# Ekonometria

## Weryfikacja modelu

Paweł Cibis
*pcibis@o2.pl*

6 kwietnia 2006

# Test *F* Fishera-Snedecora

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$$

$$H_1 : |\alpha_1| + |\alpha_2| + \ldots + |\alpha_n| \neq 0$$

$$F = \frac{(n - m - 1)}{m} * \frac{R^2}{1 - R^2}$$

$$F^*_{\alpha,m,n-m-1}$$

$F > F^*$ powoduje odrzucenie $H_0$, w przeciwnym wypadku nie ma podstaw do jej odrzucenia.

# Test *t-Studenta*

$$H_0 : \alpha_i = 0$$

$$H_1 : \alpha_i \neq 0$$

$$t_i = \frac{\alpha_i}{S(\alpha_i)}$$

$$t^*_{\alpha,n-m-1}$$

$|t_i| > t^*$ powoduje odrzucenie $H_0$, w przeciwnym wypadku nie ma podstaw do jej odrzucenia.

# Badanie istotności – Test *F* *Fishera-Snedecora*

1. Narzędzia/Analiza Danych. . . /Regresja
2. Wartość statystyki $F$ znajduje się w części „Analiza wariancji"
3. Wartość statystyki teoretycznej $F^*$:
   ROZKŁAD.F.ODW($\alpha$; $m$; $n - m - 1$)
4. Istotność F – prawdopodobieństwo tego, że $H_0$ jest prawdziwa; jeżeli wartość ta jest mniejsza od przyjętego poziomu istotności, należy odrzucić $H_0$ na rzecz $H_1$.

# Badanie istotności – Test *t-Studenta*

1. Narzędzia/Analiza Danych.../Regresja
2. Wartości statystyk $t_i$ – tabelka z oszacowaniami współczynników, kolumna „t-stat"
3. Wartość statystyki teoretycznej $t^*$: ROZKŁAD.T.ODW($\alpha$; $n - m - 1$)
4. Wartość-p – prawdopodobieństwo tego, że $H_0$ jest prawdziwa; jeżeli wartość ta jest mniejsza od przyjętego poziomu istotności, należy odrzucić $H_0$ na rzecz $H_1$.

# Uwagi

- Warunkiem badania istotności parametrów strukturalnych jest spełnienie założenia o normalności rozkładu reszt.
- W modelu nieliniowym sprowadzalnym do liniowego istotność parametrów jest oceniana dla postaci transformowanej.

# Test dla małej próby

$$H_0 : \varepsilon_t \text{ losowy}$$

$$H_1 : \varepsilon_t \text{ nielosowy}$$

$$e_t \leftarrow A, \quad e_t > 0$$

$$e_t \leftarrow B, \quad e_t < 0$$

$K_e$ – liczba serii

$K_1$ – z tablic rozkładu liczby serii dla danych: $\alpha, n_1, n_2$

$K_2$ – z tablic rozkładu liczby serii dla danych: $1 - \alpha, n_1, n_2$

$n_1, n_2$ – liczba symboli A i B (kolejność nie ma znaczenia)

# Test dla małej próby – obszary decyzyjne

Jeżeli

$$K_1 \leqslant K_e \leqslant K_2,$$

to brakuje podstaw do odrzucenia $H_0$. W przeciwnym wypadku hipotezę o losowości rozkładu elementu losowego należy odrzucić. W przypadku testu jednostronnego $H_0$ odrzucamy gdy zachodzi $K_e < K^*$. $K^*$ możemy odczytać z tablic testu dwustronnego dla poziomu istotności $2\alpha$ lub ze specjalnych tablic testu jednostronnego.

# Test dla małej próby (dwustronny) – Excel

1. Kodujemy reszty: JEŻELI(komórka$>$ 0;"A";"B").
2. Sprawdzamy czy nie ma reszt $= 0$: LICZ.JEŻELI(zakres_reszt;"=0") i w razie czego przyporządkowujemy im kody wg ustalonej wcześniej reguły.
3. Liczymy liczbę serii ($K_e$): JEŻELI($kod_{i-1} <> kod_i$;1;0) (w pierszym wierszu „1").
4. Liczymy liczbę reszt dodatnich ($n_1$): LICZ.JEŻELI(zakres_kodów;"A").
5. Liczymy liczbę reszt ujemnych ($n_2$): LICZ.JEŻELI(zakres_kodów;"B").
6. Dla przyjętego poziomu istotności $\alpha$ oraz $n_1$ i $n_2$ odczytujemy z tablic $K_1$ oraz $K_2$ i otrzymujemy wynik testu, porównując je z $K_e$.

# Test dla dużej próby

$$H_0 : \varepsilon_t \text{ losowy}$$

$$H_1 : \varepsilon_t \text{ nielosowy}$$

$$P(N_n \geqslant 1) = 1 - \Phi\left(\frac{(1 - \frac{n}{\mu})\mu^{\frac{3}{2}}}{\sigma\sqrt{n}}\right)$$

$$\mu = \frac{1 - 0,5^r}{0,5^{r+1}}$$

$$\sigma = \sqrt{2^{2r+2} - (2r+1)2^{r+1} - 2}$$

# Test dla dużej próby – Excel

Obliczanie najdłuższej serii $r$:

1. Kodujemy reszty dodatnie jako 1, a ujemne jako 0: JEŻELI(reszta>0;"1";"0").
2. Liczymy narastająco długości serii: W pierwszym wierszu wstawiamy "1", w następnych: JEŻELI($kod_{i-1} = kod_i$;$dugo_{i-1} + 1$;1);
3. Jako $r$ przyjmujemy maksimum z długości.

# Test dla dużej próby – Excel

Obliczanie prawdopodobieństwa pojawienia się serii dłuższej niż $r$:

1. Obliczamy $\mu$ i $\sigma$, a następnie $P(N_n \geqslant 1)$.
2. Jeżeli $P(N_n \geqslant 1) \geqslant \alpha$, nie ma podstaw do odrzucenia $H_0$. W przeciwnym wypadku odrzucamy hipotezę o losowości reszt.

# Test *Durbina-Watsona*

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

$$DW = \frac{\sum_{t=1}^{n-1}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

$d_L, \; d_U$ wartości krytyczne odczytane z tablic dla testu *Durbina-Watsona*

# Test *Durbina-Watsona* – obszary decyzyjne

- $0 \leqslant DW < d_L$ – odrzucamy $H_0$, autokorelacja dodatnia
- $d_L \leqslant DW \leqslant d_U$ – obszar niekonkluzywności
- $d_U < DW < 4 - d_U$ – nie ma podstaw do odrzucenia $H_0$
- $4 - d_U \leqslant DW \leqslant 4 - d_L$ – obszar niekonkluzywności
- $4 - d_L < DW \leqslant 4$ – odrzucamy $H_0$, autokorelacja ujemna

# Test *Durbina-Watsona* – Excel

1. Tworzymy wektor kolumnowy reszt.
2. Kopiujemy go do sąsiedniej kolumny o 1 wiersz niżej.
3. Liczymy różnicę odpowiadających sobie reszt w poszczególnych wierszach (oprócz pierwszego i ostatniego).
4. Liczymy sumę kwadratów różnic reszt i sumę kwadratów reszt, a następnie dzielimy je przez siebie, otrzymując statystykę DW.
5. Z tablic testu *Durbina-Watsona* odczytujemy wartości krytyczne $d_L$ oraz $d_U$ i podejmujemy decyzję o ewentualnym odrzuceniu $H_0$.

# Test *Goldfelda-Quandta* – Hipotezy

Na podstawie wykresu kwadratów reszt oceniamy, czy reszty modelu da się podzielić na 2 części – początkową i końcową – o wyraźnie różnych wartościach kwadratów reszt. Jeżeli tak, to testujemy hipotezę o równości wariancji w obu częściach:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 > \sigma_2^2$$

$$F_e = \frac{s_1^2}{s_2^2}$$

# Test *Goldfelda-Quandta* – Obszar krytyczny

$F_e > F_\alpha(n_1 - m - 1, n_2 - m - 1)$ odczytujemy z tablic rozkładu F. Jeżeli

$$F_\alpha(n_1 - m - 1, n_2 - m - 1)$$

$H_0$ należy odrzucić – element losowy jest heteroscedastyczny.

# Test *Goldfelda-Quandta* – Inna postać $H_1$

Jeżeli $H_1$ ma postać

$$H_1 : \sigma_1^2 < \sigma_2^2$$

To statystyka testowa jest postaci

$$F_e = \frac{s_2^2}{s_1^2},$$

a wartość krytyczna

$$F_\alpha(n_2 - m - 1, n_1 - m - 1).$$

Obszary krytyczne nie ulegają zmianie.
W obu przypadkach grupy reszt numerujemy tak, by licznik statystyki $F_e$ był większy od mianownika.

# Test *Goldfelda-Quandta* – Excel

1. Konstruujemy wykres kolumnowy dla kwadratów reszt.
2. Dzielimy reszty na dwie grupy.
3. Szacujemy parametry modelu dla każdej z grup (Analiza Danych lub REGLINP), liczymy kwadraty jego reszt i tworzymy ich kolejny wykres kolumnowy.
4. Dla każdej grupy liczymy sumę kwadratów reszt (SUMA.KWADRATÓW)
5. Obliczamy $s_1^2 = \frac{\sum_{i \in A} e_i^2}{n_1 - m - 1}$ oraz $s_2^2 = \frac{\sum_{i \in B} e_i^2}{n_2 - m - 1}$.
6. Zależnie od postaci $H_1$ obliczamy z odpowiednią wartość $F_e$ i $F_\alpha$ – ROZKŁAD.F.ODW i podejmujemy decyzję odnośnie $H_0$.

# Test *Hellwiga* – mała próba

$H_0$ : reszty mają rozkład normalny

$H_1$ : reszty mają inny rozkład

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_t^2}$$

$$e'_t = \frac{e_t}{s}$$

$K_1 \leqslant K \leqslant K_2$ – nie ma podstaw do odrzucenia $H_0$.

# Test *Hellwiga* – Excel

1. Sortujemy rosnąco kolumnę reszt i liczymy ich kwadraty.
2. Standaryzujemy reszty ($e'_t$) i sortujemy je rosnąco.
3. Liczymy wartości dystrybuant reszt – ROZKŁAD.NORMALNY.S(X).
4. Tworzymy przedziały („cele"): $[0, \frac{1}{n}), [\frac{1}{n}, \frac{2}{n}), \ldots, [\frac{n-1}{n}, \frac{n}{n}]$.
5. Zliczamy ile wartości dystrybuant wpada do każdej celi.
6. Obliczamy liczbę pustych cel ($K$) – LICZ.JEŻELI(zakres;0).
7. Z tablic testu Hellwiga odczytujemy wartości $K_1$ i $K_2$ i podejmujemy decyzję odnośnie $H_0$

# Test *Kołmogorowa* – duża próba

$$H_0 : \varepsilon_t \,\tilde{}\, N(0, s), \quad (t = 1, \dots, n)$$

$H_1$ : reszty mają inny rozkład

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_t^2}$$

$$D_n = sup_x|F_n(x) - F_0(x)|$$

$$\lambda_e = \sqrt{n}D_n$$

$H_0$ odrzucamy dla $\lambda_e > \lambda_0$, gdzie $\lambda_0$ odczytujemy z tablic rozkładu $\lambda_0$-*Kołmogorowa* dla $Q(\lambda) = 1 - \alpha$.

# Test *Kołmogorowa* – Excel

1. Sortujemy rosnąco kolumnę reszt i liczymy ich kwadraty.
2. Standaryzujemy reszty ($e'_t$) i sortujemy je rosnąco.
3. Liczymy wartości dystrybuant empirycznych (numer reszty / liczba reszt).
4. Liczymy wartości dystrybuant reszt – ROZKŁAD.NORMALNY.S(X).
5. Liczymy moduły różnic pomiędzy wartościami dystrybuant i wybieramy największy z wyników ($D_n$).
6. Liczymy $\lambda_e$ i odczytujemy z tablic $\lambda_0$, a następnie podejmujemy decyzję odnośnie $H_0$.

# Literatura

Strahl D., Sobczak E., Markowska M., Bal-Domańska B. *Modelowanie ekonometryczne z Excelem*. Wrocław: AE 2002.

*Ekonometria. Metody, przykłady, zadania*. Red. J. Dziechciarz. Wrocław: AE 2002.

Welfe A. *Ekonometria*. Warszawa: PWE 2003.