где:

$$S = \frac{1}{n-q}\sum_{i=1}^{q}(n_i - 1)S_i$$

$$S_0 = \frac{1}{n-q}\hat{X}^T\hat{X} = S + \frac{1}{n-q}\sum_{i=1}^{q} n_i(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T.$$

Здесь $\hat{X}$- центрированная матрица объединенной выборки, $\bar{X}$ - вектор выборочных средних объединенной выборки, $\bar{X}_i$ - вектор выборочных средних i-ой выборки.

Если положить

$$\rho = 1 - \frac{p-q+2}{2(n-q)},$$

то статистика $\eta = -2\rho \ln \lambda$ будет иметь распределение χ^2 с $\nu = p(q-1)$ степенями свободы.

Гипотеза однородности многомерных нормальных совокупностей.

Пусть имеются выборки $X^{(1)}, X^{(2)}, \ldots, X^{(q)}$ объемов n_1, n_2, $\ldots$,n_q соответственно из p-мерных нормальных совокупностей $N(\vec{\mu}_1, \Sigma_1)$, $N(\vec{\mu}_2, \Sigma_2)$, $\ldots$, $N(\vec{\mu}_q, \Sigma_q)$. Проверяется гипотеза $H_0 : \vec{\mu}_1 = \vec{\mu}_2 = \ldots = \vec{\mu}_q, \quad \Sigma_1 = \Sigma_2 = \ldots = \Sigma_q$.

Отношение правдоподобия для данной гипотезы:

$$\lambda = \frac{\prod\limits_{i=1}^{q} |S_i|^{\frac{n_i-1}{2}}}{|S_0|^{\frac{n-q}{2}}},$$

где

$$S = \frac{1}{n-q}\sum_{i=1}^{q}(n_i - 1)S_i$$

$$S_0 = \frac{1}{n-q}\hat{X}^T\hat{X} = S + \frac{1}{n-q}\sum_{i=1}^{q} n_i(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T.$$

Если положить (Т. Андерсон. Введение в многомерный статистический анализ):

$$\rho = 1 - \left(\sum_{i=1}^{q}\frac{1}{n_i - 1} - \frac{1}{n-q}\right)\frac{2p^2 + 3p - 1}{6(p+3)(q-1)} - \frac{1}{n-q}\frac{p-q+2}{p+3},$$

то статистика $\eta = -2\rho \ln \lambda$ будет иметь распределение χ^2 с числом степеней свободы

$$\nu = \frac{1}{2}(q-1)p(p+3).$$

Гипотеза о независимости множеств случайных величин

Пусть p-мерный нормальный вектор $\vec{\xi}$ разбит на q подвекторов $\vec{\xi}_1, \vec{\xi}_2, \ldots, \vec{\xi}_q$ размерности $k_1, k_2, \ldots, k_q$ соответственно. Требуется по выборке X объема n из значений вектора $\vec{\xi}$ проверить гипотезу

$$H_0 : \vec{\xi}_1, \vec{\xi}_2, \ldots, \vec{\xi}_q \text{ взаимно независимы,}$$

то есть

$$f_{\vec{\xi}} = f_{\vec{\xi}_1} \cdot f_{\vec{\xi}_2} \cdot \ldots \cdot f_{\vec{\xi}_q}$$

Если гипотезу H_0 сформулировать на основе вида ковариационной матрицы, то это означает, что матрица ковариаций имеет блочно-диагональный вид:

$$\Sigma = \begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & A_{qq} \end{pmatrix},$$

где Σ_{ii} матрица ковариаций подвектора $\vec{\xi}_i$.

Для данной гипотезы отношение правдоподобия:

$$\lambda = \frac{|S|^{\frac{n}{2}}}{\prod\limits_{i=1}^{q} |S_{ii}|^{\frac{n}{2}}},$$

где S_{ii} - выборочные матрицы ковариаций i-го подвектора , а S - выборочная ковариационная матрица вектора ξ.

Если положить (Т. Андерсон. Введение в многомерный статистический анализ, стр.326-327):

$$\rho = 1 - \frac{2(p^3 - \sum\limits_{i=1}^{q} k_i^3) + 9(p^2 - \sum\limits_{i=1}^{q} k_i^2)}{6n(p^2 - \sum\limits_{i=1}^{q} k_i^2)},$$

то статистика $\eta = -2\rho \ln \lambda$ будет иметь в асимптотике χ^2 распределение с $\nu = \frac{1}{2}(p^2 - \sum_{i=1}^{q} k_i^2)$ степенями свободы.

В случае $q = p$ получаем критерий для проверки гипотезы о независимости компонент вектора $\vec{\xi}$ или в терминах матрицы ковариаций это означает, что она имеет диагональный вид. В этом случае:

$$\lambda = \frac{|S|^{\frac{n}{2}}}{\prod\limits_{i=1}^{p} (s_i^2)^{\frac{n}{2}}},$$

$$\rho = 1 - \frac{2p + 11}{6n}, \nu = \frac{1}{2}p(p-1).$$

Гипотеза о сферичности распределения

Пусть требуется по выборке X из n значений вектора $\vec{\xi}$ проверить гипотезу H_0: $\Sigma = \sigma^2 E$, где величина σ^2 не задана. Гипотеза подобного рода называется гипотезой о сферичности распределения.

Для данной гипотезы отношение правдоподобия:

$$\lambda = \frac{|S|^{\frac{n-1}{2}}}{(tr(S/p))^{\frac{p(n-1)}{2}}}.$$

Если положить (Т. Андерсон. Введение в многомерный статистический анализ):

$$\rho = 1 - \frac{2p^2 + p + 2}{6p(n-1)},$$

то статистика $\eta = -2\rho \ln \lambda$ будет асимптотически иметь распределение χ^2 с $\nu = \frac{1}{2}p(p+1) - 1$ степенями свободы.

Гипотеза о равенстве матрицы ковариаций заданной матрице.

Пусть $X = \{(X_1^{(1)}, X_2^{(1)}, \ldots, X_p^{(1)}), (X_1^{(2)}, X_2^{(2)}, \ldots, X_p^{(2)}), \ldots, (X_1^{(n)}, X_2^{(n)}, \ldots, X_p^{(n)})\}$- выборка объема n из p-мерной нормальной совокупности $N(\vec{\mu}, \Sigma)$ случайной величины $\vec{\xi}$. Проверяется гипотеза $H_0 : \Sigma = \Sigma_0$, против альтернативы $H_1 : \Sigma \neq \Sigma_0$.

Отношение правдоподобия для данной гипотезы:

$$\lambda = \left|\bar{A} \cdot \Sigma_0^{-1}\right|^{\frac{1}{2}n} \exp\left[\frac{1}{2}pn - \frac{1}{2}nSp(\bar{A} \cdot \Sigma_0^{-1})\right],$$

где $\bar{A} = n\widehat{\Sigma}$.

Положим: $\eta = -2 \ln \lambda$, тогда при истинности H_0, статистика η будет асимптотически иметь распределение χ^2 с $\nu = \frac{1}{2}p(p+1)$ степенями свободы.

Гипотеза о равенстве вектора средних заданному вектору и матрицы ковариаций заданной матрице.

Пусть $X = \{(X_1^{(1)}, X_2^{(1)}, \ldots, X_p^{(1)}), (X_1^{(2)}, X_2^{(2)}, \ldots, X_p^{(2)}), \ldots, (X_1^{(n)}, X_2^{(n)}, \ldots, X_p^{(n)})\}$- выборка объема n из p-мерной нормальной совокупности $N(\vec{\mu}, \Sigma)$ случайного вектора $\vec{\xi}$. Проверяется гипотеза

$$H_0 : \vec{\mu} = \vec{\mu}_0, \quad \Sigma = \Sigma_0.$$

Отношение правдоподобия для H_0:

$$\lambda = \left|\bar{A} \cdot \Sigma_0^{-1}\right|^{\frac{1}{2}n} \exp\left[\frac{1}{2}n(p - Sp(\bar{A} \cdot \Sigma_0^{-1}) - (\bar{X} - \vec{\mu}_0)^T \Sigma_0^{-1}(\bar{X} - \vec{\mu}_0))\right].$$

Положим: $\eta = -2 \ln \lambda$, тогда при истинности H_0, статистика η будет асимптотически иметь распределение χ^2 с $\nu = \frac{1}{2}p(p+1) + p$ степенями свободы.

3 Метод главных компонент.

Метод главных компонент осуществляет переход к новой совокупности некоррелированных признаков $\eta_1, \ldots, \eta_m$, каждый из которых является линейной комбинацией исходных признаков ξ_1, ξ_2, $\ldots$, ξ_p. При этом линейные комбинации выбираются таким образом, чтобы среди всех возможных линейных нормированных комбинаций исходных признаков первая главная компонента η_1 обладала наибольшей дисперсией. Геометрически это выглядит как ориентация новой координатной оси η_1 вдоль направления наибольшей вытянутости эллипсоида рассеивания исследуемой выборки в пространстве признаков ξ_1, ξ_2, $\ldots$, ξ_p. Вторая главная компонента имеет наибольшую дисперсию среди всех оставшихся линейных преобразований, некоррелированных с первой главной компонентой. Она интерпретируется как направление наибольшей вытянутости эллипсоида рассеивания, перпендикулярное первой главной компоненте. Следующие главные компоненты определяются по аналогичной схеме.

В дальнейшем из полученных величин можно оставить только $m \quad (m < p)$ наиболее значимых факторов, вносящих максимальный вклад в суммарную дисперсию и использовать эти величины как некие интегральные факторы, характеризующие всю совокупность признаков.

Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_p)$ – центрированная многомерная случайная величина с матрицей ковариаций Σ, то есть $E\xi = (0, \ldots, 0)$ и $E(\xi_i, \xi_j) = \sigma_{i\,j}$. Положим

$$\eta_i = \beta^{(i)T}\xi = \beta_1^{(i)}\xi_1 + \beta_2^{(i)}\xi_2 + \ldots + \beta_p^{(i)}\xi_p, \tag{16}$$

где $\beta^{(i)}$ – векторы неизвестных коэффициентов преобразования, $i = \overline{1, m}$, $(m \leq p)$, причем параметр m заранее не определен, а определяется в процессе получения главных компонент. Будем называть величины η_i главными компонентами (факторами в терминах факторного анализа). В матричной форме приведение вектора исходных признаков ξ к главным компонентам можно записать как:

$$\eta = \beta^T\xi,$$

где

$$\beta = \begin{pmatrix} \beta_1^{(1)} & \beta_1^{(2)} & & \beta_1^{(p)} \\ \beta_2^{(1)} & \beta_2^{(2)} & & \beta_2^{(p)} \\ \ldots & \ldots & \ldots & \ldots \\ \beta_p^{(1)} & \beta_p^{(2)} & & \beta_p^{(p)}. \end{pmatrix}$$

Так как величины ξ_i центрированы, то $M(\eta_i) = 0,\ i = \overline{1, m}$, а дисперсия главных компонент определяется как:

$$D\left(\eta_i\right) = E\left(\eta_i\eta_i^T\right) = \beta^{(i)T}E\left(\xi\xi^T\right)\beta^{(i)} = \beta^{(i)T}\Sigma\beta^{(i)} \tag{17}$$