Многомерные статистические методы 2020

Домашнее задание по лекции №2. Доказательства свойств регрессионной модели

Золотарев Антон Олегович, БСТ182

9 октября 2020 г.

1 $TSS = RSS + ESS$

Чтобы доказать данное равенство, опишем общую, регрессионную и остаточную сумму квадратов по определению:

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2 + \sum_{i=1}^{n}(\widehat{y_i} - \overline{y})^2$$

Попробуем видоизменить левую часть уравнения:

$$\sum_{i=1}^{n}(y_i-\overline{y})^2 = \sum_{i=1}^{n}(y_i-\overline{y}-\widehat{y_i}+\widehat{y_i})^2 = \sum_{i=1}^{n}[(y_i-\widehat{y_i})^2-2(y_i-\widehat{y_i})(\widehat{y_i}-\overline{y})+(\widehat{y_i}-\overline{y})^2] =$$

$$= \sum_{i=1}^{n}(y_i - \widehat{y_i})^2 + \sum_{i=1}^{n}(\widehat{y_i} - \overline{y})^2 - 2 \cdot \sum_{i=1}^{n}[(y_i - \widehat{y_i})(\widehat{y_i} - \overline{y})]$$

Из действий выше очевидно, что необходимое нам равенство будет выполнено, если $2 \cdot \sum\limits_{i=1}^{n}[(y_i - \widehat{y_i})(\widehat{y_i} - \overline{y})] = 0$. Чтобы показать это, сначала вспомним оценку МНК для $\widehat{y_i}$:

$$\begin{cases} b_0 = \widehat{y_i} - b_1 \cdot \widehat{x_i} \\ b_1 = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{{S_x}^2} \end{cases}$$

Тогда имеем следующее равенство:

$$\widehat{y_i} = \overline{y} - b_1 \cdot \overline{x} + b_1 \cdot x_i = \overline{y} + (x_i - \overline{x}) \cdot \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{{S_x}^2}$$

Подставим правую часть полученного равенства в $2\cdot\sum\limits_{i=1}^{n}\left[(y_i-\widehat{y_i})(\widehat{y_i}-\overline{y})\right]$ и вынесем дробь за знак суммы как константу:

$$C\cdot\sum_{i=1}^{n}[(y_i-\widehat{y_i})(x_i-\overline{x})]=C\cdot\left[\sum_{i=1}^{n}y_i(x_i-\overline{x})-\sum_{i=1}^{n}\widehat{y_i}(x_i-\overline{x})\right]=$$

$$\text{C}\cdot\left[\sum_{i=1}^{n}y_i\cdot\sum_{i=1}^{n}(x_i-\overline{x})-\sum_{i=1}^{n}\widehat{y_i}\cdot\sum_{i=1}^{n}(x_i-\overline{x})\right]=$$

$$=C\cdot\left[\sum_{i=1}^{n}y_i\cdot 0-\sum_{i=1}^{n}\widehat{y_i}\cdot 0\right]=C\cdot 0=0$$

Теперь становится очевидно, что $\sum\limits_{i=1}^{n}(y_i-\overline{y})^2=\sum\limits_{i=1}^{n}(y_i-\widehat{y_i})^2+\sum\limits_{i=1}^{n}(\widehat{y_i}-\overline{y})^2$, откуда $TSS=RSS+ESS$. Что и требовалось доказать.

2 Какие из нижеследующих утверждений верны для МНК уравнения регрессии?

(A) График уравнения регрессии - единственная прямая, для которой сумма остатков регрессии минимальна.

(B) Среднее значение остатков регрессии равно нулю.

(C) Тангенс угла наклона линии регрессии - константа, кратная коэффициенту корреляции.

(D) Наклон линии регрессии свидетельствует о том, насколько изменится результирующая переменная при изменении факторной переменной на одну единицу.

(A) Данное утверждение верно исходя из определения уравнения регрессии для метода наименьших квадратов (см. слайды 12-13 лекции 5-6):

(B) Данное утверждение верно, покажем это:

$$\frac{\sum\limits_{i=1}^{n}(y_i-\widehat{y_i})}{n}=0$$

$$\sum_{i=1}^{n}\left[(y_i-\overline{y}+(x_i-\overline{x})\cdot\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{{S_x}^2}\right]=\sum_{i=1}^{n}(y_i-\overline{y})-c\cdot\sum_{i=1}^{n}(x_i-\overline{x})=0-c\cdot 0=0$$

(C) Данное утверждение также верно:

$$\tan\beta = b_1$$

$$b_1 = const \cdot \rho_{XY}$$

Обратимся к определениям b_1 и ρ_{XY}:

$$\begin{cases} \rho_{XY} = \dfrac{\overline{xy} - \overline{x}\cdot\overline{y}}{S_x \cdot S_y} \\ b_1 = \dfrac{\overline{xy} - \overline{x}\cdot\overline{y}}{{S_x}^2} \end{cases}$$

Отсюда следует, что:

$$b_1 = \rho_{XY} \cdot \frac{S_x}{S_y} = const \cdot \rho_{XY}$$

(D) Данное утверждение также является верным, поскольку по определению наклон линии регрессии определяется коэффиицентом b_1 и верно следующее равенство (см. слайды 16-17 лекции 5-6):

$$b_1 \cdot \triangle X = \triangle Y$$