

16. Проверка гипотезы $H_0 : b = b_0$.

Доверительные интервалы для коэффициентов регрессии. Коэффициент детерминации.

Основные гипотезы, лежащие в основе линейной регрессионной модели (возможно из можно не писать в билете):

Основные гипотезы:

1. $Y_t = a + bX_t + \varepsilon_t, t = 1, \dots, n$, - спецификация модели.
2. X_t - детерминированная величина; вектор $(X_1, \dots, X_n)'$ не коллинеарен вектору $i = (1, \dots, 1)'$.
3. (а) $E\varepsilon_t = 0, E(\varepsilon_t)^2 = V(\varepsilon_t) = \sigma^2$ - не зависит от t ($E\varepsilon = 0, V(\varepsilon) = \sigma^2 I_n$)
(б) $E(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$, некоррелированность ошибок для разных наблюдений.
Часто добавляется условие:
(с) Ошибка $\varepsilon_t, t = 1, \dots, n$, имеют совместное нормальное распределение:
 $\varepsilon_t \sim N(0, \sigma^2)$

Свойства МНК-оценок

МНК-оценки коэффициентов регрессии $\hat{a}, \hat{b}$ имеют совместное нормальное распределение:

$$\hat{a} \sim N\left(a, \sigma^2 \frac{\sum X_t^2}{n \sum x_t^2}\right), \quad \hat{b} \sim N\left(b, \sigma^2 \frac{1}{\sum x_t^2}\right), \text{ где } x_t = X_t - \bar{X}, \bar{X} = \frac{\sum X_t}{n}$$

Отсюда получаем, что $\hat{b} - b \sim N(0, \sigma_b^2)$, где $\sigma_b^2 = \frac{\sigma^2}{\sum x_t^2}$

$$\widehat{Var}(\hat{b}) = s_b^2 = s^2 \frac{1}{\sum (X_t - \bar{X})^2} = \frac{s^2}{\sum x_t^2} \Rightarrow \frac{\hat{b} - b}{s_b} \sim N(0, 1)$$

Учитывая, что распределение оценки дисперсии ошибок $s^2 = \widehat{\sigma^2}$ имеет распределение:

$$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2), \text{ получим}$$

$$\frac{s}{\sigma} \sim \sqrt{\frac{1}{n-2} \chi^2(n-2)} \Rightarrow \{\text{по определению статистики Стьюдента}\} \Rightarrow t = \frac{(\hat{b} - b)/s_b}{s/\sigma} \sim t(n-2), \text{ где } t(n-2) \\ \text{— Т-статистика Стьюдента с } (n-2) \text{ степенями свободы.}$$

$$\text{Так как } \frac{\sigma_b}{\sigma} = \frac{s_b}{s} \text{ получим } t = \frac{\hat{b} - b}{s_b} \sim t(n-2) (*)$$

$$\text{Аналогично показываем, что } t = \frac{\hat{a} - a}{s_a} \sim t(n-2).$$

Статистику (*) можно использовать для проверки гипотезы $H_0 : b = b_0$ против альтернативной гипотезы $H_1 : b \neq b_0$.

Пусть H_0 верна, тогда $t = \frac{\hat{b} - b_0}{s_b} \sim t(n-2)$

Гипотеза H_0 принимается при $|t| \leq t_\alpha(n-2)$.

Другими словами:

»

100% · (1 - α)-ый доверительный интервал для b равен:

$$\left[\hat{b} - t_e s_b, \hat{b} + t_e s_b \right], \text{ где } t_e \text{ — точка t-распределения с } (n-2) \text{ степенями свободы при уровне значимости } \alpha.$$

(Хз надо ли это писать в билете) Если гипотеза нормальности ошибок не выполняется — **3(с)**, то при некоторых условиях регулярности на поведение X_t при росте n оценки $\hat{a}, \hat{b}$ имеют асимптотически нормальное распределение, то есть

$$\hat{a} \rightarrow N\left(a, \sigma^2 \frac{\sum X_t^2}{n \sum x_t^2}\right), \quad \hat{b} \rightarrow N\left(b, \sigma^2 \frac{1}{\sum x_t^2}\right)$$

Коэффициент детерминации R^2

Рассмотрим вариацию значения Y_t вокруг среднего значения $\bar{Y}$: $\sum (Y_t - \bar{Y})^2$.

Разобьем вариацию на две части: объяснению и необъяснённую.

Пусть $\hat{Y}_t = \hat{a} + \hat{b}X_t$ — предсказанные значения Y_t .

Тогда $Y_t - \bar{Y} = (Y_t - \hat{Y}_t) + (\hat{Y}_t - \bar{Y})$ и вариация Y_t записывается в виде:

$$\sum (Y_t - \bar{Y})^2 = \sum (Y_t - \hat{Y}_t)^2 + \sum (\hat{Y}_t - \bar{Y})^2 + 2 \sum (Y_t - \hat{Y}_t)(\hat{Y}_t - \bar{Y}) \sim TSS = ESS + RSS + 2 \sum (Y_t - \hat{Y}_t)(\hat{Y}_t - \bar{Y})$$

TSS – вся дисперсия. ESS – необъясненная часть дисперсии. RSS – объясненная часть дисперсии.

Третье слагаемое равно 0, так как

$$\sum e_t(\hat{Y}_t - \bar{Y}) = \sum e_t(\hat{a} + \hat{b}X_t - \bar{Y}) = \{x_t = X_t - \bar{X}\} = (\hat{a} + \hat{b}\bar{X} - \bar{Y}) \sum e_t + \hat{b} \sum e_t x_t = 0 \text{ (вектор остатков регрессии } e_t \text{ ортогонален константе } i \text{ - единичному вектору и вектору } x)$$

Коэффициентом детерминации называется $R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS}$.

$R^2 \in [0, 1]$, если $R^2 = 0$, то регрессия ничего не дает, $R^2 = 1$ — точная подгонка.