизвестных параметров матрицы факторных нагрузок и матрицы характерных факторов равно $k(m+1)$. Чтобы задача не являлась неопределенной необходимо, чтобы $k(k+1)/2 + m(m-1)/2 \ge k(m+1)$ или $(k-m)^2 \ge (k+m)$. Поскольку, при $(k-m)^2 = (k+m)$ решение теряет статистическую значимость, то максимальное число параметров для данной модели должно удовлетворять условию $(k-m)^2 > (k+m)$. Например, при $k = 5, k = 6$ получаем, что $m \le 2$.

Для нахождения оценок матрицы факторных нагрузок и дисперсионной матрицы характерных факторов используем метод максимального правдоподобия, согласно которому можно получить следующие уравнения для оценок:

$$diag\left(\hat{A} - \overline{A}\right) = 0 \qquad \text{или} \qquad \overline{a}_{ii} = \sum_{s=1}^{m} \left(\hat{\alpha}_i^{(s)}\right)^2 + \hat{v}_i, \quad i = \overline{1,k} \tag{1}$$

$$\hat{\alpha}\hat{J} = (\overline{A} - \hat{\Sigma})\hat{\Sigma}^{-1}\hat{\alpha} \tag{2}$$

Поскольку $\hat{J}$ диагональная матрица, то из уравнения (2) вытекает, что оценки векторов факторных нагрузок $\alpha^{(i)}$ должны являться собственными векторами матрицы $(\overline{A}-\hat{\Sigma})\hat{\Sigma}^{-1}$, удовлетворяющими условию:

$$\hat{\alpha}^{(i)T}\hat{\Sigma}^{-1}\hat{\alpha}^{(i)}=\lambda_i. \qquad (3)$$

При заданном числе факторов m итерационная процедура решения системы

$$\begin{cases} J\alpha=\Sigma^{-1}(\overline{A}-\Sigma)\alpha \\ diag\overline{A}=diag(\alpha\alpha^T+\Sigma) \end{cases}$$

может быть построена следующим образом (для упрощения записи дальнейшем знаки оценок для α, Σ, J опущены):

Задаемся начальным приближением Σ_0 для Σ. Можно взять, например, в качестве диагональных элементов Σ_0 величины $(1-|\bar{\mathrm{r}}(\mathrm{i})|)\overline{\mathrm{D}}_{\mathrm{i}}$, где $r(i)$ наибольшая по абсолютной величине выборочная корреляция для i-го признака, Можно просто задать значения в долях от диагональных элементов выборочной матрицы $\overline{A}$, т.е. положить $v_{i0}=\gamma\overline{D}_i, i=\overline{1,k}$, где для γ можно рекомендовать значения в интервале (0,1; 0,5). Желательно, чтобы собственные числа матрицы $\overline{A}-\Sigma_0$ получились неотрицательными (хотя допустимо, что последние $k-m$ собственных значений в процессе итераций будут принимать отрицательные значения). Далее находим m собственных чисел и собственных векторов матрицы $(\overline{A}-\Sigma)\Sigma^{-1}$, упорядоченных по убыванию собственных чисел, нормируем их, использую условие (10) и, соответственно, получаем первое приближение для матрицы факторных нагрузок α. Далее находим следующее приближение для Σ: $\mathrm{diag}\Sigma_1=\mathrm{diag}(\overline{\mathrm{A}}-\alpha\alpha^{\mathrm{T}})$ и процесс повторяется. Итерационный процесс заканчивают, когда очередное приближение матрицы Σ мало отличается от предыдущего, то есть, если $\|\Sigma_i-\Sigma_{i-1}\|<\varepsilon$, где ε - заранее заданное число. В качестве нормы можно использовать, например, абсолютную величину следа матриц. Гипотезу о возможности представления исходного вектора признаков с помощью модели факторного анализа с числом факторов равным m можно проверить, используя следующую статистику:

$$\eta = -\left(n - \frac{1}{6}(2k+5) - \frac{2}{3}m\right)\left(\ln\frac{|\overline{A}|}{|\alpha\alpha^T + \Sigma|} - Sp(A(\alpha\alpha^T + \Sigma)^{-1} + k\right).$$

При истинности H_0: допустимо представление исходных признаков в виде m-факторной модели, статистика приближенно имеет распределение χ^2 с числом степеней свободы $\nu = \frac{1}{2}((k-m)^2 - (k+m))$.

КЛАССИФИКАЦИЯ ФАКТОРОВ. ВРАЩЕНИЕ ФАКТОРОВ

После определения факторов исследователю зачастую требуется оценить уровень информативности или вклад фактора в дисперсию каждого признака, а также выяснить, можно ли интерпретировать разумным образом полученные факторы и как это сделать.

Пусть имеется m-факторная модель. Пусть $f^{(p)}$ – некоторый фактор, $p \le m$. Общностью фактора $f^{(p)}$ (или вкладом фактора в суммарную дисперсию) признаков называется число $V^{(p)} = \sum_{j=1}^{k}\left[\alpha_j^{(p)}\right]^2$. Величину $V_0 = \sum_{p=1}^{m} V^{(p)}$ называют суммарной общностью факторов $f^{(1)},\dots,f^{(m)}$. Отношение $V^{(p)}/V_0$ называется долей фактора $f^{(p)}$ в суммарной общности.

Пусть $\alpha_{s_1}^{(p)},\alpha_{s_2}^{(p)},\dots,\alpha_{s_l}^{(p)}$ факторные нагрузки фактора $f^{(p)}$ соответствующие признакам $\xi^{(s_1)},\dots,\xi^{(s_l)}$. $\alpha^{(p)}$. Тогда число

$$K_u = \frac{\sum_{j=1}^{l}\left(\alpha_{s_j}^{(p)}\right)^2}{\left\|\alpha^{(p)}\right\|^2}$$

называется коэффициентом информативности признаков $\xi^{(s_1)},\dots,\xi^{(s_l)}$. Данное число определяет, какой вклад векторы $\xi^{(s_1)},\dots,\xi^{(s_l)}$ вносят в 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 название фактора. Если факторные нагрузки имеют более или менее равномерное распределение, то задача интерпретации фактора усложняется. В этом случае целесообразно прибегать к вращению факторов. Вспомним, что обобщенные факторы определяются с точностью до ортогонального преобразования. Следовательно, мы можем осуществлять ортогональное вращение факторного пространства, добиваясь такого расположения осей, при котором факторы допускают наиболее содержательную интерпретацию. Обычно используют следующие методы (стратегии) вращения, облегчающие интерпретацию факторов:

метод варимакс, 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;

метод квартимакс, при вращении максимизируют величину:

$$Q = \sum_{i=1}^{k} \frac{1}{m} \sum_{j=1}^{m} \left(\left(\alpha_i^{(j)}\right)^2 - \frac{1}{m} \sum_{s=1}^{m} \left(\alpha_i^{(s)}\right)^2 \right)^2,$$

которая характеризует различие строк матрицы факторных нагрузок.

Учитывая, что метод варимакс основан на упрощении столбцов, а квартимакс на упрощении строк матрицы факторных нагрузок, можно построить обобщенный метод, основанный на максимизации величины $\alpha Q + \beta V$. В частности, при $\alpha = \beta$ получим критерий, используемый в методе вращения биквартимакс.

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. Например, для трехфакторной модели:

$$C = C_{12} C_{23} C_{31} = \begin{pmatrix} \cos\varphi_1 & -\sin\varphi_1 & 0 \\ \sin\varphi_1 & \cos\varphi_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_2 & -\sin\varphi_2 \\ 0 & \sin\varphi_2 & \cos\varphi_2 \end{pmatrix} \cdot \begin{pmatrix} \sin\varphi_3 & 0 & \cos\varphi_3 \\ 0 & 1 & 0 \\ \cos\varphi_3 & 0 & -\sin\varphi_3 \end{pmatrix}$$

(в случае вращения против часовой стрелки).

Для того, чтобы избежать влияния на результаты вращения переменных с большой общностью при вращении применяют также нормировку факторных нагрузок, то есть, делят координаты векторов факторных нагрузок на соответствующие корни из общностей.

ОЦЕНИВАНИЕ ЗНАЧЕНИЙ ФАКТОРОВ

После того как найдены оценки параметров факторной модели, как правило, необходимо вычислить оценки значений обобщенных факторов для каждого наблюдения. Эти значения могут быть использованы, например, в задачах классификации или в задачах регрессионного анализа.

Если для факторного анализа использовался метод главных компонент, то задача оценивания значений факторов решается просто: согласно данному методу главные факторы выражаются через линейную комбинацию исходных признаков:

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

где векторы $\beta^{(j)}$, $j = \overline{1,m}$ - собственные векторы матрицы ковариаций A, упорядоченные по убыванию собственных значений λ_i и удовлетворяющие условию $|\beta^{(i)}|^2 = 1/\lambda_i$. Пусть X матрица значений исходных признаков размеров $n \times k$, B - матрица, составленная из векторов $\beta^{(j)}$, $j = \overline{1,m}$ (точнее, из оценок этих векторов) размеров $k \times m$, F - матрица оценок значений обобщенных факторов размеров $n \times m$, тогда:

$$F = XB.$$

Если, кроме того, использовалась процедура вращения факторов с матрицей преобразования C, то

$$F = XBC.$$

Для оценивания значений факторов канонической модели факторного анализа используют либо метод Бартлетта, либо метод Томпсона.

Согласно методу Бартлетта модель для каждого наблюдения $X^{(i)}, i = \overline{1,n}$ рассматривается как регрессия величины $X^{(i)}$ на факторы $\alpha^{(j)}, j = \overline{1,m}$ с неизвестными коэффициентами $f_i^{(j)}$:

$$X^{(i)} = f_i^{(1)}\alpha^{(1)} + f_i^{(2)}\alpha^{(2)} + \ldots + f_i^{(m)}\alpha^{(m)} + \varepsilon.$$

Соответственно для значения $X_j^{(i)}$ признака ξ_j в i-ом наблюдении справедливо:

$$X_j^{(i)} = f_i^{(1)}\alpha_j^{(1)} + f_i^{(2)}\alpha_j^{(2)} + \ldots + f_i^{(m)}\alpha_j^{(1)} + \varepsilon_j, \quad j = \overline{1,k}.$$

Заметим, что наблюдения неравноточные, характерные факторы в общем случае имеют различные дисперсии $D(\varepsilon_j) = v_j$. Следовательно, для оценки факторов следует применять обобщенный метод наименьших квадратов, согласно которому:

$$\hat{\mathrm{f}}_{\mathrm{i}} = (\hat{\mathrm{f}}_{\mathrm{i}}^{(1)}, \hat{\mathrm{f}}_{\mathrm{i}}^{(2)}, \ldots, \hat{\mathrm{f}}_{\mathrm{i}}^{(\mathrm{m})})^{\mathrm{T}} = (\alpha^{\mathrm{T}}\Sigma^{-1}\alpha)^{-1}\alpha^{\mathrm{T}}\Sigma^{-1}\mathrm{X}^{(\mathrm{i})\mathrm{T}}.$$

Последнее выражение удобнее переписать в виде:

$$\hat{f}^{(i)T} = X^{(i)}\Sigma^{-1}\alpha(\alpha^T\Sigma^{-1}\alpha)^{-1},$$

тогда матрица оценок значений факторов:

$$\hat{\mathrm{f}}_{\mathrm{i}}^{\mathrm{T}} = \mathrm{X}^{(\mathrm{i})}\Sigma^{-1}\alpha(\alpha^{\mathrm{T}}\Sigma^{-1}\alpha)^{-1}.$$

В методе Томпсона, исходят из того, что оценки обобщенных факторов есть линейная комбинация исходных признаков, коэффициенты которой определяются из условия минимума среднеквадратичного отклонения оценки от фактора, то есть:

$$\hat{f}^{(j)} = \beta^{(j)T}\xi = \beta_1^{(j)}\xi_1 + \beta_2^{(j)}\xi_2 + \ldots + \beta_k^{(j)}\xi_k, \quad j = \overline{1,m},$$

где вектор коэффициентов удовлетворяет условию:

$$M(f^{(j)} - \hat{f}^{(j)})^2 \to \min.$$

Таким образом, коэффициенты $\beta_i^{(j)}, i = \overline{1,k}$ является коэффициентами линейной регрессии и могут быть выражены через ковариации величин $\xi_1, \xi_2, \ldots, \xi_k$ и $f^{(j)}$:

$$\beta^{(j)} = A^{-1}(\xi)M(f^{(j)} \cdot \xi),$$

здесь: A - матрица ковариаций величин $\xi_1, \xi_2, \ldots, \xi_k$, $M(f^{(j)} \cdot \xi)$ - вектор ковариаций величины $f^{(j)}$ и величин $\xi_1, \xi_2, \ldots, \xi_k$.

Соответственно

$$\hat{f}^{(j)} = \xi^T A^{-1}(\xi)M(f^{(j)} \cdot \xi).$$

Заменяя $A(\xi)$ и $M(f^{(j)} \cdot \xi)$ их оценками $\hat{A}(\xi) = \alpha\alpha^T + \Sigma$ и $\hat{M}(f^{(j)} \cdot \xi) = \alpha^{(j)}$ получим оценку для $\hat{f}^{(j)}$:

$$\tilde{f}^{(j)} = \xi^T(\alpha\alpha^T + \Sigma)^{-1}\alpha^{(j)}.$$

Соответственно оценка для вектора обобщенных факторов:

$$\tilde{f}^T = \xi^T(\alpha\alpha^T + \Sigma)^{-1}\alpha .$$

Пусть X матрица значений исходных признаков размеров $n \times k$, F - матрица оценок значений обобщенных факторов размеров $n \times m$, тогда:

$$F = X(\alpha\alpha^T + \Sigma)^{-1}\alpha .$$

Задания к лабораторной работе.

Вариант 1. Базовый: выявление ключевых факторов креативности в организации

Цель: определить основные скрытые факторы, влияющие на уровень креативности сотрудников в компании.

Данные для анализа: Анкетирование 50–100 сотрудников по 12 показателям (шкала Ликерта от 1 до 5):

- свобода самовыражения;
- поддержка руководства;
- наличие времени на творческие задачи;
- доступ к ресурсам (бюджет, инструменты);
- командная коллаборация;
- поощрение инноваций;
- риск-толерантность;
- обучение и развитие;
- внутренняя мотивация;
- внешняя мотивация (премии);
- психологическая безопасность;
- гибкость процессов.

Задачи:

1. Провести эксплораторный факторный анализ (EFA) с методом главных компонент (PCA).
2. Определить оптимальное число факторов (критерий Кайзера, график «каменистой осыпи»).
3. Выполнить вращение (Varimax) для интерпретации факторов.

4. Описать выявленные факторы (например: «Организационная поддержка», «Внутренняя мотивация», «Ресурсная база»).
5. Рассчитать факторные нагрузки и коммунальности для каждой переменной.

Отчёт:

- Таблица факторных нагрузок (после вращения).
- График «каменистой осыпи».
- Интерпретация 3–4 выделенных факторов.
- Краткие рекомендации для менеджмента (например: усилить поддержку руководства для роста креативности).

Вариант 2. Сравнительный: факторы креативности в разных типах организаций

Цель: сравнить структуру факторов креативности в компаниях разного профиля (IT-стартап vs. традиционное производство).

Данные:

Две выборки по 60 респондентов:

- Группа 1: IT-стартапы (гибкие процессы, flat hierarchy).
- Группа 2: промышленные предприятия (иерархия, регламенты).

Анкета — та же (12 показателей, шкала Ликерта).

Задачи:

1. Провести отдельный EFA для каждой группы (PCA + Varimax).
2. Сравнить число и содержание факторов между группами.
3. Проверить инвариантность факторной структуры (например, через конгруэнтность факторов).
4. Выявить уникальные и общие факторы для двух типов организаций.
5. Проанализировать различия в факторных нагрузках (например, «гибкость процессов» критично для IT, но не для производства).

Отчёт:

- Две таблицы факторных нагрузок (по группам).
- Сравнение факторов (таблица: название, переменные, нагрузки).

- Выводы: как организационный контекст меняет драйверы креативности.
- Рекомендации: адаптировать менеджмент креативности под тип компании.

Общие требования к оформлению отчёта

1. **Введение:** цель, гипотезы, обоснование выбора метода.
2. **Методика:** описание выборки, инструментария, шагов анализа (EFA, вращение, регрессия).
3. **Результаты:** таблицы, графики, интерпретация факторов.
4. **Обсуждение:** связь с теорией менеджмента креативности, ограничения.
5. **Заключение:** выводы и рекомендации.
6. **Приложения:** сырые данные (анонимно), код анализа (если используется R/Python).

Софт для анализа: SPSS, R (пакеты psych, factoextra), Python (sklearn, factor_analyzer).

Список литературы

1. Айвазян С.А., Мхитарян В.С. Теория вероятностей и прикладная статистика. М.: ЮНИТИ-ДАНА, т.1, 2001.
2. Дубров А.М. Многомерные статистические методы для экономистов и менеджеров. М.: Финансы и статистика, 2003.
3. Сошникова Л.А., Тамашевич В.Н., Уебе Г., Шефер М. Многомерный статистический анализ в экономике. М.: ЮНИТИ-ДАНА, 1999
4. Иберла К. Факторный анализ. М.: Статистика, 1980.
5. Лоули Д. Факторный анализ как статистический метод. М: Мир, 1967.