Конспектировали: Ирина Долгалева и Дарья Краснова

1. Свойства многомерного нормального распределения

Случайный вектор u имеет многомерное нормальное стандартное распределение

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \sim \mathcal{N}(0, I)$$

все элементы $u_i \sim \mathcal{N}(0, 1)$ и независимы. Такой вектор удовлетворяет ряду аксиом.

Аксиомы:

1. В любом ортонормальном базисе закон распределения вектора u одинаковый
2. u_i — независимы
3. $Var(u_i) = 1$

Из этих аксиом следует два важных текстовых свойства нормального распределения.

Свойства:

1. Квадрат длины проекции вектора u на $V(\dim(V) = d)$ имеет $\sim \chi^2_d$ распределение
2. Квадраты длин проекции на V и W независимы, если $V \perp W$.

2. t — распределение

Классическое определение.

Случайная величина T

$$T = \frac{z}{\sqrt{\frac{\gamma_k}{k}}}$$

где $z \sim \mathcal{N}(0, 1), \gamma_k \sim \chi^2_k$, имеет t-распределение с k степенями свободы, где z, γ_k — независимы.

Геометрическое определение.

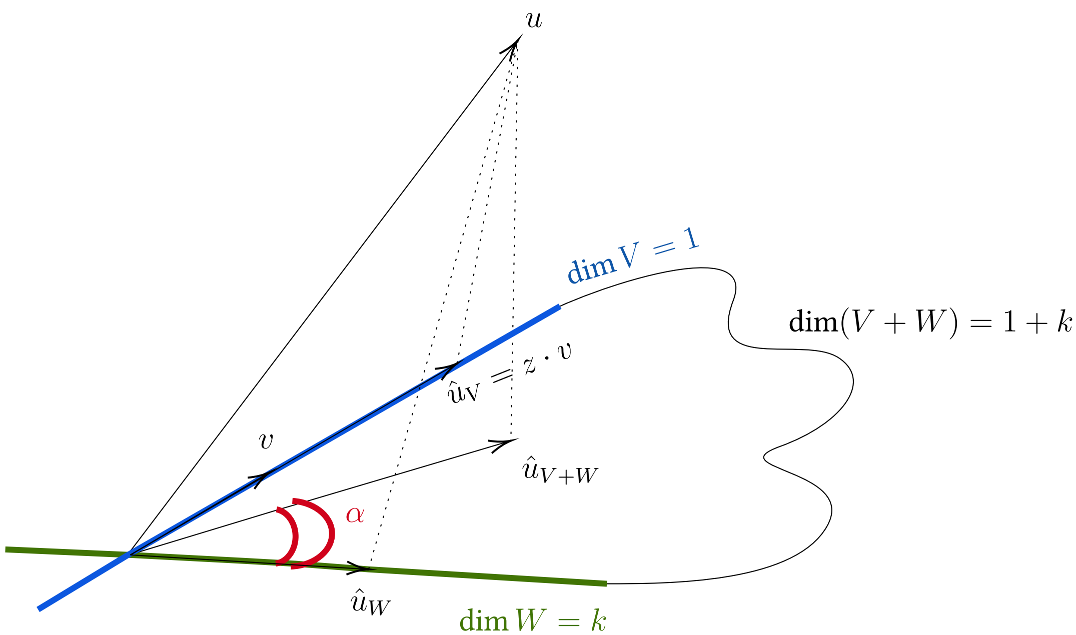
Рассмотрим V — одномерное линейное подпространство, задаваемое вектором единичной длины v и ортогональное ему подпространство W размерности k:

$$V = \underset{\|v\|=1}{\mathrm{Lin}(v)}, \dim(V) = 1$$

$$W = \mathrm{Col}(X), \dim(W) = k$$

$$V \perp W$$

Пусть u — вектор пространства $\mathbb{R}^n$, $\hat{u}_V = z \cdot v$ — проекция u на V, $\hat{u}_W$ — проекция u на W, $\hat{u}_{W+V}$ — проекция u на $W \oplus V$

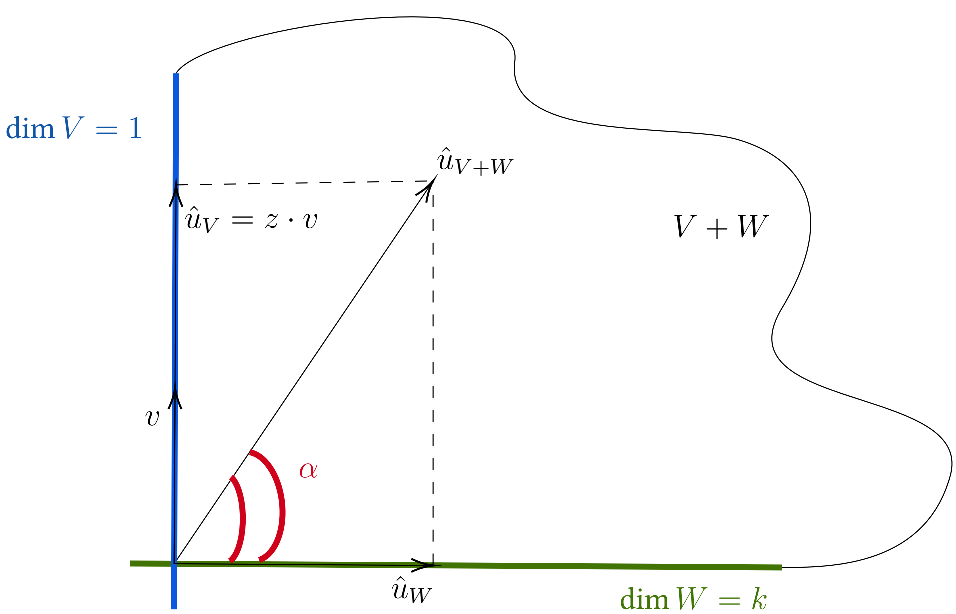


Тогда, случайная величина T

$$T = \frac{z}{\sqrt{\frac{\|\hat{u}_W\|^2}{\dim(W)}}} = \operatorname{tg}\alpha \cdot \sqrt{\dim(W)} \sim t_k$$

имеет t-распределение с k степенями свободы, где α — угол между $\hat{u}_W$ и $\hat{u}_{W+V}$.

Рассмотрим подпространство $V+W$:



v — нормаль (единичной длины) к $W \;\; \Rightarrow \;\; v \perp W$
$\hat{u}_W$ — проекция вектора u на W; $\dim(W) = k \;\; \Rightarrow \;\; \|\hat{u}_W\|^2 \sim \chi^2_k$
Получаем эквивалентные определения.

3. $\mathcal{F}$ — распределение

Классическое определение.

Случайная величина F

$$F = \frac{\gamma_a/a}{\gamma_b/b} \sim \mathcal{F}_{a,b},$$

имеет $\mathcal{F}$ — распределение с (a, b) степенями свободы, где $\gamma_a \sim \chi^2_a, \gamma_b \sim \chi^2_b$ и γ_a, γ_b независимы.

Геометрическое определение.

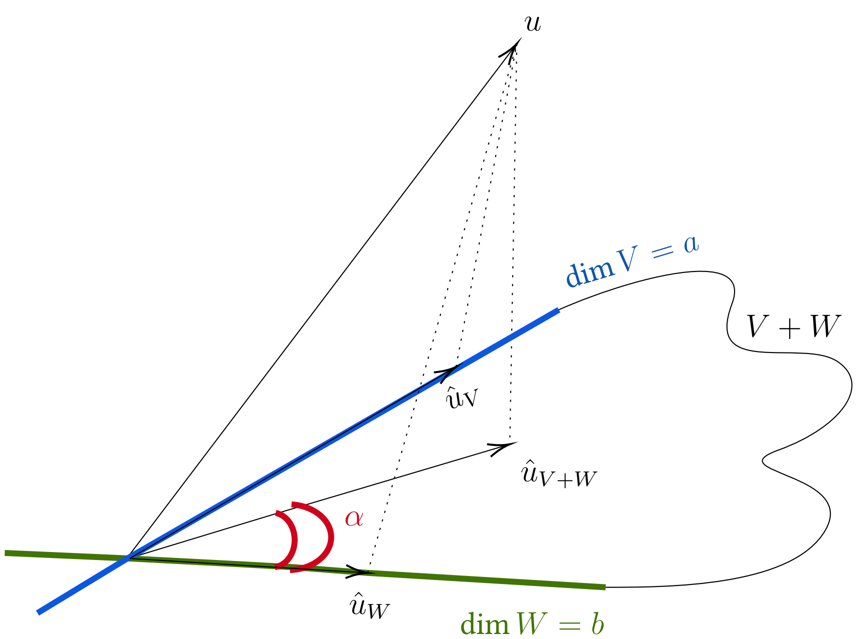
Рассмотрим V — линейное подпространство размерности a и ортогональное ему подпространство W размерности b:

$$\dim(V) = a, \ \dim(W) = b$$

$$V \perp W$$

$$\dim(V) + \dim(W) \leq n \text{ все в } \mathbb{R}^n$$

Пусть u — вектор пространства $\mathbb{R}^n$, $u \sim \mathcal{N}(0, I)$. Тогда $\hat{u}_V$ — проекция u на V, $\hat{u}_W$ — проекция вектора u на W $\hat{u}_{W+V}$ — проекция u на на $W \oplus V$



Тогда, случайная величина F

$$F = \frac{\|\hat{u}_V\|^2/\dim(V)}{\|\hat{u}_W\|^2/\dim(W)} = \mathrm{tg}^2\alpha \cdot \frac{\dim(W)}{\dim(V)} \sim \mathcal{F}\left(\dim(V), \dim(W)\right) = \mathcal{F}_{a,b}$$

имеет $\mathcal{F}$ — распределение с $(\dim(V), \dim(W))$ степенями свободы, где α — угол между $\hat{u}_W$ и $\hat{u}_{W+V}$

$\hat{u}_V$ — проекция u на V; $\dim(V) = a \ \Rightarrow \ \|\hat{u}_V\|^2 \sim \chi^2_a$
$\hat{u}_W$ — проекция u на W; $\dim(W) = b \ \Rightarrow \ \|\hat{u}_W\|^2 \sim \chi^2_b$
Получаем эквивалентные определения.

Заметим, что t-статистика — это частный случай $\mathcal{F}$-статистики:

$$T^2 = F, \dim(V) = a = 1$$

4. Связь $\mathcal{F}$ распределения и эконометрических моделей

Теорема. Рассмотрим задачу регрессии и предположим:

1. Предполагаем $y = X\beta + u$.
2. Оцениваем:

   а) длинную модель $\hat{y}^L = X\hat{\beta}^L$ (верная модель)

   $$\min_{\hat{\beta}^L} \sum_{i=1}^{n} \left(y_i - \hat{y}_i^L\right)^2$$

   б) короткую модель $\hat{y}^S = X\hat{\beta}^S$ (считаем, что последние d коэффициента — нулевые)

   $$\min_{\substack{\hat{\beta}^S \\ \hat{\beta}_k = \hat{\beta}_{k-1} = \ldots = \hat{\beta}_{k-d+1} = 0}} \sum_{i=1}^{n} \left(y_i - \hat{y}_i^S\right)^2$$

3. Стандартные предпосылки

   а) X — полного ранга, неслучайная

   б) $u \sim \mathcal{N}(0, \sigma^2 I) \quad \equiv \quad u_i \sim \mathcal{N}(0, \sigma^2), u_i$ — независимы

4. Проверяем гипотезу

H_0: Верна короткая модель
H_A: Короткая модель не верна

То,

1. $\dfrac{RSS_L}{\sigma^2} \sim \chi^2_{n-k_L}$

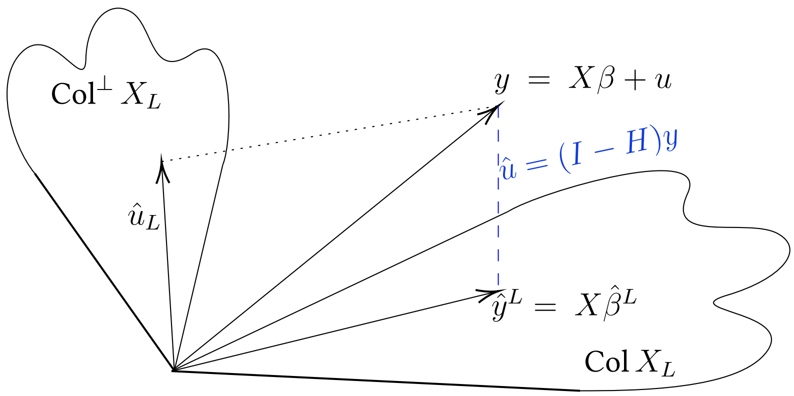
   Если верна H_0:
2. $\dfrac{RSS_S - RSS_L}{\sigma^2} \sim \chi^2_{k_L - k_S}$
3. RSS_L и $(RSS_S - RSS_L)$ — независимы
4. $\dfrac{(RSS_S - RSS_L)/(k_L - k_S)}{RSS_L/(n-k_L)} \sim \mathcal{F}_{k_L - k_S, n-k_L}$

Доказательство.

1. $\hat{u} = (I - H)y = (I - H)(X\beta + u) = [(I - H)X\beta = 0] = (I - H)u \Rightarrow$

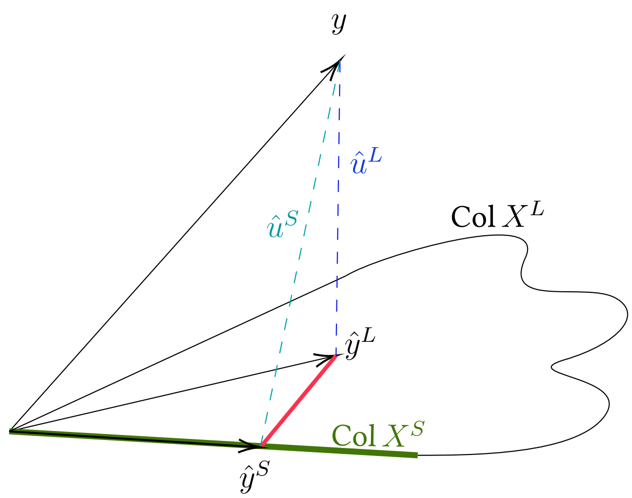
   $\hat{u}_L$ — проекция вектора u на $\mathrm{Col}^{\perp}(X_L)$, $\dim(\mathrm{Col}^{\perp}(X_L)) = n - k_L \Rightarrow$

   $RSS_L = \|\hat{u}_L\|^2$ — квадрат длины проекции u на $(n - k_L)$-мерное подпространство



2. Теперь дополнительно предположим, что еще верна короткая модель (опускаем регрессоры, отвечающие за нулевые компоненты)

$$X = \begin{bmatrix} 1 & x_1 & z_1 & q_1 & w_1 \\ 1 & x_2 & z_2 & q_2 & w_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & z_n & q_n & w_n \end{bmatrix} \Rightarrow X_S = \begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & z_n \end{bmatrix}$$



$$RSS_S - RSS_L = \|\hat{u}_S\|^2 - \|\hat{u}_L\|^2 = \|\hat{y}_S - \hat{y}_L\|^2 = \sum_{i=1}^{n} \left(\hat{y}_i^S - \hat{y}_i^L\right)^2$$

$(\hat{y}_i^S - \hat{y}_i^L)$ — проекция y на $\mathrm{Col}\, X_L \cap \mathrm{Col}^{\perp} X_S$ (ортогональное дополнение X_S в подпространстве X_L)

$\dim(\mathrm{Col}\, X_L \cap \mathrm{Col}^{\perp} X_S) = \dim(\mathrm{Col}\, X_L) - \dim(\mathrm{Col}^{\perp} X_S) = k_L - k_S \Rightarrow$

$$\frac{RSS_S - RSS_L}{\sigma^2} \sim \chi^2_{k_L - k_S}$$

3. RSS_L и $(RSS_S - RSS_L)$ — независимы, поскольку они ортогональны

4. $\dfrac{(RSS_S - RSS_L)/\left(\sigma^2\left(k_L - k_S\right)\right)}{RSS_L/\left(\sigma^2(n - k_L)\right)} = \dfrac{(RSS_S - RSS_L)/(k_L - k_S)}{RSS_L/(n - k_L)} \sim \mathcal{F}_{k_L - k_S, n - k_L}$

□

5. Матрица-мать всех регрессий

Пусть дана матрица X:

$$X = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ x_1 & x_2 & \vdots & x_k \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix};$$

Она состоит из центрированных векторов-столбцов x_j таких, что $\overline{x}_j = 0$ и $\sum_{i=1}^n x_{ij} = 0$.
Тогда матрицы W и M определяются как:

$$W = X^T X; M = W^{-1}$$

Что же находится в матрице M?
Построим регрессию x_1 на $x_1, x_2, \ldots, x_k$. Тогда:

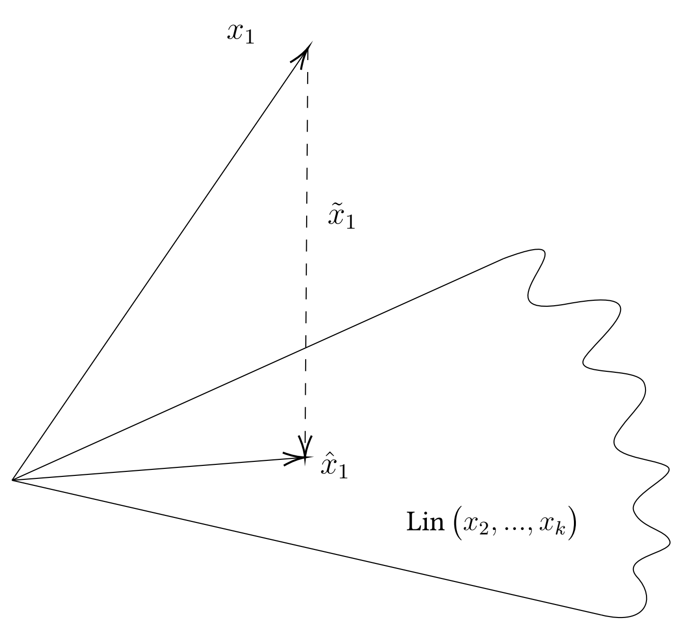
$$\hat{x}_1 = \hat{\alpha}_2 x_2 + \hat{\alpha}_3 x_3 + \ldots + \hat{\alpha}_k x_k$$

$$\tilde{x}_1 = x_1 - (\hat{\alpha}_2 x_2 + \hat{\alpha}_3 x_3 + \ldots + \hat{\alpha}_k x_k)$$

Упражнение: доказать, что:

$$\begin{cases} \langle \tilde{x}_1, x_2 \rangle = 0 \\ \langle \tilde{x}_1, x_3 \rangle = 0 \\ \quad \vdots \\ \langle \tilde{x}_1, x_k \rangle = 0 \end{cases}$$

Доказательство.
Доказательство следует напрямую из картинки:



Вектор $\tilde{x}_1$ ортогонален линейной оболочке Lin $(x_2, ..., x_k)$, а значит ортогонален всем векторам, лежащим в ней. Получаем требуемое.

□

Пользуясь этим знанием, посчитаем, чему равно $\langle \tilde{x}_1, x_1 \rangle$:

$$\langle \tilde{x}_1, x_1 \rangle = \langle \tilde{x}_1, \tilde{x}_1 + (\hat{\alpha}_2 x_2 + \hat{\alpha}_3 x_3 + \ldots + \hat{\alpha}_k x_k) \rangle = \langle \tilde{x}_1, \tilde{x}_1 \rangle = \|\tilde{x}_1\|^2$$

Введем неожиданную нормировку:

$$\check{x}_1 = \frac{\tilde{x}_1}{\|\tilde{x}_1\|^2},$$

где $\tilde{x}_1$ — остаток от регрессии x_1 на $x_2, x_3, \ldots, x_k$, $\|\tilde{x}_1\|^2$ — квадрат длины вектора остатков $\tilde{x}_1$.

Почему была выбрана именно такая нормировка? Потому что хотели подобрать такую нормировку, двукратное применение которой к вектору v давало бы сам вектор v:

$$v \xrightarrow{\text{g}} \frac{v}{\|v\|^2}; g\left(g\left(v\right)\right) = v$$

Например,

$$\begin{pmatrix} 0 \\ 5 \end{pmatrix} \xrightarrow{\text{g}} \begin{pmatrix} 0 \\ \frac{1}{5} \end{pmatrix} \xrightarrow{\text{g}} \begin{pmatrix} 0 \\ 5 \end{pmatrix}$$

Покажем, что нормировка действительно удовлетворяет приведенному выше свойству для любого вектора v:

$$g\left(g\left(v\right)\right) = \frac{v/\|v\|^2}{\left\|v/\|v\|^2\right\|^2} = \frac{v/\|v\|^2}{1/\|v\|^2 \cdot \|v\|^2} = v$$

Аналогично, запишем формулу неожиданной нормировки для вектора остатков $\tilde{x}_2$:

$$\check{x}_2 = \frac{\tilde{x}_2}{\|\tilde{x}_2\|^2},$$

где $\tilde{x}_2$ — остаток от регрессии x_2 на $x_1, x_3, \ldots, x_k$, $\|\tilde{x}_2\|^2$ — квадрат длины вектора остатков $\tilde{x}_2$.

Из векторов $\check{x}_1, \check{x}_2, \ldots, \check{x}_k$ можно составить матрицу $\check{x}$:

$$\check{x} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ \check{x}_1 & \check{x}_2 & \ldots & \check{x}_k \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Таким образом, матрица $\check{x}$ — это матрица странным образом отнормированных остатков в k регрессиях.

Как выглядит ковариационная матрица этих остатков? Для этого поймем, чему равна матрица $\check{x}^T \check{x}$.

Пусть c_{11} — $(1,1)$ элемент матрицы $\check{x}^T \check{x}$, а c_{12} — $(1,2)$ элемент. Тогда:

$$c_{11} = \langle \check{x}_1, \check{x}_1 \rangle = \|\check{x}_1\|^2 = \left\| \frac{\tilde{x}_1}{\|\tilde{x}_1\|^2} \right\|^2 = \frac{1}{\|\tilde{x}_1\|^2} = \frac{1}{RSS_1}$$

$$c_{12} = \langle \check{x}_1, \check{x}_2 \rangle = \frac{1}{\|\tilde{x}_1\|^2} \cdot \frac{1}{\|\tilde{x}_2\|^2} \langle \tilde{x}_1, \tilde{x}_2 \rangle$$

Для того, чтобы найти $\langle \tilde{x}_1, \tilde{x}_2 \rangle$, выпишем, чему они равны в явном виде:

$$\tilde{x}_1 = x_1 - (\hat{\alpha}_2 x_2 + \hat{\alpha}_3 x_3 + \ldots + \hat{\alpha}_k x_k)$$

$$\tilde{x}_2 = x_2 - (\hat{\beta}_1 x_1 + \hat{\beta}_3 x_3 + \ldots + \hat{\beta}_k x_k)$$

Тогда $\langle \tilde{x}_1, \tilde{x}_2 \rangle$ можно представить двумя способами (заменяя $\tilde{x}_1$ или $\tilde{x}_2$):

1. $\langle \tilde{x}_1, \tilde{x}_2 \rangle = \langle x_1 - (\hat{\alpha}_2 x_2 + \hat{\alpha}_3 x_3 + \ldots + \hat{\alpha}_k x_k), \tilde{x}_2 \rangle = -\hat{\alpha}_2 \langle \tilde{x}_2 \rangle = -\hat{\alpha}_2 \|\tilde{x}_2\|^2$
2. $\langle \tilde{x}_1, \tilde{x}_2 \rangle = \langle \tilde{x}_1, x_2 - (\hat{\beta}_1 x_1 + \hat{\beta}_3 x_3 + \ldots + \hat{\beta}_k x_k) \rangle = -\hat{\beta}_1 \langle \tilde{x}_1, \tilde{x}_1 \rangle = -\hat{\beta}_1 \|\tilde{x}_1\|^2$

При этом мы пользовались тем, что:

$$\tilde{x}_1 \perp x_2, x_3, \ldots, x_k$$

$$\tilde{x}_2 \perp x_1, x_3, \ldots, x_k$$

Вернемся к нахождению c_{12}:

$$c_{12} = \langle \check{x}_1, \check{x}_2 \rangle = \frac{1}{\|\tilde{x}_1\|^2} \cdot \frac{1}{\|\tilde{x}_2\|^2} \langle \tilde{x}_1, \tilde{x}_2 \rangle = \frac{-\hat{\beta}_1}{\|\tilde{x}_2\|^2} = \frac{-\hat{\alpha}_2}{\|\tilde{x}_1\|^2}$$

Таким образом, бесплатно получили следующее:

$$\frac{\text{(коэффициент при } x_2 \text{ в регрессии } x_1 \text{ на ост.)}}{RSS_1} = \frac{\text{(коэффициент при } x_1 \text{ в регрессии } x_2 \text{ на ост.)}}{RSS_2}$$

Упражнение: доказать, что $\check{x}^T X = I$.

Доказательство.

Для этого найдем $(1,1)$ и $(1,2)$ элементы матрицы $\check{x}^T X = I$:

$$b_{11} = \langle \check{x}_1, x_1 \rangle = \left\langle \frac{\tilde{x}_1}{\|\tilde{x}_1\|^2}, x_1 \right\rangle = \langle \tilde{x}_1, x_1 \rangle \cdot \frac{1}{\|\tilde{x}_1\|^2} = \langle \tilde{x}_1, \tilde{x}_1 \rangle \cdot \frac{1}{\|\tilde{x}_1\|^2} = 1$$

$$b_{12} = \langle \check{x}_1, x_2 \rangle = \langle \tilde{x}_1, x_2 \rangle \cdot \frac{1}{\|\tilde{x}_1\|^2} = 0$$

□

Упражнение: пусть

$$W = X^T X$$

$$M = \check{x}^T \check{x}$$

Доказать, что $M \cdot W = I$ (то есть $M = W^{-1}$).

Доказательство.

Из предыдущей части лекции известно, чему равны матрицы M и W:

$$M = \begin{bmatrix} \dfrac{1}{\langle \tilde{x}_1, \tilde{x}_1 \rangle} & \dfrac{-\hat{\alpha}_2}{\langle \tilde{x}_1, \tilde{x}_1 \rangle} & \dfrac{-\hat{\alpha}_3}{\langle \tilde{x}_1, \tilde{x}_1 \rangle} & \dots \\ \dfrac{-\hat{\beta}_1}{\langle \tilde{x}_2, \tilde{x}_2 \rangle} & \dfrac{1}{\langle \tilde{x}_2, \tilde{x}_2 \rangle} & \dfrac{-\hat{\beta}_3}{\langle \tilde{x}_2, \tilde{x}_2 \rangle} & \vdots \\ \dfrac{-\hat{\gamma}_1}{\langle \tilde{x}_3, \tilde{x}_3 \rangle} & \dfrac{-\hat{\gamma}_2}{\langle \tilde{x}_3, \tilde{x}_3 \rangle} & \ddots & \vdots \\ \vdots & \vdots & \dots & \dfrac{1}{\langle \tilde{x}_k, \tilde{x}_k \rangle} \end{bmatrix}$$

$$W = \begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_2, x_1 \rangle & \dots \\ \langle x_1, x_2 \rangle & \dots & \dots \\ \vdots & \ddots & \vdots \\ \langle x_1, x_k \rangle & \dots & \dots \end{bmatrix}$$

Найдем элтемент a_{11} и a_{12} — $(1,1)$ и $(1,2)$ элементы матрицы $M \cdot W$:

$$a_{11} = \frac{\langle x_1, x_1 \rangle}{\langle \tilde{x}_1, \tilde{x}_1 \rangle} - \frac{\hat{\alpha}_2 \langle x_1, x_2 \rangle}{\langle \tilde{x}_1, \tilde{x}_1 \rangle} - \dots - \frac{\hat{\alpha}_k \langle x_1, x_k \rangle}{\langle \tilde{x}_1, \tilde{x}_1 \rangle} = \frac{1}{\langle \tilde{x}_1, \tilde{x}_1 \rangle} \Big(\langle x_1, x_1 \rangle - \hat{\alpha}_2 \langle x_1, x_2 \rangle - \dots - \hat{\alpha}_k \langle x_1, x_k \rangle \Big)$$

Воспользуемся двумя равенствами:

1. $x_1 - (\hat{\alpha}_2 x_2 + \hat{\alpha}_3 x_3 + \dots + \hat{\alpha}_k x_k) = \tilde{x}_1$
2. $\langle \tilde{x}_i, x_i \rangle = \langle \tilde{x}_i, \tilde{x}_i \rangle$

Поэтому:

$$a_{11} = \frac{\langle x_1, \tilde{x}_1 \rangle}{\langle \tilde{x}_1, \tilde{x}_1 \rangle} = 1$$

Находим a_{12}:

$$a_{12} = \frac{1}{\langle \tilde{x}_1, \tilde{x}_1 \rangle} \Big(\langle x_2, x_1 \rangle - \hat{\alpha}_2 \langle x_2, x_2 \rangle - \dots - \hat{\alpha}_k \langle x_2, x_k \rangle \Big) = \frac{\langle x_2, \tilde{x}_1 \rangle}{\langle \tilde{x}_1, \tilde{x}_1 \rangle} = 0,$$

так как $\langle x_2, \tilde{x}_1 \rangle = 0$, потому что остаток от регрессии x_1 на $x_2, x_3, \dots, x_k$ ортогонатен $x_2, x_3, \dots, x_k$. Аналогично можно найти все элементы этой матрицы. Таким образом, получили, что $M \cdot W = I$.

$\square$

Итог: если $W \cdot (n-1)$ — выборочная ковариационная матрица X, то $W^{-1} = M$:

$$M = W^{-1} = \begin{bmatrix} \dfrac{1}{RSS_1} & \dfrac{-\hat{\alpha}_2}{RSS_1} & \dots & \dfrac{-\hat{\alpha}_k}{RSS_1} \\ \dfrac{-\hat{\beta}_1}{RSS_2} & \dfrac{1}{RSS_2} & \dfrac{-\hat{\beta}_3}{RSS_2} & \vdots \\ \dfrac{-\hat{\gamma}_1}{RSS_3} & \dfrac{-\hat{\gamma}_2}{RSS_3} & \ddots & \vdots \\ \vdots & \vdots & \dots & \dfrac{1}{RSS_k} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{RSS_1} & \dfrac{-\hat{\beta}_1}{RSS_2} & \dfrac{-\hat{\gamma}_1}{RSS_3} & \dots \\ \dfrac{-\hat{\alpha}_2}{RSS_1} & \dfrac{1}{RSS_2} & \dfrac{-\hat{\gamma}_2}{RSS_3} & \vdots \\ \dfrac{-\hat{\alpha}_3}{RSS_3} & \dfrac{-\hat{\beta}_3}{RSS_2} & \ddots & \vdots \\ \vdots & \vdots & \dots & \dfrac{1}{RSS_k} \end{bmatrix}$$

P.S. Напутствие последующим конспектирующим:

1. Картинки можно легко и быстро рисовать в Matcha.
2. Если нужно переиспользовать / доработать картинки из этой лекции, их можно найти, кликнув мышкой сюда.