

Testowanie hipotez statystycznych (część II)

Wykład 10

Natalia Nehrebecka Stanisław Cichocki

20 grudnia 2014

Plan zajęć

- 1 Testowanie hipotez łącznych
 - Hipotezy łączne - test F

- 2 Pytania teoretyczne

Hipotezy łączne są ważne z punktu widzenia:

- rozważań teoretycznych
- doboru zmiennych do modelu

Uwaga:

Hipotezy łączne nie są równoważne iloczynowi hipotez prostych!

Typowa hipoteza łączna

dana jest układem równań:

$$H_0 : \mathbf{H}\beta = \mathbf{h}$$

gdzie: \mathbf{H} - macierz o pełnym rzędzie wierszowym $= g$.

Liczba równań w tym układzie nazywana jest **liczbą ograniczeń**

Układ równań:

- zawiera równania liniowo niezależne
- nie jest sprzeczny

Zadanie

W modelu:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$$

testowana jest hipoteza H_0 postaci:

$$H_0 : \begin{cases} \beta_0 = 0 \\ \beta_1 = \beta_2 \\ \beta_2 + \beta_3 = 1 \end{cases}$$

Znaleźć macierze \mathbf{H} i \mathbf{h} za pomocą których hipotezę H_0 można zapisać jako $\mathbf{H}\boldsymbol{\beta} = \mathbf{h}$.

Statystyka testowa

$$F = \frac{(S_R - S)/g}{S/(N - K)} = \frac{(R^2 - R_R^2)/g}{(1 - R^2)/(N - K)} \sim F(g, N - K)$$

gdzie:

$S_R = \mathbf{e}'_R \mathbf{e}_R$ - suma kwadratów reszt w modelu z narzuconymi ograniczeniami, wynikającymi z hipotezy zerowej $H_0 : \mathbf{H}\beta = \mathbf{h}$

$S = \mathbf{e}'\mathbf{e}$ - suma kwadratów reszt w modelu bez ograniczeń

Interpretacja

Ograniczenia w modelu utrudniają minimalizację kwadratów reszt, z tego powodu $\mathbf{e}'_R \mathbf{e}_R \geq \mathbf{e}'\mathbf{e}$, zatem:

- wprowadzenie ograniczeń zawsze powoduje pogorszenie dopasowania (mierzone R^2)
- znaczne pogorszenie dopasowania \Rightarrow wysoka wartość statystyki F
- wysoka wartość statystyki \Rightarrow odrzucenie hipotezy zerowej

Statystyka krytyczna

Odczytujemy z tablic rozkładu F wartość krytyczna (α - poziom istotności) :

$$F^* = F(g, N - K)$$

Podjęcie decyzji

- $F \geq F^*$ - odrzucamy hipotezę zerowa
- $F < F^*$ - nie ma podstaw do odrzucenia hipotezy zerowej

Przykład (1/3)

```
. xi: regress dochod wiek wiek_2 plec i.miasto i.wykształcenie
i.miasto      _Imiasto_1-8      (naturally coded; _Imiasto_1 omitted)
i.wykształcenie  _Iwykształc_1-3  (naturally coded; _Iwykształc_1 omitted)
```

Source	SS	df	MS	Number of obs =	1083
Model	86357666.3	12	7196472.19	F(12, 1070) =	11.65
Residual	661123469	1070	617872.401	Prob > F =	0.0000
				R-squared =	0.1155
				Adj R-squared =	0.1056
Total	747481135	1082	690832.842	Root MSE =	786.05

dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
wiek	36.75887	15.43179	2.38	0.017	6.478863 67.03888
wiek_2	-.4268203	.1969836	-2.17	0.030	-.8133382 -.0403023
plec	-357.2113	48.13809	-7.42	0.000	-451.6671 -262.7555
_Imiasto_2	69.51889	107.7679	0.65	0.519	-141.9416 280.9793
_Imiasto_3	158.9295	83.95361	1.89	0.059	-5.802896 323.6619
_Imiasto_4	376.3989	101.0848	3.72	0.000	178.052 574.7458
_Imiasto_5	162.5285	83.82589	1.94	0.053	-1.953291 327.0103
_Imiasto_6	219.3064	79.72788	2.75	0.006	62.86565 375.7471
_Imiasto_7	237.4997	96.73371	2.46	0.014	47.69038 427.309
_Imiasto_8	302.6196	86.98608	3.48	0.001	131.9369 473.3022
_Iwykształc~2	120.6223	79.8898	1.51	0.131	-36.1362 277.3807
_Iwykształc~3	562.1802	104.2811	5.39	0.000	357.5616 766.7988
_cons	-69.92963	297.1081	-0.24	0.814	-652.9102 513.0509

Przykład (2/3)

- Czy zmienne objaśniające są łącznie istotne?
- Czy wykształcenie średnie i wyższe są łącznie istotne w modelu?

Przykład (3/3)

```
. xi: regress dochod wiek wiek_2 plec i.miasto
i.miasto      _Imiasto_1-8      (naturally coded; _Imiasto_1 omitted)
```

Source	SS	df	MS	Number of obs =	1083
Model	62279436.7	10	6227943.67	F(10, 1072) =	9.74
Residual	685201699	1072	639180.689	Prob > F =	0.0000
Total	747481135	1082	690832.842	R-squared =	0.0833
				Adj R-squared =	0.0748
				Root MSE =	799.49

dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiek	40.66732	15.66769	2.60	0.010	9.924507	71.41014
wiek_2	-.4767307	.1996926	-2.39	0.017	-.8685633	-.0848981
plec	-361.8976	48.95487	-7.39	0.000	-457.9558	-265.8393
_Imiasto_2	101.4421	108.9616	0.93	0.352	-112.36	315.2443
_Imiasto_3	235.5312	83.65064	2.82	0.005	71.39359	399.6687
_Imiasto_4	453.9829	101.1928	4.49	0.000	255.4246	652.5413
_Imiasto_5	228.7915	84.07216	2.72	0.007	63.82688	393.7562
_Imiasto_6	294.8323	79.05281	3.73	0.000	139.7165	449.9481
_Imiasto_7	330.7518	96.72832	3.42	0.001	140.9535	520.5501
_Imiasto_8	421.8666	85.47443	4.94	0.000	254.1504	589.5827
_cons	-33.80157	297.1458	-0.11	0.909	-616.8549	549.2518

- 1 Jak należy testować hipotezę postaci $H_0 : \mathbf{H}\beta = \mathbf{h}$, używając do tego sum kwadratów reszt z modelu bez ograniczeń i z ograniczeniami?
- 2 Wyjaśnić, jakie korzyści i niebezpieczeństwa łączą się z narzucaniem ograniczeń na model.