Прикладная статистика в машинном обучении

Квиз #3B

16 ноября 2020 г.

Критические значения.

- $t_{500-6;0.025} = -1.964778$
- $Z_{0.025} = -1.959964$
- $t_{500-6,0.05} = -1.647944$
- $Z_{0.05} = -1.644854$
- $F_{2,1;0.95} = 199.5$
- $F_{1,1;0.95} = 161.4476$.

Условие для задач 1-4.

Рассмотрим модель $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \beta_3 X_{2i} + \beta_4 \ln X_{3i} + \beta_5 X_{4i} + u_i$, которая оценивается по 500 наблюдениям при помощи МНК. Оценённая модель имеет следующий вид:

$$\hat{y}_i = \underset{(0.95)}{8} - \underset{(13)}{142} X_{1i} + \underset{(5)}{11.2} X_{1i}^2 + \underset{(2.7)}{4} X_{2i} + \underset{(0.001)}{0.17} \ln X_{3i} - \underset{(0.09)}{2} X_{4i}.$$

В скобках указаны стандартные ошибки оценок коэффициентов. Будем считать, что все предпосылки теоремы Гаусса-Маркова выполнены, и $u \sim \mathcal{N}(0, \sigma^2 I)$. Также известно, что

$$\hat{\operatorname{Var}}(\hat{\beta}) = \begin{pmatrix} 0.9025 & -0.04 & 0.001 & -0.001 & 0.47 & -0.1 \\ -0.04 & 169 & 0.09 & 1 & 0.9 & 0.4 \\ 0.001 & 0.09 & 25 & 0.017 & -0.48 & 9 \\ -0.001 & 1 & 0.017 & 7.29 & -0.002 & 4.78 \\ 0.47 & 0.9 & -0.48 & -0.002 & 10^{-5} & 1.17 \\ -0.1 & 0.4 & 9 & -4.78 & 1.17 & 0.0081 \end{pmatrix}$$

Задача 1.

a) Проверьте коэффициенты β_0, β_1 и β_2 на значимость на 5% уровне.

b) Проверьте коэффициенты β_0, β_3 и β_4 на значимость на 5% уровне.

c) Проверьте коэффициенты β_1, β_2 и β_5 на значимость на 5% уровне.

d) Проверьте коэффициенты β_2, β_3 и β_5 на значимость на 5% уровне.

e) Проверьте коэффициенты β_0, β_4 и β_5 на значимость на 5% уровне.

f) Проверьте коэффициенты β_0, β_1 и β_3 на значимость на 5% уровне.

g) Проверьте коэффициенты β_3, β_4 и β_5 на значимость на 5% уровне.

Задача 2.

a) Проверьте следующие гипотезы на уровне значимости 5%.

$$\begin{cases} H_0 : \beta_1 = -150, \\ H_1 : \beta_1 \neq -150 \end{cases} \text{и} \begin{cases} H_0 : \beta_2 = 12, \\ H_1 : \beta_2 > 12. \end{cases}$$

b) Проверьте следующие гипотезы на уровне значимости 5%.

$$\begin{cases} H_0 : \beta_3 = 2, \\ H_1 : \beta_3 \neq 2 \end{cases} \text{и} \begin{cases} H_0 : \beta_5 = -1, \\ H_1 : \beta_5 < -1. \end{cases}$$

c) Проверьте следующие гипотезы на уровне значимости 5%.

$$\begin{cases} H_0 : \beta_4 = 0.1, \\ H_1 : \beta_4 \neq 0.1 \end{cases} \text{и} \begin{cases} H_0 : \beta_1 = -140, \\ H_1 : \beta_1 < -140. \end{cases}$$

d) Проверьте следующие гипотезы на уровне значимости 5%.

$$\begin{cases} H_0 : \beta_4 = 0.1, \\ H_1 : \beta_4 \neq 0.1 \end{cases} \text{и} \begin{cases} H_0 : \beta_2 = 11, \\ H_1 : \beta_2 > 11. \end{cases}$$

Задача 3.

a) Постройте 95%-ый доверительный интервал для β_1.

b) Постройте 95%-ый доверительный интервал для β_2.

c) Постройте 95%-ый доверительный интервал для β_3.

d) Постройте 95%-ый доверительный интервал для β_4.

e) Постройте 95%-ый доверительный интервал для β_5.

Задача 4.

a) Постройте 95%-ый доверительный интервал для $\mathbb{E}(y_{501}|X_1 = X_2 = 0, X_3 = X_4 = 1)$.

b) Постройте 95%-ый доверительный интервал для $\mathbb{E}(y_{501}|X_1 = X_2 = X_3 = X_4 = 1)$.

c) Постройте 95%-ый доверительный интервал для $\mathbb{E}(y_{501}|X_1 = X_2 = 9, X_3 = 1, X_4 = 0)$.

d) Постройте 95%-ый доверительный интервал для $\mathbb{E}(y_{501}|X_1 = X_2 = 4, X_3 = 1, X_4 = 0)$.

Задача 5.

Рассмотрим модель $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$, оцениваемую при помощи МНК. Предположим, что все предпосылки теоремы Гаусса-Маркова выполнены, и $u \sim \mathcal{N}(0, \sigma^2 I)$. Модель оценили на следующих данных:

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 5 & 0 \\ 1 & 2 & 1 \\ 1 & 3 & 1 \end{bmatrix}, y = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}.$$

Оказалось, что $\hat{\beta} = \begin{pmatrix} -1.76 & 0.72 & 1.08 \end{pmatrix}'$.

a) Проверьте регрессию на значимость в целом на уровне значимости 5%.

b) Проверьте гипотезу

$$\begin{cases} H_0 : \beta_1 = \beta_2, \\ H_1 : \beta_1 \neq \beta_2 \end{cases}$$

на уровне значимости 5%, если

$$(X'X)^{-1}_R = \begin{pmatrix} 6.8 & -1.5 \\ -1.5 & 0.4 \end{pmatrix}, \hat{\beta}_R = \begin{pmatrix} -1.7 \\ 0.8 \end{pmatrix}$$

Задача 6.

Рассмотрим модель $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$, оцениваемую при помощи МНК. Предположим, что все предпосылки теоремы Гаусса-Маркова выполнены, и $u \sim \mathcal{N}(0, \sigma^2 I)$. Модель оценили на следующих данных:

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 4 \\ 1 & 1 & 0 \\ 1 & 0 & 2 \end{bmatrix}, y = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}.$$

Оказалось, что $\hat{\beta} = \begin{pmatrix} 1.4 & -0.8 & 0.3 \end{pmatrix}'$.

a) Проверьте регрессию на значимость в целом на уровне значимости 5%.

b) Проверьте гипотезу

$$\begin{cases} H_0 : \beta_1 = \beta_2, \\ H_1 : \beta_1 \neq \beta_2 \end{cases}$$

на уровне значимости 5%, если

$$(X'X)^{-1}_R = \begin{pmatrix} 1.15 & -0.3 \\ -0.3 & 0.1 \end{pmatrix}, \hat{\beta}_R = \begin{pmatrix} 1.2 \\ 0.1 \end{pmatrix}$$

Задача 7.

Рассмотрим линейную модель $y = X\beta + u$, оцениваемую при помощи МНК. Пусть $\mathbb{E}(u) = 0$, $\mathrm{Var}(u) = \sigma^2 I$, число наблюдений равно n, число регрессоров, включая константный, равно k. Найдите следующие величины и укажите их размеры

a) $\mathbb{E}(y)$, $\mathrm{Var}(\hat{\beta})$, $\mathrm{Cov}(u, \hat{u})$, $\mathrm{Cov}(\hat{\beta}, \hat{u})$.

b) $\mathbb{E}(\hat{\beta})$, $\mathrm{Var}(\hat{y})$, $\mathrm{Cov}(\hat{y}, \hat{u})$, $\mathrm{Cov}(y, \hat{u})$.

c) $\mathbb{E}(\hat{y})$, $\mathrm{Var}(\hat{u})$, $\mathrm{Cov}(\hat{u}, \hat{\beta})$, $\mathrm{Cov}(u, \hat{y})$.

Задача 8.

Рассмотрим модель парной регрессии

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

оцениваемую при помощи МНК. Пусть известно, что

y_i	x_i
1	1
2	9
4	7
5	2

$$(X'X)^{-1} = \begin{bmatrix} 0.75 & -0.1 \\ -0.1 & 0.02 \end{bmatrix}$$

a) Постройте 95%-ый доверительный интервал для $\hat{\beta}_0$, используя стандартные ошибки HC_0.

b) Постройте 95%-ый доверительный интервал для $\hat{\beta}_0$, используя стандартные ошибки HC_3.

c) Постройте 95%-ый доверительный интервал для $\hat{\beta}_1$, используя стандартные ошибки HC_0.

d) Постройте 95%-ый доверительный интервал для $\hat{\beta}_1$, используя стандартные ошибки HC_3.