Методический семинар 15.03.2012, часть1

Тестирование регрессионных остатков на наличие выбросов, нормальность, гомоскедастичность. Зачем это нужно?

Демидова О.А.

Множественная линейная регрессия

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i,\ i = 1,\ldots,n -$$

общий вид модели множественной регрессии.

$X_2,\ldots,X_k$ – факторы (независимые переменные),

Y – зависимая переменная,

u_i – возмущения,

n – число наблюдений.

$$\hat{Y}_i = b_1 + b_2 X_{2i} + \ldots + b_k X_{ki},\ i = 1,\ldots,n -$$

оцененная функция регрессии.

Множественная линейная регрессия

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i\,,\; i = 1,\ldots,n -$$

Обозначим

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix},\; X_2 = \begin{pmatrix} X_{21} \\ X_{22} \\ \ldots \\ X_{2n} \end{pmatrix},\ldots, X_k = \begin{pmatrix} X_{k1} \\ X_{k2} \\ \ldots \\ X_{kn} \end{pmatrix},\; u = \begin{pmatrix} u_1 \\ u_2 \\ \ldots \\ u_n \end{pmatrix}$$

Тогда уравнение регрессии можно переписать в векторном виде

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

Множественная линейная регрессия

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i, \; i = 1, \ldots, n -$$

Если ввести матрицу наблюдений X размера (nxk) и вектор коэффициентов β размера (kx1)

$$X = \begin{pmatrix} 1 & X_{21} & \ldots & X_{k1} \\ 1 & X_{22} & \ldots & X_{k2} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & X_{2n} & \ldots & X_{kn} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \ldots \\ \beta_k \end{pmatrix},$$

то уравнение регрессии можно переписать в матричном виде:

$$Y = X\beta + u$$

Множественная линейная регрессия

Используя матричный вид уравнения регрессии,

$$Y = X\beta + u$$

можно получить компактную формулу для оценок МНК коэффициентов $\beta_1,\ldots,\beta_k$ (без доказательства):

$$b = (X'X)^{-1} X'Y,$$

где ´ - операция транспонирования, $^{-1}$ - операция взятия обратной матрицы,

$$b = \begin{pmatrix} b_1 \\ \ldots \\ b_k \end{pmatrix}.$$

Теорема Гаусса – Маркова для случая множественной линейной регрессии

Если модель множественной линейной регрессии

$$Y = \beta_1 + \beta_2 X_2 + ... + \beta_k X_k + u$$

1) Правильно специфицирована (т.е. выбрана правильная функциональная форма, включены необходимые факторы и нет лишних)

2) Не существует линейной связи между регрессорами

3) Возмущения имеют нулевое мат. ожидание $E(u_i) = 0$,

4) Дисперсии возмущений одинаковы $D(u_j) = \sigma_u^2$, $j = 1,\ldots,n$ (гомоскедастичность)

5) Возмущения с разными номерами не коррелируют

$Cov(u_i, u_j) = 0$

Тогда оценки МНК являются BLUE (Best Linear Unbiased Estimator).

Формулы для вычисления оценок коэффициентов и стандартных отклонений

$$Y = X\beta + u$$

$$b = \begin{pmatrix} b_1 \\ \dots \\ b_k \end{pmatrix}.$$

$$b = (X'X)^{-1} X'Y,$$

Ковариационная матрица:

$$V\hat{a}r(b) = \hat{\sigma}_u^2 (X'X)^{-1}, \quad \hat{\sigma}_u^2 = \frac{RSS}{n-k}$$

$$\hat{D}(b_j) = \hat{\sigma}_u^2 (X'X)^{-1}_{jj}, \; j = 1,\dots, k$$

Классическая линейная регрессия

Предположение о нормальности распределения возмущений

$$u_i \sim N(0, \sigma_u^2), \quad \textit{тогда}$$

$$b_j \sim N(\beta_j, \sigma_u^2 (X'X)^{-1}_{jj}), \; j = 1, \ldots, k$$

И мы можем проверять гипотезы о значимости коэффициента и строить доверительные интервалы.

Множественная линейная регрессия

```
. reg EARNINGS S EXP

      Source |       SS       df       MS              Number of obs =     540
-------------+------------------------------           F(  2,   537) =   67.54
       Model |  22513.6473     2  11256.8237           Prob > F      =  0.0000
    Residual |  89496.5838   537  166.660305           R-squared     =  0.2010
-------------+------------------------------           Adj R-squared =  0.1980
       Total |  112010.231   539  207.811189           Root MSE      =   12.91

------------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   2.678125   .2336497    11.46   0.000     2.219146    3.137105
         EXP |   .5624326   .1285136     4.38   0.000     .3099816    .8148837
       _cons |  -26.48501    4.27251    -6.20   0.000    -34.87789   -18.09213
------------------------------------------------------------------------------
```

$$\widehat{EARNINGS} = -26.49 + 2.68S + 0.56EXP$$

Пример оцененной множественной регрессии.

Проверка выполнения условий теоремы Гаусса-Маркова и нормальности распределения возмущений

- Тестирование правильной спецификации модели (следующая лекция)
- Тестирование наличия выбросов (outliers)
- Тестирование нормальности распределения остатков
- Тестирование гомоскедастичности остатков