Weryfikacja hipotez statystycznych

Część 2. Hipotezy łączne

Hipoteza złożona

- Hipoteza łączna

$$H_0 : \; R\beta = q$$

- W praktyce weryfikowana jest hipoteza

$$H_0 : \; R\beta - q = 0$$

przeciwko alternatywie

$$H_1 : \; R\beta - q \neq 0$$

Przykładowe ograniczenia

1. Jeden ze współczynników wektora β jest równy zero, $\beta_j = 0$.

$$R = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix}, q = [0]$$

2. Dwa współczynniki równania regresji są sobie równe, $\beta_j = \beta_k$:

$$R = \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \end{bmatrix}, q = [0]$$

3. Kilka współczynników równania regresji sumuje się do liczby k, $\beta_2 + \beta_3 + \beta_4 = k$:

$$R = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & \dots & 0 \end{bmatrix}, q = [k]$$

Przykładowe ograniczenia

4. Kilka współczynników równania regresji jest równych zero, $\beta_1 = 0, \quad \beta_2 = 0, \quad \beta_3 = 0$:

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \end{bmatrix} \qquad q = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

5. Kilka ograniczeń liniowych nałożonych na współczynniki regresji, $\beta_2 + \beta_3 = 1, \quad \beta_4 + \beta_6 = 0, \quad \beta_5 + \beta_6 = 0.$

$$R = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \qquad q = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

6. Wszystkie współczynniki modelu są równe zero.

Ograniczenia parametrów

- Równanie zarobków

$$ln(zarobki) = \beta_0 + \beta_1 plec + \beta_2 wiek + \beta_3 wiek2 + \varepsilon$$

- Weryfikowana jest hipoteza o braku wpływu wieku na wysokość zarobków

$$H_0 : \begin{cases} \beta_2 = 0 \\ \beta_3 = 0 \end{cases}$$

- W tym przypadku macierz ograniczeń ma następującą postać

$$R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad q = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Ograniczenia parametrów

- Funkcja produkcji Cobb-Douglasa

$$y = \beta_0 + \beta_1 l + \beta_2 k + \varepsilon$$

- Weryfikowana jest hipoteza o stałych przychodach skali

$$H_0 : \left\{ \beta_1 + \beta_2 = 1 \right.$$

- W tym przypadku macierz ograniczeń ma następującą postać

$$R = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad q = 1$$

- Ograniczenia nakładane na parametry modelu

$$m = R\beta - q$$

- Rozkład wektora ograniczeń

$$m \sim \mathcal{N}(E(R\beta - q), R\Sigma_b R')$$

- Przy spełnionych ograniczeniach

$$m \sim \mathcal{N}(0, R\Sigma_b R')$$

Statystyka testu Walda

- Przy założeniu znanej wariancji

$$W = m'(R\Sigma_b R')m = \frac{(Rb-q)'[R(X'X)^{-1}R'/J]^{-1}(Rb-q)}{\sigma^2} \sim \chi^2_J$$

- Gdy wariancję należy oszacować

$$\frac{((Rb-q)'[R(X'X)^{-1}R']^{-1}(Rb-q))/J}{\frac{e'e}{N-k}} \sim F_{J,N-k}$$

Wyprowadzenie

- Maksymalizowana jest wartość funkcja z ograniczeniem

$$\min_{b_R} (y - X'b_R)'(y - X'b_R) \text{ pw. } Rb_R = q$$

- Odpowiadająca mu funkcja Lagrange'a ma postać

$$L(b_R) = (y - X'b_R)'(y - X'b_R) - \lambda'(Rb_R - q)$$

Własności estymatora z ograniczeniami

- Estymator z narzuconymi ograniczeniami

$$b_R = b - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(Rb - q)$$

- Wariancja estymatora z narzuconymi ograniczeniami

$$var(b_R) = var(b) - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(X'X)^{-1}\sigma^2$$

Weryfikacja ograniczeń

- Reszty przy ograniczeniach

$$e_R = e + X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(Rb - q)$$

- Oszacowanie wariancji

$$e_R' e_R = e'e + (Rb - q)'[R(X'X)^{-1}R']^{-1}(Rb - q)$$

- Po wstawieniu warunku do wzoru na statystykę F

$$F = \frac{(e_R' e_R - e'e)/J}{e'e/(N-k)}$$

Regresja podzielona

- Zapiszmy podzieloną regresję

$$y = X'\beta + \varepsilon = X_1'\beta_1 + X_2'\beta_2 + \varepsilon$$

- Estymator dla β_1 wynosi

$$b_1 = (X_1' M_{X_2} X_1)^{-1} X_1' M_{X_2} y$$

$$y = X_1\beta_1 + \varepsilon_1 \tag{1}$$

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon_2 \tag{2}$$

- Wnioski

Model Solowa

Próba	Non-oil	Intermediate	OECD
obs	98	75	22
CONST	5,84(1,59)	5,36(1,55)	7,97(2,48)
$ln(I/GDP)$	1,42(0,14)	1,31(0,17)	0,50(0,43)
$ln(n+g+\delta)$	-1,97(0.56)	-2,01(0.53)	-0,76(0.84)
R^2	0,59	0,59	0,01
sse	0,69	0,61	0,38
CONST	6,87(0,12)	7,10(0,15)	8,62(0,53)
$ln(I/GDP) - ln(n+g+\delta)$	1,48(0,12)	1,43(0,14)	0,56(0,36)
R^2	0,59	0,59	0,06
sse	0,69	0,61	0,37
p-value testu ograniczeń	0,38	0,26	0,79
implikowana α	0,60(0,02)	0,59(0,02)	0,36(0,15)

Żrdóło: Mankiw, Romer, Weil (1992)

Source	SS	df	MS
Model	709.024852	16	44.3140532
Residual	1845.2104	16145	.114289898
Total	2554.23525	16161	.158049332

Number of obs	=	16162
F(16, 16145)	=	387.73
Prob > F	=	0.0000
R-squared	=	0.2776
Adj R-squared	=	0.2769
Root MSE	=	.33807

lzarobki	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_Iplec_2	-.2831618	.0056319	-50.28	0.000	-.2942009	-.2721226
wiek	.0282415	.0019491	14.49	0.000	.024421	.0320619
wiek2	-.0002838	.0000251	-11.31	0.000	-.000333	-.0002346
_Iwyksztal~2	-.2503001	.0148188	-16.89	0.000	-.2793466	-.2212536
_Iwyksztal~3	-.2569524	.0093107	-27.60	0.000	-.2752023	-.2387025
_Iwyksztal~4	-.232063	.0124627	-18.62	0.000	-.2564912	-.2076347
_Iwyksztal~5	-.3774886	.0091844	-41.10	0.000	-.3954911	-.3594862
_Iwyksztal~6	-.4872175	.0107408	-45.36	0.000	-.5082708	-.4661643
_Iwyksztal~7	-.6097186	.0781733	-7.80	0.000	-.7629469	-.4564903
_Iklm_12_1	-.0336191	.0092216	-3.65	0.000	-.0516944	-.0155437
_Iklm_12_2	-.0724722	.0088013	-8.23	0.000	-.0897237	-.0552207
_Iklm_12_3	-.0694849	.0107082	-6.49	0.000	-.0904742	-.0484956
_Iklm_12_4	-.117864	.0136519	-8.63	0.000	-.1446233	-.0911048
_Iklm_12_5	-.1120795	.0196574	-5.70	0.000	-.1506103	-.0735488
_Iklm_12_6	-.0972314	.0777095	-1.25	0.211	-.2495506	.0550878
_Iklm_12_9	-.1370085	.0069922	-19.59	0.000	-.1507141	-.123303
_cons	5.744772	.0376436	152.61	0.000	5.670987	5.818558

- Weryfikowana jest hipoteza o braku wpływu wieku na wysokość zarobków

```
. test wiek + wiek2=0

 ( 1)  wiek + wiek2 = 0

       F(  1, 16145) =  211.09
            Prob > F =    0.0000
```

- Weryfikowana jest hipoteza o braku wpływu miejsca zamieszkania na wysokość zarobków

```
. test _Iklm_12_1 _Iklm_12_2 _Iklm_12_3 _Iklm_12_4 _Iklm_12_5 _Iklm_12_6 _Iklm_12_9

 ( 1)  _Iklm_12_1 = 0
 ( 2)  _Iklm_12_2 = 0
 ( 3)  _Iklm_12_3 = 0
 ( 4)  _Iklm_12_4 = 0
 ( 5)  _Iklm_12_5 = 0
 ( 6)  _Iklm_12_6 = 0
 ( 7)  _Iklm_12_9 = 0

       F(  7, 16145) =   60.43
            Prob > F =    0.0000
```

```
Constrained linear regression                      Number of obs =    16162
                                                   F( 15, 16146) =   394.38
                                                   Prob > F      =   0.0000
 ( 1)  wiek + wiek2 = 0                            Root MSE      =   .34026
------------------------------------------------------------------------------
    lzarobki |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Iplec_2 |  -.2785073   .0056592   -49.21   0.000        -.2896   -.2674145
        wiek |  -.0000767   3.75e-06   -20.49   0.000     -.0000841   -.0000694
       wiek2 |   .0000767   3.75e-06    20.49   0.000      .0000694    .0000841
_Iwyksztal~2 |  -.2512886   .0149147   -16.85   0.000     -.2805232   -.2220541
_Iwyksztal~3 |  -.2584555   .0093705   -27.58   0.000     -.2768226   -.2400883
_Iwyksztal~4 |  -.2378895    .012537   -18.97   0.000     -.2624634   -.2133156
_Iwyksztal~5 |  -.3826277   .0092371   -41.42   0.000     -.4007334   -.3645219
_Iwyksztal~6 |  -.4976874   .0107861   -46.14   0.000     -.5188294   -.4765454
_Iwyksztal~7 |  -.6622477    .078596    -8.43   0.000     -.8163046   -.5081908
  _Iklm_12_1 |  -.0311869   .0092799    -3.36   0.001     -.0493765   -.0129974
  _Iklm_12_2 |  -.0708824   .0088577    -8.00   0.000     -.0882445   -.0535204
  _Iklm_12_3 |  -.0673703   .0107767    -6.25   0.000     -.0884938   -.0462468
  _Iklm_12_4 |  -.1138371   .0137376    -8.29   0.000     -.1407644   -.0869099
  _Iklm_12_5 |  -.1101222   .0197844    -5.57   0.000     -.1489019   -.0713425
  _Iklm_12_6 |  -.0923709   .0782127    -1.18   0.238     -.2456764    .0609347
  _Iklm_12_9 |  -.1365614   .0070375   -19.40   0.000     -.1503557   -.1227671
       _cons |   6.267933   .0110455   567.46   0.000      6.246283    6.289584
------------------------------------------------------------------------------
```

- $RSS_U = 1845.2104$
- $RSS_R = 1869.3333$
- więc wartość statystyki F wynosi

$$F = \frac{(e_R' e_R - e' e)/J}{e' e/(N-k)} = \frac{(1869{,}3333 - 1845{,}2104)/1}{1845{,}2104/16145} \approx 211$$