Дисперсионный анализ

Н.П. Алексеева,

СПбГУ, математико-механический факультет

2020 г.

Линейная модель с фиксированными эффектами

Предполагаем, что у нас есть r нормально распределенных популяций $\mathcal{N}(\mu_1, \sigma), \ldots, \mathcal{N}(\mu_r, \sigma)$, где через μ_i обозначены генеральные средние внутри каждой популяции.
Переменная x_{ij} означает j-е наблюдение в i-й подпопуляции.
Это предположение можно записать в виде:

$$x_{ij} = \mu_i + e_{ij}, \quad i = 1, \ldots, r, \ j = 1, \ldots, n_i, \tag{1}$$

где "ошибки" $e_{ij} \sim \mathcal{N}(0, \sigma)$. Перепишем модель однофакторного дисперсионного анализа в виде

$$x_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \ldots, r, \ j = 1, \ldots, n_i, \tag{2}$$

где μ генеральное среднее объединенной популяции, α_i главные эффекты подпопуляций.

Модель с фиксированными эффектами

Наилучшими оценками параметров модели

$$x_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \ldots, r, \ j = 1, \ldots, n_i,$$

в смысле метода наименьших квадратов являются

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij}, \quad \hat{\alpha}_i = \bar{x}_{i\cdot} - \bar{x} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} - \bar{x}.$$

Ошибкой в этой модели является выражение

$$Q_2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu} - \hat{\alpha}_i)^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i\cdot})^2.$$

Ее число степеней свободы равно $\nu_2 = n - r$. Принятие гипотезы $H_0 : \alpha_1 = \ldots = \alpha_r = 0$ означает справедливость эквивалентной гипотезы $H_0 : \mu_1 = \ldots = \mu_r = \mu$.

Проверка нулевой гипотезы $H_0 : \alpha_1 = \ldots = \alpha_r = 0$

При $H_0 : \alpha_1 = \ldots = \alpha_r = 0$ имеет место усеченная модель вида

$$x_{ij} = \mu + e_{ij}, \ i = 1, \ldots, r; \ j = 1, \ldots, n_i. \tag{3}$$

Наилучшей оценкой параметра μ в смысле метода наименьших квадратов является $\hat{\mu} = \bar{x}$. Ошибкой $\widetilde{Q_2}$ с числом степеней свободы $\widetilde{\nu_2} = n - 1$ является выражение

$$\widetilde{Q_2} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \hat{\mu})^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = Q.$$

Статистика критерия правдоподобия имеет вид:

$$F = \frac{(\widetilde{Q_2} - Q_2)/(\widetilde{\nu_2} - \nu_2)}{Q_2/\nu_2} = \frac{Q_1/(r-1)}{Q_2/(n-r)}. \tag{4}$$

Значимость определяется при помощи доверительного уровня вероятности

$$\alpha_* = P\{F > F_*\},$$

где F_* наблюдаемое значение статистики F. При $\alpha_* < \alpha = 0.05$ гипотеза $H_0 : \alpha_1 = \ldots = \alpha_r = 0$ отвергается.

Матричная модель дисперсионного анализа

Пусть $Y = (y_{11}, \ldots, x_{1n_1}, y_{21}, \ldots, x_{2n_2}, \ldots, y_{r1}, \ldots, x_{rn_r})'$ –вектор наблюдений, $\sum\limits_{i=1}^{r} n_i = n$, матрица плана размерности n на r имеет вид:

$$X = \left[\begin{array}{ccccc} 1 & 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & 0 & \ldots & 0 \\ \hline 1 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 0 & 1 & \ldots & 0 \\ \hline \ldots & \ldots & \ldots & \ldots & \ldots \\ \hline 1 & -1 & -1 & \ldots & -1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & -1 & -1 & \ldots & -1 \end{array}\right], \quad \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \ldots \\ \alpha_{r-1} \end{bmatrix}, \quad Y = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{r1} \\ \ldots \\ y_{rn_r} \end{bmatrix}$$

Матричный вид модели однофакторного дисперсионного анализа

$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, \ldots, r, \; j = 1, \ldots, n_i, \quad$ где $\sum\limits_{i=1}^{r} \alpha_i = 0$

$$Y = X\beta + \epsilon, \quad \text{где} \quad \epsilon \sim \mathbb{N}(\mathbb{O}, \sigma^2\mathbb{I})$$

Оценки параметров по методу наименьших квадратов имеют вид:

$$\hat{\beta} = (\mathrm{X}'\mathrm{X})^{-}\mathrm{X}'\mathrm{Y}\,.$$

Для проверки гипотезы $\mathrm{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_\mathrm{r} = 0$ используют модель с ограничением на параметры при помощи матрицы H размерности r (число параметров) на $\mathrm{s} = \mathrm{r} - 1$

$$\mathrm{H}'\beta = \theta_0, \quad \text{где} \quad \mathrm{H}' = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} \quad \text{и} \quad \theta_0 = \begin{bmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{bmatrix}$$

Нужно показать, что $\mathrm{Z} = \mathrm{H}'\hat{\beta} \sim \mathcal{N}_\mathrm{s}(\mathrm{H}'\beta, \sigma^2\mathrm{D})$, где $\mathrm{D} = \mathrm{H}'(\mathrm{X}'\mathrm{X})^{-}\mathrm{H}$ и $\mathrm{R}_0^2 = (\mathrm{Y} - \mathrm{X}\hat{\beta})'(\mathrm{Y} - \mathrm{X}\hat{\beta}) \sim \sigma^2\chi^2(\mathrm{n} - \mathrm{r})$ распределены независимо. Кроме того $(\mathrm{Z} - \theta_0)'\mathrm{D}^{-1}(\mathrm{Z} - \theta_0) = \mathrm{R}_1^2 - \mathrm{R}_0^2 \sim \sigma^2\chi^2(\mathrm{s})$, где $\mathrm{R}_1^2 = (\mathrm{Y} - \mathrm{X}\beta^*)'(\mathrm{Y} - \mathrm{X}\beta^*)$, β^* оценка параметров усеченной модели. Тогда при справедливости нулевой гипотезы

$$\mathrm{F} = \frac{\mathrm{n} - \mathrm{r}}{\mathrm{s}} \frac{\mathrm{R}_1^2 - \mathrm{R}_0^2}{\mathrm{R}_0^2} \sim \mathrm{F}(\mathrm{s}, \mathrm{n} - \mathrm{r})\,.$$

Для получения оценок МНК дифференцируем квадратичную форму.

$$\mathcal{L}_1 = (\mathrm{Y} - \mathrm{X}\beta)^{\mathrm{T}}(\mathrm{Y} - \mathrm{X}\beta) = \left(\mathrm{y}_1 - \sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{x}_{1\mathrm{i}}\beta_{\mathrm{i}}\right)^2 + \ldots + \left(\mathrm{y}_{\mathrm{n}} - \sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{x}_{\mathrm{ni}}\beta_{\mathrm{i}}\right)^2 .$$

Система нормальных уравнений имеет вид

$$\begin{cases} -\frac{\partial \mathcal{L}_1}{\partial \beta_1} = 2\left(\mathrm{x}_{11}\left(\mathrm{y}_1 - \sum\limits_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{x}_{1\mathrm{i}}\beta_{\mathrm{i}}\right) + \ldots + \mathrm{x}_{\mathrm{n}1}\left(\mathrm{y}_{\mathrm{n}} - \sum\limits_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{x}_{\mathrm{ni}}\beta_{\mathrm{i}}\right)\right) = 0, \\ \ldots \\ -\frac{\partial \mathcal{L}_1}{\partial \beta_{\mathrm{m}}} = 2\left(\mathrm{x}_{1\mathrm{m}}\left(\mathrm{y}_1 - \sum\limits_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{x}_{1\mathrm{i}}\beta_{\mathrm{i}}\right) + \ldots + \mathrm{x}_{\mathrm{nm}}\left(\mathrm{y}_{\mathrm{n}} - \sum\limits_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{x}_{\mathrm{ni}}\beta_{\mathrm{i}}\right)\right) = 0, \end{cases}$$

$$\frac{\partial \mathcal{L}_1}{\partial \beta} = 0 \Longleftrightarrow \quad \mathrm{X}^{\mathrm{T}}(\mathrm{Y} - \mathrm{X}\beta) = 0\,.$$

При ОМК с ограничением на параметры дифференцируется выражение

$$\mathcal{L}_2 = \lambda'(1,s)H'(s,m)\beta(m,1) = (\lambda_1,\ldots,\lambda_s)\begin{bmatrix} h_{11} & \ldots & h_{m1} \\ \ldots & \ldots & \ldots \\ h_{1s} & \ldots & h_{ms} \end{bmatrix}\begin{bmatrix} \beta_1 \\ \ldots \\ \beta_m \end{bmatrix} =$$

$$= (\lambda_1,\ldots,\lambda_s)\begin{bmatrix} \sum\limits_{i=1}^{m} h_{i1}\beta_i \\ \ldots \\ \sum\limits_{i=1}^{m} h_{is}\beta_i \end{bmatrix} = \sum_{j=1}^{s}\lambda_j\left(\sum_{i=1}^{m} h_{ij}\beta_i\right).$$

Система нормальных уравнений

$$\begin{cases} \frac{\partial\mathcal{L}_2}{\partial\beta_1} = \sum\limits_{j=1}^{s}\lambda_j h_{1j} = 0, \\ \qquad\ldots \\ \frac{\partial\mathcal{L}_2}{\partial\beta_m} = \sum\limits_{j=1}^{s}\lambda_j h_{mj} = 0, \end{cases} \iff$$

$$\iff \begin{bmatrix} h_{11} & \ldots & h_{1s} \\ \ldots & \ldots & \ldots \\ h_{m1} & \ldots & h_{ms} \end{bmatrix}\begin{bmatrix} \lambda_1 \\ \ldots \\ \lambda_s \end{bmatrix} = H\lambda = 0\,.$$

Определение. A^- - обобщенная обратная матрица, такая, что для любого вектора Y, при котором система $AX = Y$ совместна, $X = A^-Y$ является ее решением.
A^- существует $\iff AA^-A = A \iff A^-AA^- = A^-$.

Задача 1

Пусть A — симметричная матрица n на n ранга r, B — матрица m на n, BY — m линейных функций от $Y = (y_1, \ldots, y_n)^T$, y_i независимые нормально распределенные случайные величины. Для независимости BY и $Y^T AY$ достаточно условия $BA = 0_{m,n}$, где $0_{m,n}$ матрица m на n из нулей.

Спектральное разложение симметричной матрицы A ранга r:

$$A = \lambda_1 P_1 P_1^T + \ldots + \lambda_r P_r P_r^T, \quad \lambda_i \neq 0, \ i = 1, \ldots, r. \tag{5}$$

Пусть $BA = 0$. Умножим (5) слева на B и справа для примера на P_1.

$$BAP_1 = \lambda_1 BP_1 P_1^T P_1 + \ldots + \lambda_r BP_r P_r^T P_1 = 0_{m,1}.$$

Из ортогональности собственных векторов получаем $BP_1 = 0_{m,1}$. Аналогично для остальных $BP_i = 0_{m,1}$, т.е. линейные функции BY не зависят от функций $P_i^T Y$ (если $a^T b = 0$, то для Y с независимыми компонентами $\mathbb{E}(a^T Y)(Y^T b) = a^T I b = a^T b = 0$) и не зависят от $Y^T AY$, так как

$$Y^T AY = \sum \lambda_i (P_i^T Y)^2.$$

Задача 2

а) $X = X(X'X)^{-}X'X$,
б) $I - X(X'X)^{-}X'$ идемпотентна.

$$\text{а)} \quad G = X(I - (X'X)^{-}X'X),$$

$$\Longrightarrow \quad G'G = (I - (X'X)^{-}X'X)'X'X(I - (X'X)^{-}X'X) =$$

$$= (I - (X'X)^{-}X'X)'(X'X - X'X(X'X)^{-}X'X) =$$

$$= (I - (X'X)^{-}X'X)'(X'X - X'X) = 0, \quad \Longrightarrow \quad G = 0.$$

$$\text{б)} \quad (I - X(X'X)^{-}X')(I - X(X'X)^{-}X') =$$

$$= I - X(X'X)^{-}X' - X(X'X)^{-}X' + \underbrace{X(X'X)^{-}X'X}_{X}(X'X)^{-}X' =$$

$$= I - X(X'X)^{-}X'.$$

Задача 3

Пусть $\beta = (\beta_1, \ldots, \beta_m)'$ вектор параметров, X - матрица плана ранга r, $(X'X)^-$ обобщенная обратная, $\hat{\beta} = (X'X)^- X'Y$, $\hat{\beta}$ решение уравнения $(X'X)\beta = X'Y$. Матрица H' размерности s на m ранга s. Показать, что распределены независимо $Z = H'\hat{\beta}$ и

$$R_0^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta}).$$

Введем матрицу C такую, что $H = X'C$. Тогда

$$H'\hat{\beta} = \overbrace{C'X(X'X)^- X'}^{B} Y = BY.$$

$$R_0^2 = Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}' \underbrace{X'X\hat{\beta}}_{X'Y} = Y'Y - Y'X\hat{\beta}, \qquad (6)$$

$$R_0^2 = Y'Y - Y'X(X'X)^- X'Y = Y'(\underbrace{I - X(X'X)^- X'}_{A})Y = Y'AY\,.$$

Таким образом $BA = C'X(X'X)^- X'(I - X(X'X)^- X') =$

$$= C'(X(X'X)^- X' - \underbrace{X(X'X)^- X'X}_{X}(X'X)^- X') = 0\,.$$

Теорема Фишера-Кочрена

Пусть $Y = (y_1, \ldots, y_n)$ с независимыми компонетами, $y_i \sim \mathcal{N}(0,1)$, квадратичные формы $Q_1, \ldots, Q_k$ рангов $n_1, \ldots, n_k$, такие что

$$Y^T Y = Q_1 + \ldots + Q_k\,.$$

Тогда $n = n_1 + \ldots + n_k \implies Q_i \sim \chi^2(n_i)$ и $Q_1, \ldots, Q_k$ независимы.

Необходимость очевидна. Достаточность. Пусть $n = n_1 + \ldots + n_k$ и $Q_i = Y^T A_i Y$, где матрица A_i ранга n_i. Тогда существует матрица B_i размерности n_i на n, при помощи которой квадратичная форма приводится к диагональному виду, то есть $Q_i = (B_i Y)^T \Delta_i (B_i Y)$, где Δ_i диагональная матрица с элементами ± 1. Соберем k матриц B_i в одну матрицу B размерности n на n.

$$[B_1^T(n, n_1)| \ldots |B_k^T(n, n_k)] \cdot \begin{bmatrix} \Delta_1 & \ldots & 0 \\ \vdots & \ldots & \vdots \\ 0 & \ldots & \Delta_k \end{bmatrix} \cdot \begin{bmatrix} B_1(n_1, n) \\ \vdots \\ B_k(n_k, n) \end{bmatrix} = \sum_i B_i^T \Delta_i B_i$$

$$\implies Y^T Y = \sum_i Q_i = \sum_i (B_i Y)^T \Delta_i (B_i Y) =$$

$$= Y^T (\sum_i B_i^T \Delta_i B_i) Y = Y^T B^T \Delta B Y, \Rightarrow B^T \Delta B = I\,.$$

Ранг матрицы B должен быть равен n, так как с одной стороны, он не может быть больше n, а с другой ранг произведения матриц не превышает минимального ранга сомножителей,

$$n = \mathrm{rank}(I) = \mathrm{rank}((B^T \Delta)B) \leq \min(\mathrm{rank}(B^T \Delta), \mathrm{rank}(B)).$$

Тогда $\Delta = (B^T)^{-1} B^{-1}$ положительно определенная матрица, следовательно, $\Delta = I$ и матрица B является ортогональной. Тогда компоненты вектора $X = BY$ независимы и нормальны, кроме того $Y^T Y = X^T X$.

$$\begin{aligned} Q_1 &= x_1^2 + \ldots x_{n_1}^2, \\ Q_2 &= x_{n_1+1}^2 + \ldots x_{n_1+n_1}^2, \\ Q_3 &= x_{n_1+n_2+1}^2 + \ldots x_{n_1+n_2+n_3}^2, \\ &\ldots \end{aligned}$$

Осюда все Q_i независимы и имеют распределение $\chi^2(n_i)$ соответственно.

Таким образом достаточность $n = n_1 + \ldots + n_k$ установлена.

Для того чтобы квадратичная форма Y^TAY имела распределение χ^2, необходимо и достаточно условие идемпотентности матрицы $A = A^2$, причем $df = \text{rank}(A) = \text{Tr}A$.

Достаточность. Пусть $A = A^2$. Применим теорему Фишера-Кочрена к квадратичной форме ранга n

$$Y^TY = Y^TAY + Y^T(I - A)Y$$

и воспользуемся неравенством Сильвестра для $A = A^2$ и $C = I - A$,

$$\text{rank}(A) + \text{rank}(C) - \text{rank}(AC) \leq n$$
$$\text{rank}(A) + \text{rank}(I - A) - \underbrace{\text{rank}(A(I - A))}_{=0} \leq n$$
$$n = \text{rank}(I) = \text{rank}(A + I - A) \leq \text{rank}(A) + \text{rank}(I - A)\,,$$

Таким образом,

$$n = \text{rank}(A) + \text{rank}(I - A),$$

и по теореме Фишера-Кочрена $Y^TAY \sim \chi^2(\text{rank}(A))$.

Для доказательства необходимости (пусть квадратичная форма Y^TAY имеет распределение χ^2) заметим, что существует ортогональная матрица C такая, что при преобразовании $Y = CX$

$$Y^TAY = X^TC^TACX = \lambda_1 x_1^2 + \ldots + \lambda_m x_m^2,$$
$$Y^TY = X^TX = x_1^2 + \ldots + x_n^2,$$

где $\lambda_1, \ldots, \lambda_m$ ненулевые собственные числа матрицы A, $m = \mathrm{rank}(A)$. Так как $Y^TAY \sim \chi^2(p)$, то характеристическая функция суммы $\lambda_1 x_1^2 + \ldots + \lambda_m x_m^2$ имеет вид

$$[(1 - 2i\lambda_1 t)\ldots(1 - 2i\lambda_m t)]^{-\frac{1}{2}} = (1 - 2it)^{-\frac{p}{2}}.$$

Сравнение х.ф. дает $p = m$, все $\lambda_i = 1$, C^TAC - диагональная матрица с нулями или единицами на главной диагонали.

$$C^TAC = C^TACC^TAC = C^TA^2C, \implies A^2 = A.$$

Число степеней свободы равно $p = m = \mathrm{rank}(A)$, то есть TrA, так как $A^2 = A$.

Задача 4

а) $Z = H'\hat{\beta} \sim \mathcal{N}_s(H'\beta, \sigma^2 D)$, где $D = H'(X'X)^- H$ и
б) $R_0^2 \sim \sigma^2\chi^2(n-r)$, где r – ранг матрицы плана X.

$$E(\hat{\beta} - E\hat{\beta})(\hat{\beta} - E\hat{\beta})' =$$
$$= E((X'X)^- X'(Y - EY))((X'X)^- X'(Y - EY))' =$$
$$= (X'X)^- X'\sigma^2 I((X'X)^- X')' =$$
$$= \sigma^2(X'X)^- X'X(X'X)^- = \sigma^2(X'X)^- .$$ Отсюда $D = H'(X'X)^- H$.

б) $R_0^2 = Y'(I - X(X'X)^- X')Y,$

матрица $I - X(X'X)^- X'$ идемпотентна, $\Longrightarrow$ $R_0^2 \sim \sigma^2\chi^2(n-r)$,

где $n - r = TrI - TrX(X'X)^- X' =$

$$= n - Tr(X'X)^- X'X = n - rang(X'X) = n - rang(X) = n - r .$$

Задача 5

Минимизация выражения $(\mathrm{Y}-\mathrm{X}\beta)'(\mathrm{Y}-\mathrm{X}\beta)$ по β при условии $\mathrm{H}'\beta=\theta_0$ приводит к минимизации $\mathcal{L}=(\mathrm{Y}-\mathrm{X}\beta)'(\mathrm{Y}-\mathrm{X}\beta)+\lambda(\mathrm{H}'\beta-\theta_0)$ и к системе уравнений:

$$\begin{cases} \mathrm{X}'\mathrm{X}\beta+\mathrm{H}\lambda=\mathrm{X}'\mathrm{Y}, \\ \mathrm{H}'\beta=\theta_0 \end{cases} \tag{7}$$

λ^* и β^* – ее решение. $\mathrm{Z}=\mathrm{H}'\hat{\beta}\ \sim\ \mathcal{N}_{\mathrm{s}}(\mathrm{H}'\beta,\sigma^2\mathrm{D})$, где $\mathrm{D}=\mathrm{H}'(\mathrm{X}'\mathrm{X})^{-}\mathrm{H}$, $\mathrm{R}_0^2=(\mathrm{Y}-\mathrm{X}\hat{\beta})^{\mathrm{T}}(\mathrm{Y}-\mathrm{X}\hat{\beta})$.

$$\mathrm{R}_1^2=(\mathrm{Y}-\mathrm{X}\beta^*)'(\mathrm{Y}-\mathrm{X}\beta^*)=(\mathrm{Z}-\theta_0)'\mathrm{D}^{-1}(\mathrm{Z}-\theta_0)+\mathrm{R}_0^2\,.$$

Заметим, что

$$\begin{aligned}\mathrm{R}_1^2&=(\mathrm{Y}-\mathrm{X}\beta^*+\mathrm{X}\hat{\beta}-\mathrm{X}\hat{\beta})'(\mathrm{Y}-\mathrm{X}\beta^*+\mathrm{X}\hat{\beta}-\mathrm{X}\hat{\beta})=\\&=(\mathrm{Y}-\mathrm{X}\hat{\beta})'(\mathrm{Y}-\mathrm{X}\hat{\beta})+(\hat{\beta}-\beta^*)'\mathrm{X}'\mathrm{X}(\hat{\beta}-\beta^*)',\end{aligned}$$

так как $(\hat{\beta}-\beta^*)'\mathrm{X}'(\mathrm{Y}-\mathrm{X}\hat{\beta})=0$ по определению $\hat{\beta}$.

Достаточно показать, что

$$(\hat{\beta}-\beta^*)'\mathrm{X}'\mathrm{X}(\hat{\beta}-\beta^*)'=(\mathrm{Z}-\theta_0)'\mathrm{D}^{-1}(\mathrm{Z}-\theta_0).$$

$(\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*)' = (Z - \theta_0)'D^{-1}(Z - \theta_0)$

Пусть $H = X'XC$, тогда

$$D = H'(X'X)^- H = C'X'X(X'X)^- X'XC = C'X'XC.$$

Вычитаем из $X'X\hat{\beta} = X'Y$ уравнение $X'X\beta^* + H\lambda^* = X'Y$.

$$X'X(\hat{\beta} - \beta^*) = H\lambda^* = X'XC\lambda^*, \quad \hat{\beta} - \beta^* = (X'X)^- X'XC\lambda^*,$$

$$(\hat{\beta} - \beta^*)'X'X(\hat{\beta} - \beta^*) = \underbrace{\lambda^{*'}C'X'X(X'X)^-}_{(\hat{\beta}-\beta^*)'} X'XC\lambda^* =$$

$$= \lambda^{*'}C'X'XC\lambda^* = \lambda^{*'}D\lambda^*$$

$$Z - \theta_0 = H'\hat{\beta} - H'\beta^* = H'(\hat{\beta} - \beta^*) = C'X'X(X'X)^- X'XC\lambda^* =$$

$$= C'X'XC\lambda^* = D\lambda^*$$

$$\Longrightarrow \lambda^* = D^{-1}(Z - \theta_0),$$

$$\Longrightarrow \lambda^{*'}D\lambda^* =$$

$$= (Z - \theta_0)'D^{-1}DD^{-1}(Z - \theta_0) = (Z - \theta_0)'D^{-1}(Z - \theta_0).$$