TEORIA

ZESTAW II

Dana jest zależność:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + ... \alpha_k X_k + \varepsilon \text{ ,}$$

Estymujemy parametry takiego modelu. Otrzymujemy

$$Y = a_0 + a_1 X_1 + a_2 X_2 + ... a_k X_k + \varepsilon \text{ ,} \qquad \textbf{(*)}$$

- **WSPÓŁCZYNNIK KORELACJI LINIOWEJ PEARSONA**

Miara statystyczna objaśniająca poziom oraz kierunek zależności liniowej pomiędzy zmiennymi – Czy kolejne realizacje dwóch zmiennych (kolejne obserwacje) odchylają się w stosunku do swoich średnich w tą samą stronę (znak +), w przeciwne (znak -), czy niezależnie (0).

$$r_{XY} = \sum_{i=1}^{N} \frac{(x_i - \bar{x}) \cdot (y_i - \bar{y})}{S_X \cdot S_Y}$$

Uwaga: Przy budowie modelu ekonometrycznego można dobierać zmienne wg wartości wsp. korelacji pomiędzy zmienną X oraz Y. Wskaźnik $r_{XY} = 0$ sugeruje zupełny brak związku liniowego pomiędzy zmiennymi, wartości bliskie –1 lub 1 wskazują natomiast na możliwość opisania zmienności Y prawie w 100%.

Uwaga 2. Model jest *koincydentny*, jeśli **sgn r_i = sgn a_i**

- **PARA KORELACYJNA, REGULARNA PARA KORELACYJNA**

$$R_0 = \begin{bmatrix} r_1 \\ ... \\ r_k \end{bmatrix}, \qquad R = \begin{bmatrix} 1 & r_{12} & ... & r_{1k} \\ r_{21} & 1 & ... & r_{2k} \\ ... & ... & ... & .. \\ r_{k1} & r_{k2} & ... & 1 \end{bmatrix}.$$

r_{ji}- współczynnik korelacji pomiędzy X_i i X_j
r_i - współczynnik korelacji pomiędzy X_i i Y

(R,R_0)- **para korelacyjna**

Jeżeli spełniony jest warunek: $0<r_1<r_2<…<r_k$,
(R,R_0)- **regularna para korelacyjna**,

UWAGA- jeżeli współczynnik $r_i<0$, to należy zmienną X_i w modelu podmienić przez zmienną $X_i' = -X_i$.

- **WARIANCJA SKŁADNIKA LOSOWEGO, BŁĘDY SZACUNKU PARAMETRÓW**

Ponieważ parametry modelu estymowane są jedynie na podstawie próby, a nie całej populacji, oraz dane wejściowe są zmiennymi losowymi (zauważ, że obserwowany *y* składa się m.in. ze **składnika losowego** ε, a więc zależy od zm. losowej, a więc *y* jest zm. losową), to parametry również są zmiennymi losowymi.

Nieobciążony i zgodny estymator wariancji σ^2 składnika losowego w modelu (*) szacowanym MNK jest:

$$S^2 = \frac{e^T e}{n-(k+1)} = \frac{(y - Xa)^T (y - Xa)}{n-(k+1)} = \frac{y^T y - a^T X^T y}{n-(k+1)},$$

Nieobciążony i zgodny estymator macierzy kowariancji estymatora **a**:

$$\hat{D}^2(a) = [d_{ij}] = S^2 (X^T X)^{-1}.$$

Obliczając wartości elementów diagonalnych d_{jj} macierzy $D^2(a)$ otrzymujemy oceny wariancji estymatorów poszczególnych parametrów modelu (*).

$S_{a_j} = \sqrt{d_{jj}}$ - **średni błąd szacunku** parametru α_j.

$\left| \frac{S_{a_j}}{a_j} \right| \cdot 100\%$ - **średni względny błąd szacunku** parametru α_j.

- **WSPÓŁCZYNNIK DETERMINACJI R^2 ($R^2 \in [0,1]$)**

Po zbudowaniu modelu współczynnik determinacji pozwala zmierzyć, w jakim stopniu model umożliwia objaśnienie zmienności zmiennej *Y.*

Współczynnik determinacji zwykły

- W modelu musi występować wyraz wolny
- Interpretacja jest poprawna pod warunkiem, że badane związki są liniowe.
- R^2 przyjmuje wartości z przedziału (0, 1)

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = 1 - \frac{(\mathbf{y}-\hat{\mathbf{y}})^T(\mathbf{y}-\hat{\mathbf{y}})}{(\mathbf{y}-\mathbf{1}\cdot\bar{y})^T(\mathbf{y}-\mathbf{1}\cdot\bar{y})} = 1 - \frac{\mathbf{e}^T\mathbf{e}}{\mathbf{y}^T\mathbf{y} - n\cdot\bar{y}^2} = 1 - \frac{\mathbf{y}^T\mathbf{y} - \mathbf{a}^T\mathbf{X}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y} - n\cdot\bar{y}^2}$$

Ponieważ $\varphi^2 = \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$ jest to stosunek sumy kwadratów odchyleń wartości obserwowanych y_t od teoretycznych $\hat{y}_t$ do sumy kwadratów odchyleń obserwowanych y_t od średniej $\bar{y}$, to φ^2 (**współczynnik indeterminacji**) podaje, jaki procent zmienności y nie jest objaśniony przez model, natomiast R^2 (czyli 1-φ^2) podaje, jaki % zmienności y jest objaśniane przez model.

Inny sposób liczenia takiego współczynnika determinacji:

$$R^2 = R_0^T R^{-1} R_0$$

Współczynnik determinacji skorygowany

Gdy liczba szacowanych parametrów jest niewiele mniejsza od liczby dostępnych obserwacji, to współczynnik determinacji może być zawyżony (zafałszowany względem rzeczywistej jego interpretacji) i wtedy stosujemy skorygowany współczynnik determinacji $\overline{R^2}$, który jest mniejszy od R^2,

$$\overline{R^2} \le R^2 : \quad \overline{R^2} = R^2 - \frac{k}{n-(k+1)}(1-R^2)$$

Współczynnik determinacji niescentrowany

Gdy szacowany model jest bez wyrazu wolnego, wtedy stosuje się **niescentrowny współczynnik determinacji** $R_N^2 \in [0,1]$:

$$R_N^2 = 1 - \frac{\mathbf{e}^T\mathbf{e}}{\mathbf{y}^T\mathbf{y}}$$

- **EFEKT KATALIZY**

Współczynnik determinacji jest miarą dopasowania modelu ekonometrycznego do danych empirycznych, lecz informacja, jaką niesie o modelu, może być fałszywa, jeśli w modelu występują zmienne, które nazywamy **katalizatorami.**

ZAŁOŻENIE: (R,R_0)- regularna para korelacyjna.
Zmienna X_j z pary zmiennych (X_i,X_j), i<j, jest katalizatorem jeżeli

$r_{ij} < 0$ lub $r_{ij} > \frac{r_i}{r_j}$

$\eta = R^2 - H$ **- natężenie efektu katalizy**,

Gdzie H jest integralną pojemnością informacyjną zestawu zmiennych objaśniających modelu.

$W_\eta = \frac{\eta}{R^2}\cdot 100\%$ **- względne natężenie efektu katalizy**

Zadanie 1

Dla 20 typów radioodbiorników oszacowano zależność ceny radioodbiornika od cech charakteryzujących jakość. Oszacowano model liniowy postaci:

$\hat{y}_t = -1636 + 314x_t + 449z_t + 364w_t$,

w którym:

y_t - cena radioodbiornika typu t, t = 1,2,…,20

x_t - liczba zakresów fal, $x_t \in \{2,3,4\}$

z_t - liczba lamp, $z_t \in \{3,4,5,6\}$

w_t - rodzaj skrzynki, $w_t \in \{0,1\}$, w_t= 1 gdy skrzynka drewniana, w_t= 0 gdy skrzynka z tworzywa sztucznego.

Podać interpretację parametrów

Zadanie 2

Dla pewnego zakładu oszacowano model

$\hat{y}_t = 1{,}65 + 0{,}45x_t + 0{,}75z_t - 0{,}38p_t, t = 1,2,...,348,$

w którym:

x_t= 0 gdy staż pracy krótszy niż 3 lata,

x_t= 1 gdy staż pracy 3 lata lub dłuższy,

z_t=0 wykształcenie podstawowe,

z_t=1 wykształcenie średnie,

p_t=0 mężczyzna,

p_t=1 kobieta,

Porównać zarobki w 3 dowolnych klasach pracowników, jeśli klasą nazywamy grupę pracowników o jednakowym stażu, wykształceniu i płci. Na przykład w klasie (0,0,0) są mężczyźni o stażu krótszym niż 3 lata i wykształceniu podstawowym.

Zadanie 3

Czy podana para jest parą korelacyjną?Czy jest regularną parą korelacyjną?

$$R_0 = \begin{bmatrix} -0{,}01 \\ 0{,}9 \\ 0{,}68 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & -0{,}09 & 0{,}47 \\ -0{,}09 & 1 & -0{,}65 \\ 0{,}47 & -0{,}65 & 1 \end{bmatrix}.$$

Zadanie 4

Podczas szacowania MNK parametrów modelu ekonometrycznego postaci

$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t$, otrzymano następujące wyniki obliczeń:

$$(X^T X)^{-1} = \begin{bmatrix} 0{,}125 & 0 & 0 \\ 0 & 0{,}6 & 0{,}2 \\ 0 & 0{,}2 & 0{,}4 \end{bmatrix}, \quad X^T y = \begin{bmatrix} 12 \\ 4 \\ 5 \end{bmatrix}, \quad y^T y = 60.$$

Podać oszacowania parametrów tego modelu. Wyznaczyć średnie błędy szacunku i średnie względne błędy szacunku. Obliczyć i zinterpretować wartość współczynnika determinacji.

Zadanie 5

Obserwowano kształtowanie się przyrostów popytu na masło (Δy_t), przyrostów cen margaryny (Δx_{1t}) i przyrostów cen masła (Δx_{2t}) w 10 kolejnych okresach. Zebrano następujące dane:

t	1	2	3	4	5	6	7	8	9	10
Δy_t	1	2	3	0	-5	-4	-4	0	1	2
Δx_{1t}	0	0	0	0	0	0	0	1	1	1
Δx_{2t}	-1	-1	-1	0	1	1	1	1	0	0

Założono, że popyt na masło zależy w sposób liniowy od cen masła i cen margaryny. Zbudować liniowy model ekonometryczny objaśniający kształtowanie się zmian popytu na masło w zależności od zmian cen masła i margaryny. Przeprowadzić dyskusję dotyczącą uwzględnienia wyrazu wolnego w konstruowanym modelu. Oszacować parametry wyspecyfikowanego modelu i zinterpretować otrzymane wyniki. Ocenić dopasowanie modelu do danych empirycznych. Czy model jest koincydentny?

Zadanie 6

Dla pewnego liniowego modelu z wyrazem wolnym (n=8) dane są:

$$R = \begin{bmatrix} 1 & 0,5 \\ 0,5 & 1 \end{bmatrix}, R_0 = \begin{bmatrix} 0,4 \\ 0,9 \end{bmatrix},$$

$$\sum_t (y_t - \hat{y}_t)^2 = 0,56, \sum_t (y_t - \overline{y})^2 = 3,$$

Czy dane te nie są sprzeczne??