Конспектировал: Ефим Лубошников

1. Свойства математического ожидания

Вспомним, какими свойствами обладает математическое ожидание случайной величины и случайного вектора. Обозначим: y_1, y_2 - случайные величины, y, z - случайные вектора, a - скалярная константа, A, B - матрицы констант.

Случайная величина	Случайный вектор
$E(ay_1) = aE(y_1)$	$E(AyB) = AE(y)B$
$Var(y_1) = E(y_1^2) - E(y_1)^2$	$Var(y) = E(yy^T) - E(y)E(y^T)$
$Var(ay_1) = a^2Var(ay_1)$	$Var(Ay) = A(E(yy^T) - E(y)E(y^T))A^T$
$Var(a + y_1) = Var(y_1)$	$Var(a + y) = Var(y)$
$Cov(y_1, y_2) = E(y_1y_2) - E(y_1)E(y_2)$	$Cov(y, z) = E(yz^T) - E(y)E(z^T)$
	$Cov(Ay, Bz) = ACov(y, z)B^T$
	$Cov(y, z) = Cov(z, y)^T$
	Если $y, z \in \mathbb{R}^n$, то справедливо:
	$Cov(y + z, w) = Cov(y, w) + Cov(z, w)$
	$Var(y + z) = Var(y) + Var(z) + Cov(y, z) + Cov(z, y)$,

Так выглядит дисперсия случайного вектора $y \in \mathbb{R}^n$:

$$Var(y) = \begin{bmatrix} Var(y_1) & Cov(y_1, y_2) & \vdots & Cov(y_1, y_n) \\ Cov(y_2, y_1) & Var(y_2) & \vdots & Cov(y_2, y_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(y_n, y_1) & Cov(y_n, y_2) & \vdots & Var(y_n) \end{bmatrix}$$

2. Предпосылки для теоремы Гаусса-Маркова

Рассмотрим задачу регрессии $y = X\beta + u$ в случае двух регрессоров: $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + u_i$

Предпосылки:

1. Предполагаем β -неизвестная константа

2. X должен иметь полный ранг, чтобы оценки МНК существовали. X может быть:

   a) известная константа, $y_t = \beta_1 + \beta_2 t + u_t$

б) наблюдение - случайная величина, времянной ряд.

3. Гомоскедастичность $Var(u_i|X) = \sigma^2 I$

4. Отсутствие автокорреляции $Cov(u_i, u_j|x) = 0, i \neq j$

5. $E(u|X) = 0$

Обозначим:

$\hat{y}$ - предсказанный y на тренировочной выборке.

$\hat{y}_{test}$ - предсказанный y на тестовой выборке.

2.1. Упражнение мега-матрица

Задание: Заполнить матрицу Var, Cov для соответсвующих элементов матрицы:

	y	$\hat{y}$	u	$\hat{u}$	$\hat{\beta}$
y					
$\hat{y}$					
u					
$\hat{u}$					
$\hat{\beta}$					
y_{test}					
$\hat{y}_{test}$					
$\hat{u}_{test}$					

Решение:

Будем пользоваться результатами, полученными ранее:

$\hat{\beta} = (X^TX)^{-1}X^Ty$

$\hat{y} = X\hat{\beta}$

$\hat{u} = y - \hat{y}$

$y_{test} = X_{test}\beta + u_{test}$

$\hat{y}_{test} = X_{test}\hat{\beta}$

Предварительно найдём мат.ожидания:

$E(y|X) = E(X\beta + u|X) = X\beta + E(u|X) = X\beta$

$E(\beta|X) = E((X^TX)^{-1}X^Ty|X) = E((X^TX)^{-1}X^TX\beta|X) = (X^TX)^{-1}X^TX\beta = \beta$

Получим элементы мега-матрицы:

$Var(y|X) = Var(X\beta + u|X) = Var(u|X) = \sigma^2 I$

$Var(\beta|X) = Var((X^TX)^{-1}X^Ty|X) = (X^TX)^{-1}X^TVar(y|X)((X^TX)^{-1}X^T)^T = (X^TX)^{-1}X^T\sigma^2X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$

$Cov(\hat{\beta}, \hat{u}|X) = Cov((X^TX)^{-1}X^Ty, y - \hat{y}|X) = Cov((X^TX)^{-1}X^Ty, y|X) - Cov((X^TX)^{-1}X^Ty, \hat{y}|X) = (X^TX)^{-1}X^TVar(y|X) - Cov(\beta, X\beta|X) = \sigma^2(X^TX)^{-1}X^T - \sigma^2(X^TX)^{-1}X^T = 0$

На контрольной работе - вывести 2 элемента мега-матрицы.

Выведем ещё пару свойств для RSS, ESS, TSS:

Как можно представить RSS?

$RSS = \sum_{i=1}^{n} \hat{u_i}^2 = \hat{u}^T\hat{u} = ((I-H)y)^T((I-H)y) = y^T(I-H)^T(I-H)y = y^T(I-H)y$

Далее, найдем математическое ожидание RSS при фиксированном X:

$E(RSS|X) = E(y^T(I-H)y|X) =$ |всегда можно взять $trace$, если результат скаляр| $= E(tr(y^T(I-H)y|X)) = E(tr(I-H)yy^T|X)) = tr((I-H)E(yy^T)) = tr(I-H)(Var(y|X) + E(y|X)E(y^T|X)) = tr(I-H)(\sigma^2 I + X\beta\beta^T X^T)) = tr[(I-H)(\sigma^2 I + X\beta\beta^T X^T))] = tr(I-H)\sigma^2 + tr((I-H)(X\beta\beta^T X^T)) = tr(I-H)\sigma^2 = (n-k)\sigma^2$

Слагаемое $(I-H)X = 0$, так как $(I-H)$ - ортогональное дополнение к X.

$E(RSS|X) = (n-k)\sigma^2$

$\Rightarrow \hat{\sigma}^2 = \frac{RSS}{n-k}$ - оценка разумная.

2.2. Теорема Гаусса - Маркова

Теорема 2.1. Если выполняются все предпосылки про β, u, X, существуют несмещенные оценки $\hat{\beta} = (X^TX)^{-1}X^Ty$ и $\hat{\beta}_{alt} = A_{alt}^T y$, то:

$Var(\hat{\beta}_{alt_j}) \geq Var(\hat{\beta}_j|X)$

$Var(\hat{\beta}_{alt}) - Var(\hat{\beta}|X)$ являетеся положительно определенной.

Замечание. Первая формулировка является частным и упрощенным вариантом второй формулировки.

Доказательство. $\hat{\beta} = ((X^TX)^{-1}X^T)^Ty = A^Ty$

$\hat{\beta}_1$ - 1 строка из A^T. Также $\hat{\beta}_1$ - <первый столбец в A, y> A задаёт веса, с которыми надо брать y, чтобы получить $\hat{\beta}$.

Рассмотрим, каким образом собрана матрица A:

$$\begin{pmatrix} C_1 & C_2 & \dots & C_k \\ C_1 & C_2 & \dots & C_k \\ \vdots & \vdots & \vdots & \vdots \\ C_1 & C_2 & \dots & C_k \end{pmatrix}$$

$$\begin{pmatrix} * & \dots \\ * & \dots \\ \vdots & \dots \\ * & \dots \end{pmatrix}$$

Первая строка матрицы A формируется, как линейная комбинация C_i:

$a_1 = *C_1 + *C_2 + ... + *C_k$

Веса, с которыми надо барть y, чтобы получить β являются линейной комбинацией строк в X.

Дисперсия $Var(\hat{\beta}_1|X) = Var(a_1^T y|X) = a_1^T \sigma^2 I a_1 = \sigma^2 a_1^T a_1 = \sigma^2 * \left\| a_1 \right\|^2$

$Var(\hat{\beta}_1{}^{alt}|X) = \sigma^2 \left\| a_1 \right\|^2 \geq Var(\hat{\beta}_1|X)$

Краткая схема доказательства:

1)$a_1 \in lin(...)$

Это следует из $A = X(X^T X)^{-1}$

2)$a_1^{alt} - a_1 \perp lin(...)$

Это следует из $(a_1^{alt})^T X\beta = \beta_1 = a_1^T X\beta$

Рассмотрим доказательство для общего случая:

$\delta = Var(\hat{\beta}^{alt}|X) - Var(\hat{\beta}|X)$

$\delta-$ положительно определённая.

$\omega^T \omega \geq 0$

$Var(\omega^T \hat{\beta}_{alt}|X) \geq Var(\omega^T \hat{\beta}|X)$

$Var(a_{alt}^T Y|X) \geq Var(a^T y|X)$

1)$a \in lin(X)$;

2)$a^T - a \perp lin(X)$

□

Теоремы без доказательства

Если верны предпосылки о β, X, u, и известно, что $u \sim N(0; \sigma^2 I)$, то:

1) $\hat{\beta}_j|X \sim N(\beta_j, Var(\hat{\beta}_j|X))$

2) $\frac{\hat{\beta}_j - \beta_j}{Var(\hat{\beta}_j|X)} = \frac{\hat{\beta}_j - \beta_j}{\kappa(\hat{\beta}_j)} \sim t_{n-k}$

3) Если $\exists$ 2 вложенные модели: "short" - частный случай "long", то:

$$\frac{\frac{RSS_s - RSS_l}{k_l - k_s}}{\frac{RSS_l}{n - k_l}} \sim F_{k_l - k_s, n - k_l}$$

Задача

Дано: 200 наблюдений.

Модель A: $\hat{y}_i = 2.7 - 3.2x_i, \quad R^2 = 0.6$

(1.1) (0.6)

Модель B: $\hat{y}_i = 3.2 - 2.8x_i + 4.2z_i + 0.1x_i^2, \quad R^2 = 0.7$

$(1.2) \quad (0.7) \quad (1.1) \quad (0.5)$

В скобках ниже указаны стандартные ошибки

Задание: 1. Проверить гипотезу при уровне значимости $\alpha = 5\%$

H_0: верна модель А.

H_A: модель А - неверна, но верна В.

2.Постройте 95% доверительный интервал для β_x.

Решение

Из теоремы 2 :

$$\frac{\hat{\beta}_x - \beta_x}{\kappa(\hat{\beta}_x)} \sim t_{200-2}$$

Т-распределение с 198 степенями свободы очень похоже на нормальное.

$t_{198} = N(0;1)$

$-t_{cr} \leq \frac{\hat{\beta}_x - \beta_x}{\kappa(\hat{\beta}_x)} \leq t_{cr}$

Дробь с вероятностью 95% лежит от -1.96 до 1.96. $\beta_x \in [\hat{\beta}_x - t_{cr}\kappa(\hat{\beta}_x); \hat{\beta}_x + t_{cr}\kappa(\hat{\beta}_x)]$

$\beta_x \in [-3.2 - 1.96 * 0.6; -3.2 + 1.96 * 0.6] \approx [-4.4; 2.0]$

Предположим, β_x=0, попадает в $[-4.4; 2.0]$.

$H_0 : \beta_x = 0$

$H_A : \beta \neq 0, \alpha = 5\%$

H_0 отвергается, β_x значительно отличается от нуля.

$$R^2 = 1 - \frac{RSS}{TSS}$$

Воспользуемся 3 теоремой:

$$\frac{\frac{RSS_s - RSS_l}{k_l - k_s}}{\frac{RSS_l}{n - k_l}} \sim F_{k_l - k_s, n - k_l}$$

$$\frac{\frac{RSS_s - RSS_l}{4-2}}{\frac{RSS_l}{200-196}} = \frac{\frac{RSS_s - RSS_l}{TSS}/2}{\frac{RSS_l}{TSS}/196} = \frac{(0.4 - 0.3) * 196}{2 * 0.3} \approx 3.04$$

$F_{cr} = 3.04$

Сравним F_{cr} с $F_{2,196}$, видим, что гипотеза H_0 отвергается, поэтому выбираем модель В для оценки.