

PORÓWNANIE PRODUKCJI ENERGII ELEKTRYCZNEJ W LATACH 1995-2002 W WYBRANYCH WOJEWÓDZTWACH

Streszczenie

W pracy badano dynamikę produkcji energii elektrycznej w latach 1995-2002 ze źródeł ciepłych i wodnych. Wykorzystano metodę Gujarati'ego [1970] porównania prostych regresji opisujących dynamikę w wybranych województwach. Zwrócono też uwagę na sprawdzenie założeń wymaganych do zastosowania tej metody.

Słowa kluczowe: porównanie równań regresji, testowanie hipotez

Wstęp

W pracach dotyczących inżynierii rolniczej niektóre związki pomiędzy cechami badanych jednostek eksperymentalnych opisuje się równaniami regresji. Przykłady można znaleźć w licznych publikacjach. Mieszkalski [1999] opisał prostymi regresji niektóre cechy fizyczne nasion łubinu (średnią masę pojedynczego nasiona, masę właściwą, gęstość usypową i in.) w zależności od wilgotności dla trzech odmian łubinu. Również Mazur i Grochowicz [2000] oszacowali dynamikę wilgotności nasion bobiku, fasoli, grochu, łubinu (3 odmiany) i soi w zależności od temperatur w jakich zachodziła absorpcja wody oraz czasów absorpcji wody równaniami regresji wielokrotnej.

Eksperymentatora może interesować odpowiedź na pytanie, czy dana cecha zmienia się jednakowo, czy też nie dla badanych grup jednostek (np. odmian, frakcji). Odpowiednia metoda statystyczna może być do tego celu zastosowana. Prezentowany tu sposób porównania zależności regresyjnych opiera się na obliczeniach wykonanych przy wykorzystaniu pakietu statystycznego i w odróżnieniu od podobnych metod nie wymaga znajomości rachunku macierzowego ani przeprowadzania żmudnych obliczeń.

Cel i zakres pracy

Celem pracy było przeprowadzenie analizy zmian produkcji energii elektrycznej ze źródeł wodnych i źródeł ciepłych poprzez oszacowanie zależności regresyjnych opisujących zmiany w produkcji energii elektrycznej w ciągu 8 lat. Następnie uwzględniając odpowiednie założenia porównano proste regresji odpowiadające wybranym województwom.

Dane dotyczące produkcji energii elektrycznej pochodzą z bazy udostępnionej przez Główny Urząd Statystyczny (www.stat.gov.pl) i obejmują kolejne lata 1995-2002.

Opis metody

Założmy, że pewną zależność między dwiema cechami Y (opisywaną) oraz t (opisującą) dla l grup można wyrazić przy pomocy l prostych regresji. Interesuje nas sprawdzenie czy badane zjawisko przebiega z jednakową dynamiką w badanych grupach, czy też nie. Odpowiada to statystycznemu sprawdzeniu przy pomocy odpowiedniego testu wzajemnego położenia prostych – tj. zbadaniu czy proste te pokrywają się, są równoległe, czy też przebiegają w zupełnie różny sposób.

Opisywana metoda oprócz założenia dotyczącego normalności rozkładu wymaga założenia o równości wariancji składników losowych dla badanych grup. W przypadku, gdy w grupach mamy jednakową (n) ilość obserwacji można to sprawdzić przy pomocy testu Hartley'a [Seber 1977] formułując hipotezę

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_l^2 \quad (1)$$

przeciwko hipotezie

$$H_1 : \text{nie wszystkie wariancje s\aa r\o wne}$$

i obliczaj\aa c warto\u015b\u0107 statystyki testowej

$$H = \frac{\max(S_i^2)}{\min(S_i^2)}, \quad (2)$$

gdzie S_i^2 , $i=1, \dots, l$ s\aa estymatorami wariancji b\u0142\u0119d\o w (\u015brednimi kwadratami dla b\u0142\u0119d\o w), kt\o re mo\u017ana odczyta\u0107 z tabel analizy wariancji dla omawianych l regresji.

Warto\u015b\u0107 krytyczna $H_{\alpha;l;n-1}$ uzyskuje si\u0119 z tabel kwantyli statystyki Hartley'a [Krysicki i in. 1999].

W przypadku, gdy nie odrzucimy hipotezy zerowej (1) mo\u017cemy przyja\u0107, \u017ce wariancje s\aa r\o wne i prowadzi\u0107 dalsz\aa analiz\u0119 w kolejnych etapach:

1. Budowa modelu pe\u0142nego.

Definiujemy $l-1$ zmiennych zero-jedynkowych (o jedn\aa mniej ni\u017c liczba por\o wnywanych prostych) jako

$$Z_i = \begin{cases} 1 & \text{dla obserwacji dotycz\aa cych (i+1)-ej grupy} \\ 0 & \text{dla obserwacji z pozosta\u0142ych grup,} \end{cases}$$

gdzie $i=1, \dots, l-1$.

Nast\u0119pnie budujemy model regresji [Gujarati 1970]

$$Y_p = \alpha_1 + \beta_1 t + Z_1(\alpha_2 + \beta_2 t) + \dots + Z_{l-1}(\alpha_l + \beta_l t) + \varepsilon, \quad (3)$$

przy czym ε jest b\u0142\u0119dem losowym natomiast α_i, β_i , ($i=1, 2, \dots, l$) s\aa parametrami.

Powy\u017cszy model b\u0119dziemy nazywa\u0107 pe\u0142nym (st\aa d oznaczenie Y_p).

2. Sformu\u0142owanie hipotez.

Hipoteza zerowa w postaci

$$H_0 : \alpha_2 = \beta_2 = \alpha_3 = \beta_3 = \dots = \alpha_l = \beta_l = 0 \quad (4)$$

oznacza, \u017ce wszystkie proste da si\u0119 opisa\u0107 wsp\o lnym r\o wnanie\u0107 (lub tym samym proste nie r\o \u017ani\aa si\u0119 istotnie) przeciwko hipotezie alternatywnej

$$H_1 : \text{nie wszystkie } \alpha_i \text{ oraz } \beta_i \text{ s\aa jednocze\u015bnie zerami (} i=2, \dots, l \text{)}$$

m\o wi\aa c, \u017ce wsp\o lczynniki prostych nie s\aa jednakowe.

3. Analiza modelu przy prawdziwo\u015b\u0107ci hipotezy zerowej.

Model przy prawdziwo\u015b\u0107ci hipotezy zerowej (4) jest nast\u0119puj\aa cy:

$$Y_{H_0} = \alpha_0 + \beta_0 t + \varepsilon. \quad (5)$$

4. Obliczenie warto\u015b\u0107ci statystyki testowej i sformu\u0142owanie wniosk\o w.

Hipotez\u0119 H_0 testujemy przy pomocy statystyki [Draper, Smith 1998]

$$F = \frac{SSR(P) - SSR(H_0)}{k - q} : MSE(P), \quad (6)$$

gdzie wartości $SSR(P)$ i $SSR(H_0)$ oznaczają regresyjne sumy kwadratów odpowiednio $SSR(P)$ dla modelu pełnego - $SSR(H_0)$ dla modelu (5) przy założeniu hipotezy zerowej (4), zaś $MSE(P)$ oznacza średni kwadrat dla błędu w modelu pełnym (3). Liczby k oraz q są stopniami swobody odpowiadającymi $SSR(P)$ i $SSR(H_0)$. Wszystkie te wartości można otrzymać z tabel analizy wariancji dla regresji przeprowadzając estymację współczynników dla modeli (3) i (5) metodą najmniejszych kwadratów w pakiecie statystycznym.

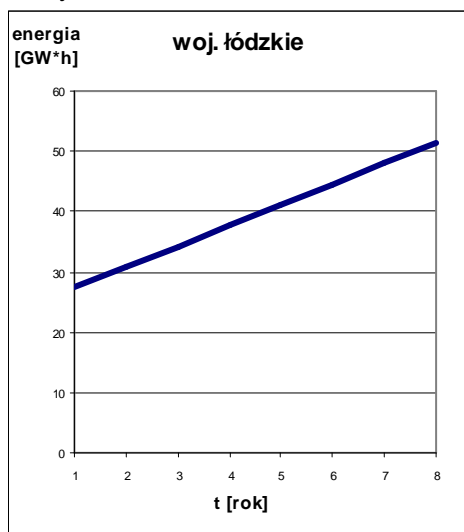
Wartość krytyczną $F_{\alpha; k-q; e}$ odczytujemy z tablic F Snedecor'a, przy czym α jest poziomem istotności natomiast e jest ilością stopni swobody odpowiadających $MSE(P)$. Wnioski wyciągamy jak w każdym teście F.

Wyniki badań

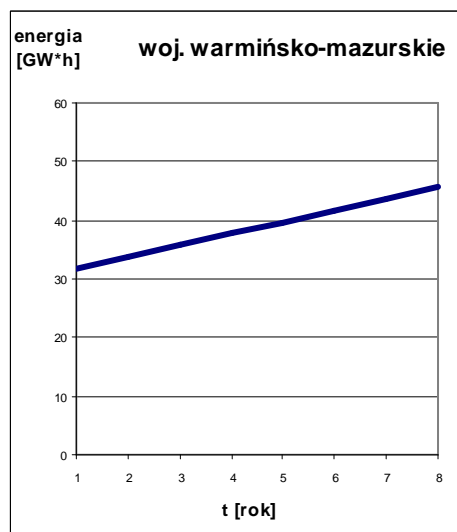
Do obliczeń wykorzystano program Statistica 6.0.

1. Produkcja energii elektrycznej ze źródeł wodnych.

Wstępna analiza danych nie pozwoliła na dopasowanie uniwersalnego modelu regresji (prostej, paraboli) wspólnego dla wszystkich województw. Okazało się jednak, że dynamikę wielkości produkcji energii elektrycznej (w GW*h) ze źródeł wodnych w ciągu rozpatrywanych 8 lat można dobrze opisać prostymi regresji dla województw łódzkiego i warmińsko-mazurskiego – Rys 1a i 1b. Latom 1995-2002 odpowiadają liczby od 1 do 8 na osi poziomej.



Rys. 1a) $\hat{y} = 3,41t + 24,06$



Rys. 1b) $\hat{y} = 1,96t + 29,89$

Rys. 1. Oszacowane proste regresji opisujące produkcję energii elektrycznej ze źródeł wodnych w kolejnych latach

Fig. 1. Estimated regression lines describing electric energy generation from the water sources in successive years

Sprawdzając założenie dotyczące równości wariancji tj. hipotezę (1) dla $l = 2$ otrzymujemy estymatory wariancji $S_1^2 = 56,32$ oraz $S_2^2 = 25,89$, a wartość statystyki (2) wynosi $H_{obl} = 2,18$. Wartość H_{obl} jest mniejsza od wartości krytycznej $H_{0,05;2;7} = 4,99$ nie ma więc podstaw do odrzucenia hipotezy (1) o równości wariancji. Możemy przyjąć, że wariancje są równe i przystąpić do porównania współczynników wcześniej opisaną metodą.

Tworzymy model pełny $Y_p = \alpha_1 + \beta_1 t + Z_1(\alpha_2 + \beta_2 t) + \varepsilon$, gdzie Y_p oznacza ilość energii, t jest zmienną oznaczającą czas oraz

$$Z_1 = \begin{cases} 1 & \text{dla obserwacji dotyczących woj. warmińsko-mazurskiego} \\ 0 & \text{dla obserwacji dotyczących woj. łódzkiego.} \end{cases}$$

Hipoteza zerowa $H_0 : \alpha_2 = \beta_2 = 0$ pozwoli na sprawdzenie czy proste są jednakowe – tzn. czy dynamika produkcji energii elektrycznej w ciągu ośmiu lat dla obu województw jest taka sama. Wartość statystyki testowej (6) $F_{obl} = 0,56$ jest mniejsza od wartości krytycznej $F_{0,05;2;12} = 3,89$. Nie ma podstaw do odrzucenia hipotezy zerowej. Możemy stwierdzić, że proste oszacowane dla województw łódzkiego i warmińsko-mazurskiego nie różnią się istotnie.

2. Produkcja energii elektrycznej ze źródeł ciepłych.

Podobnie jak w przypadku źródeł wodnych wstępna analiza danych nie wskazuje na możliwość doboru modelu dobrze opisującego zmiany produkcji energii elektrycznej ze źródeł ciepłych w zależności od czasu we wszystkich województwach. Oszacowano proste regresji oraz estymatory wariancji błędów dla następujących trzech województw:

$$\text{woj. opolskie} \quad \hat{y} = 563,64t + 4966,86 \quad S_1^2 = 660191,$$

$$\text{woj. lubuskie} \quad \hat{y} = 55,51t + 432,82 \quad S_2^2 = 7454,$$

$$\text{woj. warmińsko-mazurskie} \quad \hat{y} = 7,11t + 199,39 \quad S_3^2 = 137,07.$$

Wartość statystyki (2) $H_{obl} = 4816,45$ jest większa od wartości krytycznej $H_{0,05;3;7} = 6,94$.

Odrzucamy zatem hipotezę zerową (1) o równości trzech wariancji. Założenie nie jest spełnione - nie możemy dokonać porównania prostych opisaną metodą.

Wnioski

- Dla województw łódzkiego i warmińsko-mazurskiego dynamika produkcji energii elektrycznej ze źródeł wodnych w badanych latach nie różni się istotnie.
- Nie można porównać przebiegu produkcji energii elektrycznej ze źródeł ciepłych w województwach opolskim, lubuskim i warmińsko-mazurskim.

Uwagi

- Przedstawiona metoda pozwala na porównywanie nie tylko prostych, ale również wielomianów lub powierzchni regresji.
- W przypadku różnych ilości obserwacji w badanych grupach do sprawdzenia założenia o jednakowości wariancji składników losowych zamiast testu Hartley'a należy użyć innych testów np. Bartlett'a [Seber 1977].
- W analogiczny sposób można testować inne hipotezy zerowe niż (4) np.: $H_0 : \alpha_2 = \alpha_3 = \dots = \alpha_l = 0$ (wszystkie proste przecinają oś pionową w tym samym punkcie) lub $H_0 : \beta_2 = \beta_3 = \dots = \beta_l = 0$ (wszystkie proste są równoległe).

Bibliografia

Draper N.R., Smith H., 1998, Applied Regression Analysis, Wiley Series in Probability and Statistics.

Gujarati D., 1970, Use of Dummy Variables in Testing for Equality Between Sets of Coefficients in Linear Regressions: A Generalization, The American Statistician: p. 18-22.

Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M., 1999, Rachunek Prawdopodobieństwa i Statystyka Matematyczna w Zadaniach, cz. II – Statystyka Matematyczna, PWN, str. 297.

Mazur J., Grochowicz J., 2000, Wpływ temperatury na zdolność absorpcji wody przez nasiona roślin strączkowych zanurzonych całkowicie w wodzie, Inżynieria Rolnicza 5(16), str. 163-170.

Mieszkalski L., 1999, Badania podstawowych właściwości fizycznych nasion łubinów, Problemy Inżynierii Rolniczej Nr 1/99, str. 51-58.

Seber G. A. F., 1977, Linear Regression Analysis, John Wiley & Sons

COMPARISON OF ELECTRIC ENERGY GENERATION WITHIN 1995-2002 IN SELECTED PROVINCES

Summary

The study dealt with the dynamics of electric energy generation within 1995-2003 from the thermal and water sources. The method by Gujarati (1970) was applied to compare the regression lines describing the dynamics in selected provinces of Poland. An attention was paid to check of the assumptions required to application of mentioned method.

Key words: electric energy generation, comparison of regression equations, testing of hypothesis.

Recenzent – Janusz Boss