Klasyczny Model Regresji Liniowej
Testowanie hipotez

Stanisław Cichocki

Natalia Nehrebecka

Wykład 9

Plan wykładu

- 1. Estymator wariancji błędu losowego
- 2. Estymator macierzy wariancji i kowariancji **b**
- 3. Kombinacja liniowa parametrów
- 4. Dodatkowe założenie KMRL
- 5. Testowanie hipotez prostych
  - Rozkład estymatora **b**
  - Testowanie hipotez prostych przy użyciu statystyki t

Plan wykładu

- 1. Estymator wariancji błędu losowego
- 2. Estymator macierzy wariancji i kowariancji **b**
- 3. Kombinacja liniowa parametrów
- 4. Dodatkowe założenie KMRL
- 5. Testowanie hipotez prostych
  - Rozkład estymatora **b**
  - Testowanie hipotez prostych przy użyciu statystyki t

Estymator wariancji błędu losowego

```
. regress dochod wiek wiek_2 plec srednie wyzsze

      Source |       SS       df       MS              Number of obs =    1083
-------------+------------------------------           F(  5,  1077) =   22.98
       Model |  72048793.8     5  14409758.8            Prob > F      =  0.0000
    Residual |   675432341  1077  627142.378            R-squared     =  0.0964
-------------+------------------------------           Adj R-squared =  0.0922
       Total |   747481135  1082  690832.842            Root MSE      =  791.92

------------------------------------------------------------------------------
      dochod |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |    36.06131   15.48328     2.33   0.020     5.680494    66.44212
      wiek_2 |   -.3998842   .1973767    -2.03   0.043    -.7871707   -.0125977
        plec |   -338.0671   48.25867    -7.01   0.000    -432.7588   -243.3755
     srednie |    208.5538   77.72619     2.68   0.007     56.04182    361.0657
      wyzsze |    708.2862   99.55596     7.11   0.000     512.9406    903.6318
       _cons |   -26.64989   298.3288    -0.09   0.929    -612.0215    558.7217
------------------------------------------------------------------------------
```

Plan wykładu

- 1. Estymator wariancji błędu losowego
- 2. Estymator macierzy wariancji i kowariancji **b**
- 3. Kombinacja liniowa parametrów
- 4. Dodatkowe założenie KMRL
- 5. Testowanie hipotez prostych
  - Rozkład estymatora **b**
  - Testowanie hipotez prostych przy użyciu statystyki t

Estymator macierzy wariancji i kowariancji b

```
. regress dochod wiek wiek_2 plec srednie wyzsze

      Source |       SS       df       MS              Number of obs =    1083
-------------+------------------------------           F(  5,  1077) =   22.98
       Model |  72048793.8     5  14409758.8            Prob > F      =  0.0000
    Residual |   675432341  1077  627142.378            R-squared     =  0.0964
-------------+------------------------------           Adj R-squared =  0.0922
       Total |   747481135  1082  690832.842            Root MSE      =  791.92

------------------------------------------------------------------------------
      dochod |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |    36.06131   15.48328     2.33   0.020     5.680494    66.44212
      wiek_2 |   -.3998842   .1973767    -2.03   0.043    -.7871707   -.0125977
        plec |   -338.0671   48.25867    -7.01   0.000    -432.7588   -243.3755
     srednie |    208.5538   77.72619     2.68   0.007     56.04182    361.0657
      wyzsze |    708.2862   99.55596     7.11   0.000     512.9406    903.6318
       _cons |   -26.64989   298.3288    -0.09   0.929    -612.0215    558.7217
------------------------------------------------------------------------------
```

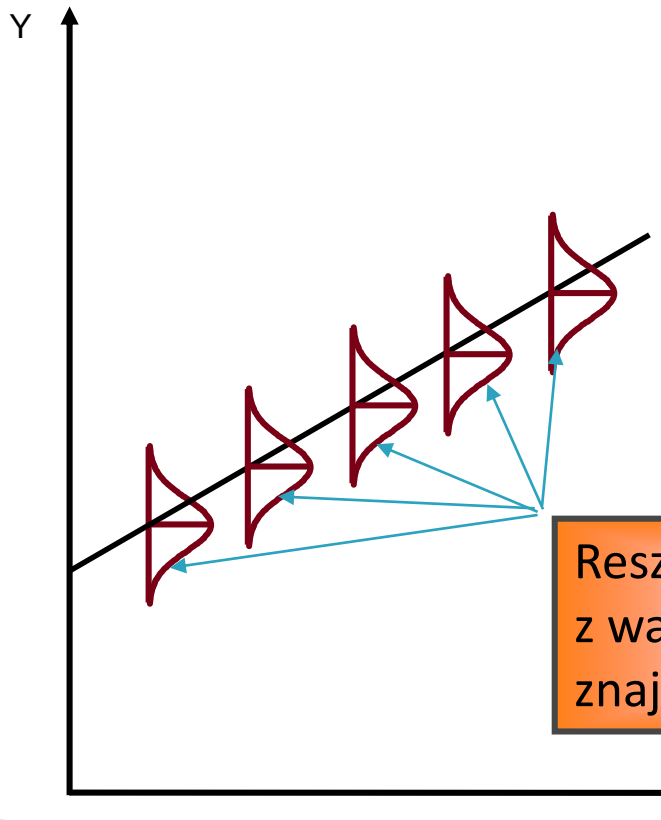
Plan wykładu

- 1. Estymator wariancji błędu losowego
- 2. Estymator macierzy wariancji i kowariancji **b**
- 3. Kombinacja liniowa parametrów
- 4. Dodatkowe założenie KMRL
- 5. Testowanie hipotez prostych
  - Rozkład estymatora **b**
  - Testowanie hipotez prostych przy użyciu statystyki t

Plan wykładu

- 1. Estymator wariancji błędu losowego
- 2. Estymator macierzy wariancji i kowariancji **b**
- 3. Kombinacja liniowa parametrów
- 4. Dodatkowe założenie KMRL
- 5. Testowanie hipotez prostych
  - Rozkład estymatora **b**
  - Testowanie hipotez prostych przy użyciu statystyki t

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.

Plan wykładu

- 1. Estymator wariancji błędu losowego
- 2. Estymator macierzy wariancji i kowariancji **b**
- 3. Kombinacja liniowa parametrów
- 4. Dodatkowe założenie KMRL
- 5. Testowanie hipotez prostych
  - Rozkład estymatora **b**
  - Testowanie hipotez prostych przy użyciu statystyki t

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

Testowanie hipotez prostych

- Rozkład estymatora **b:**

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

- Rozkład pojedynczego elementu tego wektora b_k :

$$b_k \sim N(\beta_k, [\Sigma_b]_{kk})$$

Testowanie hipotez prostych

- Korzystając z rozkładu b_k:

$$\frac{b_k - \beta_k}{\sqrt{[\Sigma_b]_{kk}}} = \frac{b_k - \beta_k}{se(b_k)} \sim N(0,1)$$

- Tej statystyki nie da się policzyć ponieważ macierz Σ_b jest nieznana
- Oszacowaniem tej macierzy jest $\hat{\Sigma}_b$ ale zastosowanie jej w powyższym wzorze wpłynie na rozkład statystyki
- Tak zmodyfikowana statystyka (będziemy ją nazywać **t**) będzie miała rozkład t-studenta

Testowanie hipotez prostych

- Hipoteza prosta: dotyczy pojedynczego parametru modelu albo kombinacji liniowej parametrów

- Załóżmy, że H0: $\beta_k = \beta_k^*$, spełnione są założenia KMRL, błąd losowy ma rozkład normalny i H0 jest prawdziwa, wtedy

$$t = \frac{b_k - \beta_k^*}{\hat{\Lambda}} \sim t_{N-K}$$
$$se(b_k)$$

Testowanie hipotez prostych

- Najczęściej testujemy H0: $\beta_k = \beta_k^*$ przy hipotezie alternatywnej H1: $\beta_k \neq \beta_k^*$ stosując dwustronny obszar krytyczny

- Możliwe także jest testowanie H0: $\beta_k = \beta_k^*$ przy hipotezie alternatywnej H1: $\beta_k > \beta_k^*$ lub H1: $\beta_k < \beta_k^*$ używając jednostronnych obszarów krytycznych

Testowanie istotności poszczególnych zmiennych

- Testowanie prostych hipotez przebiega w następujących krokach:
- Dla modelu:

$$y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \varepsilon_i$$

- którego oszacowaniem jest:

$$\hat{y}_i = b_1 + b_2 X_{2i} + \ldots + b_K X_{Ki}$$

- Krok 1. Stawiamy tak zwaną hipotezę zerową co do wartości nieznanego parametru β_K

$$H_0 : \beta_K = 0 \quad (\text{zmienna } X_{Ki} \text{ jest nieistotna})$$

- Hipotezie tej towarzyszy hipoteza alternatywna:

$$H_1 : \beta_K \neq 0 \quad (\text{zmienna } X_{Ki} \text{ jest istotna})$$

Testowanie istotności poszczególnych zmiennych

Krok 2. Przy założeniu, że postawiona hipoteza zerowa jest prawdziwa, wyznaczamy statystykę testową z rozkładu *t - Studenta* o N - K stopniach swobody postaci:

$$t = \frac{b_K}{\hat{se}(b_K)}$$

Gdzie:

$\hat{se}(b_K)$ - oszacowanie odchylenia standardowego estymatora b_K

Testowanie istotności poszczególnych zmiennych

Krok 3. Odczytujemy z tablic rozkładu *t-Studenta* wartość krytyczną (α - poziom istotności[1)])

$$t^* = t\left(N - K; 1 - \frac{\alpha}{2}\right)$$

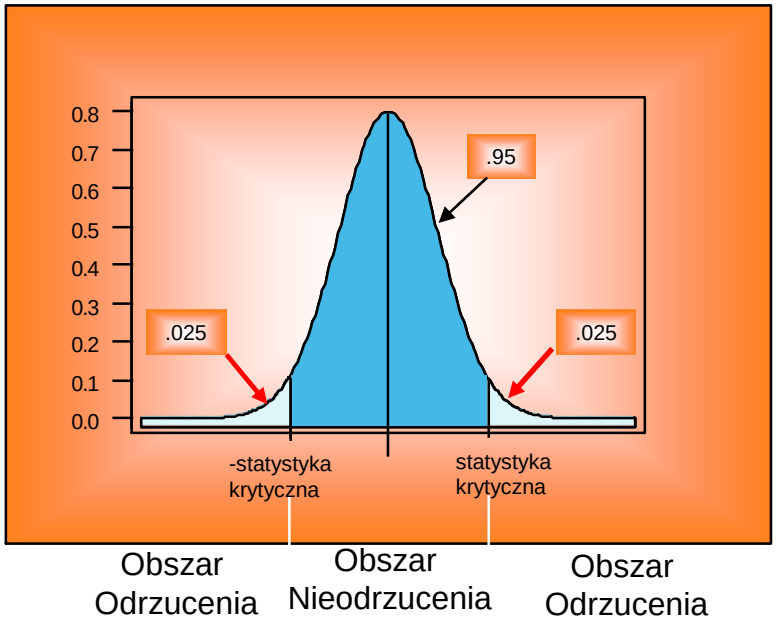
Stopni swobody

Rząd kwantyla

[1)] *maksymalne dopuszczalne prawdopodobieństwo popełnienia błędu polegającego na odrzuceniu prawdziwej hipotezy zerowej*

Testowanie istotności poszczególnych zmiennych

Krok 4. Podjęcie decyzji



- ✓ $|t| \geq t^*$ - odrzucamy hipotezę zerową, czyli zmienna X_{Ki} jest istotna.
- ✓ $|t| < t^*$ - nie ma podstaw do odrzucenia hipotezy zerowej, czyli zmienna X_{Ki} jest nieistotna.

Testowanie istotności poszczególnych zmiennych

- Przykład

xi: reg wynagrodzenie i.plec i.wyksztalcenie godziny wiek szara dorywcza

```
      Source |       SS       df       MS              Number of obs =   26352
-------------+------------------------------           F(  9, 26342) =   66.80
       Model |  3.7557e+11     9  4.1730e+10           Prob > F      =  0.0000
    Residual |  1.6457e+13 26342   624728699           R-squared     =  0.0223
-------------+------------------------------           Adj R-squared =  0.0220
       Total |  1.6832e+13 26351   638768004           Root MSE      =   24995

------------------------------------------------------------------------------
wynagrodze~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _Iplec_1 |  -1795.469   325.0304    -5.52   0.000    -2432.546   -1158.391
_Iwyksztal~2 |  -4386.364   683.3584    -6.42   0.000    -5725.783   -3046.944
_Iwyksztal~3 |  -5950.806   557.4312   -10.68   0.000    -7043.401   -4858.211
_Iwyksztal~4 |  -8167.496   538.4532   -15.17   0.000    -9222.893   -7112.099
_Iwyksztal~5 |  -9698.71    578.6504   -16.76   0.000     -10832.9   -8564.524
     godziny |  -.3193543   14.63862    -0.02   0.983    -29.01183    28.37312
        wiek |  -95.59548   13.53115    -7.06   0.000    -122.1173    -69.0737
   _Iszara_1 |   11363.98   1571.524     7.23   0.000      8283.71    14444.25
    dorywcza |  -8008.054   742.8795   -10.78   0.000    -9464.138   -6551.97
       _cons |   18979.6   974.8196    19.47   0.000      17068.9     20890.3
```

Testowanie istotności poszczególnych zmiennych

- W popularnych pakietach ekonometrycznych obok wyliczonej wartości statystyki **t** podawane jest również odpowiadające mu **prawdopodobieństwo p** , że $\beta_k = 0$. Oznaczane ono jest z angielskiego przez $p - value$.
- W przypadku hipotez dwustronnych:

$$p = 2[1 - F(k^*)]$$

gdzie: F- dystrybuanta rozkładu, k* - wartość statystyki testowej

- W przypadku hipotez jednostronnych:

$$p = 1 - F(k^*)$$

gdzie: F- dystrybuanta rozkładu, k* - wartość statystyki testowej

Testowanie istotności poszczególnych zmiennych

- ✓ Jeśli p-value $\leq \alpha$ (poziomu istotności), to odrzucamy hipotezę zerową.
- ✓ Jeśli p-value $> \alpha$ (poziomu istotności), to brak podstaw do odrzucenia hipotezy zerowej.

Testowanie istotności poszczególnych zmiennych

- Przykład

xi: reg wynagrodzenie i.plec i.wyksztalcenie godziny wiek szara dorywcza

```
      Source |       SS       df       MS              Number of obs =   26352
-------------+------------------------------           F(  9, 26342) =   66.80
       Model |  3.7557e+11     9  4.1730e+10            Prob > F      =  0.0000
    Residual |  1.6457e+13 26342   624728699           R-squared     =  0.0223
-------------+------------------------------           Adj R-squared =  0.0220
       Total |  1.6832e+13 26351   638768004           Root MSE      =   24995

------------------------------------------------------------------------------
wynagrodze~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _Iplec_1 |  -1795.469   325.0304    -5.52   0.000    -2432.546   -1158.391
_Iwyksztal~2 |  -4386.364   683.3584    -6.42   0.000    -5725.783   -3046.944
_Iwyksztal~3 |  -5950.806   557.4312   -10.68   0.000    -7043.401   -4858.211
_Iwyksztal~4 |  -8167.496   538.4532   -15.17   0.000    -9222.893   -7112.099
_Iwyksztal~5 |  -9698.71    578.6504   -16.76   0.000     -10832.9   -8564.524
     godziny |  -.3193543   14.63862    -0.02   0.983    -29.01183    28.37312
        wiek |  -95.59548   13.53115    -7.06   0.000    -122.1173   -69.0737
   _Iszara_1 |   11363.98   1571.524     7.23   0.000      8283.71   14444.25
    dorywcza |  -8008.054   742.8795   -10.78   0.000    -9464.138   -6551.97
       _cons |   18979.6   974.8196    19.47   0.000      17068.9    20890.3
```

Pytania teoretyczne

1. Pokazać, że s^2 jest nieobciążonym estymatorem σ^2 .
2. Udowodnić, że $s^2(X'X)^{-1}$ jest nieobciążonym estymatorem VAR(b).
3. Podać postać estymatora $\delta'\beta$ dla kombinacji liniowej i udowodnić, że jest on nieobciążony.
4. Wyprowadzić rozkład małopróbkowy estymatora MNK. Jakie założenie, poza standardowymi założeniami KMRL, należy w tym przypadku przyjąć?
5. Jaką postać ma statystyka służąca do testowania hipotezy o tym, że $\beta_k = \beta_k^*$?

Dziękuję za uwagę