Проверка гипотез, относящихся к коэффициентам регрессии

Рассмотрим классическую нормальную линейную модель множественной регрессии (выполнены 6 условий Гаусса-Маркова)

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \dots + a_{r-1} x_{ir-1} + a_r + \xi_i \qquad i = \overline{1,n}$$

1. $M\xi_i = 0 \quad \forall i = \overline{1,n}$

2. $D\xi_i = \sigma^2 \quad \forall i = \overline{1,n}$

3. $\operatorname{cov}\left(\xi_i, \xi_j\right) = 0 \quad \forall i \neq j$

4. Случайные ошибки не зависят от объясняющих переменных

5. n>r

6. $\xi_i \sim N\left(0, \sigma^2\right)$

Проверка гипотез, относящихся к коэффициентам регрессии

Рассмотрим классическую нормальную линейную модель множественной регрессии (выполнены 6 условий Гаусса-Маркова)

В этой модели можно проверять гипотезы и строить доверительные прогнозы

Проверка гипотез, относящихся к коэффициентам регрессии

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \ldots + a_{r-1} x_{ir-1} + a_r + \xi_i$$

$H_0 : a_i = 0$ -переменная y не зависит от переменной x_i
(i-й коэффициент не значим)

Проверка гипотез, относящихся к коэффициентам регрессии

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \dots + a_{r-1} x_{ir-1} + a_r + \xi_i$$

$H_0 : a_i = 0$ -переменная y не зависит от переменной x_i
(i-й коэффициент не значим)

1) $s_i = \sqrt{\left(X^T X\right)^{-1}_{ii}}\, s$ называется **стандартной ошибкой коэффициента** $\hat{a}_i$

Проверка гипотез, относящихся к коэффициентам регрессии

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \ldots + a_{r-1} x_{ir-1} + a_r + \xi_i$$

$H_0 : a_i = 0$ -переменная у не зависит от переменной x_i
(і-й коэффициент не значим)

1) $s_i = \sqrt{\left(X^T X\right)^{-1}_{ii}}\, s$ называется **стандартной ошибкой коэффициента** $\hat{a}_i$

2) $T_i = \dfrac{\hat{a}_i}{s_i}$ - статистика Стьюдента

Проверка гипотез, относящихся к коэффициентам регрессии

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \ldots + a_{r-1} x_{ir-1} + a_r + \xi_i$$

1) $H_0 : a_i = 0$ -переменная y не зависит от переменной x_i
(i-й коэффициент не значим)

1) $s_i = \sqrt{\left(X^T X\right)^{-1}_{ii}}\, s$ называется **стандартной ошибкой коэффициента** $\hat{a}_i$

2) $T_i = \dfrac{\hat{a}_i}{s_i}$ - статистика Стьюдента

Проверка гипотез, относящихся к коэффициентам регрессии

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \dots + a_{r-1} x_{ir-1} + a_r + \xi_i$$

3) $p_i = P(T_i / H_0)$ - вероятность получить значение T_i в условиях H_0 Эту величину называют Р-значение

Проверка гипотез, относящихся к коэффициентам регрессии

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \ldots + a_{r-1} x_{ir-1} + a_r + \xi_i$$

3) $p_i = P(T_i / H_0)$ - вероятность получить значение T_i в условиях H_0 Эту величину называют Р-значение

4) Задают уровень значимости α

Обычно уровень значимости 0,1; 0,05; 0,01

Если Р-значение меньше α, принимается гипотеза $H_1 : a_i \neq 0$

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	-26,93164811	4,523407834	-5,95384	4,73E-09
N	2,674036105	0,231999296	11,52605	1,28E-27
Nrab	0,59409725	0,137923673	4,307435	1,96E-05

$$s_i = \sqrt{\left(X^T X\right)^{-1}_{ii}}\, s \qquad T_i = \frac{\hat{a}_i}{s_i} \qquad p_i = P(T_i / H_0)$$

$$y = a_1 x_1 + a_2 x_2 + \dots \; a_{r-1} x_{r-1} + a_r + \xi$$

$$H_0 : a_1 = a_2 = \dots \; = a_{r-1} = 0$$

-означает незначимость уравнения регрессии в целом. Переменная у
-не зависит от факторов, включенных в модель

2.

$$y = a_1 x_1 + a_2 x_2 + \ldots\ a_{r-1} x_{r-1} + a_r + \xi$$

$$H_0 : a_1 = a_2 = \ldots\ = a_{r-1} = 0$$

- означает незначимость уравнения регрессии в целом

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - r}{r - 1}$$

$p = P(F / H_0)$ - вероятность получить значение F в условиях H_0. Эту величину называют P-значение

Проверка гипотез, относящихся к коэффициентам регрессии

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \dots + a_{r-1} x_{ir-1} + a_r + \xi_i$$

Задают уровень значимости α

Обычно уровень значимости 0,1; 0,05; 0,01

Если Р-значение меньше α , гипотеза

$H_0 : a_1 = a_2 = \dots = a_{r-1} = 0$ отвергается

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	2	22875,36105	11437,68	66,73107	1,31E-26
Остаток	537	92041,60034	171,3996		
Итого	539	114916,9614			

$H_0 : a_1 = a_2 = 0$ отвергается, уравнение регрессии значимо

Доверительные интервалы для прогнозируемых значений

объясняющие переменные $x_1 \ldots x_{r-1}$ приняли значения $x_1^*, x_2^* \ldots x_{r-1}^*$

Какое значение примет объясняемая переменная *y*?

$$\hat{y}^* = \hat{a}_1 x_1^* + \ldots + \hat{a}_{r-1} x_{r-1}^* + \hat{a}_r$$ – точечный прогноз

$$\hat{y}^* - s_{\hat{y}^*} t_{1-\frac{\alpha}{2}}(n-r) \le y^* \le \hat{y}^* + s_{\hat{y}^*} t_{1-\frac{\alpha}{2}}(n-r)$$

– доверительный интервал для прогнозируемого значения.