Klasyczny Model Regresji Liniowej
Testowanie hipotez

Stanisław Cichocki

Natalia Nehrebecka

Wykład 9

Plan wykładu

- 1. Kombinacja liniowa parametrów
  - Estymator elementu β_k
- 2. Prognozowanie
- 3. Testowanie hipotez
  - Rozkład estymatora **b**

Plan wykładu

- 1. Kombinacja liniowa parametrów
  - Estymator elementu β_k
- 2. Prognozowanie
- 3. Testowanie hipotez
  - Rozkład estymatora **b**

Kombinacja liniowa parametrów

- Czasami interesuje nas nie tylko wielkość oszacowanych parametrów ale także oszacowania pewnych funkcji tych parametrów
- W przypadku MNK łatwo jest przeanalizować własności estymatora kombinacji liniowej parametrów β
- Kombinacja liniowa:

$$\delta' \beta = \sum_{k=1}^{K} \delta_k \beta_k$$

gdzie: δ - nielosowy wektor współczynników

Kombinacja liniowa parametrów

- Nieobciążony estymator tej kombinacji to $\delta' b$
- Wariancja tej kombinacji:

$$Var(\delta' b) = E[(\delta' b - \delta' \beta)(\delta' b - \delta' \beta)']$$

$$= E[\delta'(b-\beta)(b-\beta)'\delta] = \delta' E[(b-\beta)(b-\beta)']\delta$$

$$= \delta' \Sigma \delta$$

- Estymator wariancji tej kombinacji: $\delta' \hat{\Sigma}_b \delta$

Kombinacja liniowa parametrów

- Przykład

dochod - zmienna zależna,
wiek, wiek_2 oraz interakcje miedzy wykształceniem i płcią - zmienne niezależne

```
xi: regress dochod wiek wiek_2  i.plec*i.wyksztalcenie
```

Source	SS	df	MS
Model	81648217.6	7	11664031.1
Residual	665832918	1075	619379.458
Total	747481135	1082	690832.842

Number of obs	=	1083
F(7, 1075)	=	18.83
Prob > F	=	0.0000
R-squared	=	0.1092
Adj R-squared	=	0.1034
Root MSE	=	787.01

dochod	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
wiek	36.38318	15.39846	2.36	0.018	6.168745	66.59762
wiek_2	-.4049352	.1962222	-2.06	0.039	-.7899572	-.0199131
_Iplec_1	**-144.4044**	143.4615	-1.01	0.314	-425.9008	137.0919
_Iwykszta1~2	**274.2703**	105.1538	2.61	0.009	67.94046	480.6002
_Iwykszta1~3	**1040.998**	137.1701	7.59	0.000	771.8461	1310.149
IpleXwyk~2	**-143.4455**	153.4394	-0.93	0.350	-444.5201	157.6292
IpleXwyk~3	**-682.341**	197.7395	-3.45	0.001	-1070.34	-294.3418
_cons	-121.1625	300.6773	-0.40	0.687	-711.1435	468.8184

Kombinacja liniowa parametrów

- Szczególny przypadek tej kombinacji liniowej: kombinacja, która daje k-ty element wektora β
- Wtedy:

$$\delta' = [0....0 \;\; 1 \;\; 0...0]$$

$$\delta' \beta = \beta_k$$

- Nieobciążonym estymatorem β_k jest b_k
- Estymatorem wariancji β_k jest $\delta' \hat{\Sigma}_b \delta = [\hat{\Sigma}_b]_{kk}$

Kombinacja liniowa parametrów

- Przykład

dochod - zmienna zależna,
wiek, wiek_2 oraz interakcje miedzy wykształceniem i płcią - zmienne niezależne

```
xi: regress dochod wiek wiek_2  i.plec*i.wyksztalcenie

      Source |       SS       df       MS              Number of obs =    1083
-------------+------------------------------           F(  7,  1075) =   18.83
       Model |  81648217.6     7  11664031.1           Prob > F      =  0.0000
    Residual |   665832918  1075  619379.458           R-squared     =  0.1092
-------------+------------------------------           Adj R-squared =  0.1034
       Total |   747481135  1082  690832.842           Root MSE      =  787.01

------------------------------------------------------------------------------
      dochod |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |   36.38318   15.39846     2.36   0.018     6.168745    66.59762
      wiek_2 |  -.4049352   .1962222    -2.06   0.039    -.7899572   -.0199131
    _Iplec_1 |  -144.4044   143.4615    -1.01   0.314    -425.9008    137.0919
_Iwyksztal~2 |   274.2703   105.1538     2.61   0.009     67.94046    480.6002
_Iwyksztal~3 |   1040.998   137.1701     7.59   0.000     771.8461    1310.149
_IpleXwyk_~2 |  -143.4455   153.4394    -0.93   0.350    -444.5201    157.6292
_IpleXwyk_~3 |   -682.341   197.7395    -3.45   0.001     -1070.34   -294.3418
       _cons |  -121.1625   300.6773    -0.40   0.687    -711.1435    468.8184
```

Plan wykładu

- 1. Kombinacja liniowa parametrów
  - Estymator elementu β_k
- 2. Prognozowanie
- 3. Testowanie hipotez
  - Rozkład estymatora **b**

Prognoza

- Prognoza: przewidywana przez model wartość y_f dla danego wektora zmiennych objaśniających x_f
- Błąd prognozy:

$$e_f = \hat{y}_f - y_f$$

- Nieobciążona prognoza y_f w KMRL to:

$$\hat{y}_f = x_f b$$

Prognoza

- Prognoza ta jest nieobciążona bo:

$$E(e_f) = E(\hat{y}_f - y_f) = E(x_f b - x_f \beta) = x_f E(b) - x_f \beta$$

$$= x_f \beta - x_f \beta = 0$$

Prognoza

- Przykład

wydg- zmienna zależna
dochg – zmienna niezależna

```
reg wydg dochg

 Source |        SS        df        MS               Number of obs =   31679
-------------+------------------------------               F(  1, 31677) =21732.03
       Model |  2.3577e+10      1  2.3577e+10              Prob > F      =  0.0000
    Residual |  3.4367e+10 31677  1084914.37               R-squared     =  0.4069
-------------+------------------------------               Adj R-squared =  0.4069
       Total |  5.7944e+10 31678  1829163.02               Root MSE      =  1041.6

------------------------------------------------------------------------------
        wydg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dochg |   .5879668   .0039884   147.42   0.000     .5801493    .5957843
       _cons |   712.8104   10.01991    71.14   0.000      693.171    732.4498
------------------------------------------------------------------------------
```

Prognoza

- Pytanie

  Ile wynoszą wydatki jeśli dochód będzie równy 5000 zł?

- Odpowiedź:

$$wydg_f = 712,81 + 0,58 * 5000 = 3612,81$$

Prognoza

- Wariancja prognozy:

$$Var(\hat{y}_f) = Var(x_f b) = x_f Var(b) x'_f = x_f \Sigma_b x'_f$$

Prognoza

- Wariancja błędu prognozy (mierzy precyzję prognozy):

$$Var(e_f) = E[(\hat{y}_f - y_f)(\hat{y}_f - y_f)'] = E[x_f (b-\beta)(b-\beta)' x'_f]$$

$$-2x_f E[(b-\beta)\varepsilon_f] + E[\varepsilon^2{}_f] = x_f \Sigma_b x'_f + \sigma^2 = Var(\hat{y}_f) + \sigma^2$$

- Gdyż: $$E[(b-\beta)\varepsilon_f] = (X'X)^{-1} X' E(\varepsilon\varepsilon_f) = 0$$

Prognoza

- Wariancja błędu prognozy:

$$Var(\hat{e}_f) = Var(\hat{y}_f) + \sigma^2$$

błąd estymacji

błąd losowy

Prognoza

- Estymator wariancji błędu prognozy:

$$x_f \hat{\Sigma}_b x'_f + s^2$$

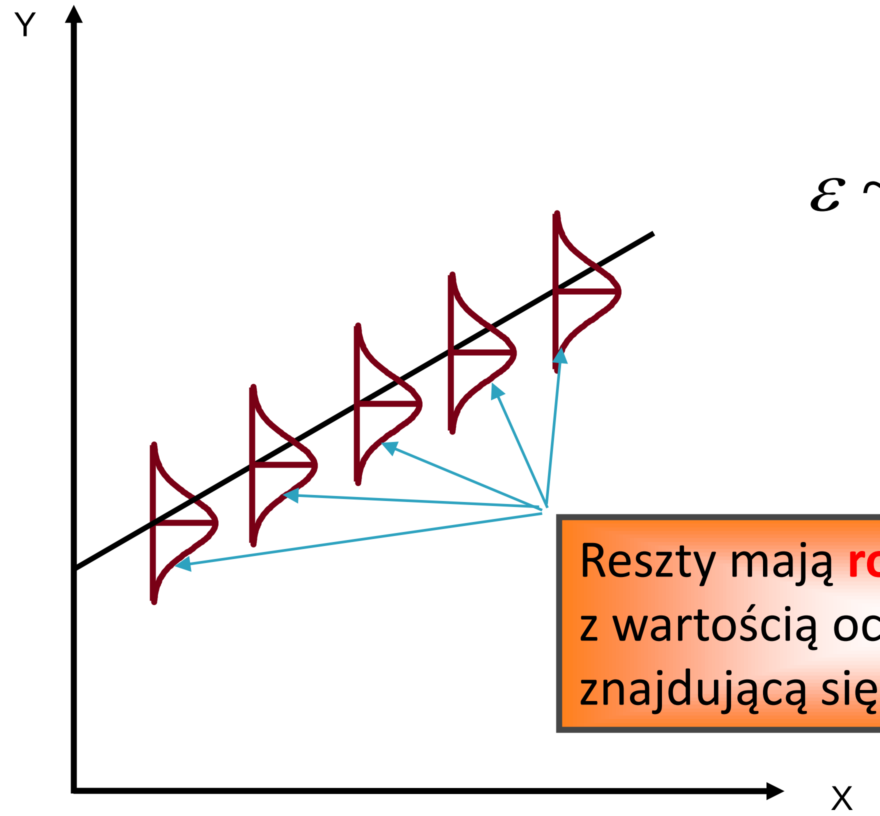
Plan wykładu

- 1. Kombinacja liniowa parametrów
  - Estymator elementu β_k
- 2. Prognozowanie
- 3. Testowanie hipotez
  - Rozkład estymatora **b**

Testowanie hipotez

- Badamy czy hipotezy teoretyczne (wynikające z teorii) znajdują potwierdzenie w danych
- Hipotezy narzucają pewne ograniczenia na wartości parametrów
- Oszacowania parametrów powinny spełniać te ograniczenia w przybliżeniu
- Jeśli oszacowania parametrów odbiegają od postulowanych związków wynikających z teorii to odrzucamy hipotezę jako sprzeczną z danymi
- Uwzględnienie w modelu wiedzy z hipotezy prawdziwej poprawia precyzję oszacowań
- Uwzględnienie w modelu wiedzy z hipotezy fałszywej prowadzi do obciążenia estymatora
- Do testowania hipotez wykorzystujemy testy statystyczne

Dodatkowo założenie klasycznego modelu regresji liniowej



$$\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

Reszty mają **rozkład normalny** z wartością oczekiwaną znajdującą się na linii regresji.

Testowanie hipotez

- Rozkład estymatora **b:**

$$b \sim N(\beta, \sigma^2 (X' X)^{-1})$$

Dziękuję za uwagę