

Ekonometria dla II E oraz MSEM at — Kolokwium

29-01-2021

Rozwiązanie należy przesłać do 14:55 na skrzynkę rafal.wozniak@uw.edu.pl w formie skanu lub zdjęcia ręcznie napisanych kodów na kartce.

Zadanie 0

Proszę odręcznie napisać zdanie: "Daję słowo honoru, że nie będę ściągąć."

Zadanie 1 (40%)

Podczas 9-miesięcznego badania w latach 2008-09 zebrano dane o poziomie płacy 397 nauczycieli akademickich z amerykańskich college'y podzielonych na trzy rozłączne grupy: adiunktów, wykładowców i profesorów. W modelu logarytm płacy ($\ln placa$) był tłumaczony zmiennymi zerojedynekowymi: dla tytułu profesora (*profesor*), dla habilitacji (*habilitacja*), zmienna określająca czy dana osoba pracuje na wydziale nauk stosowanych (*stosowana*), zmiennej dla płci (*kobieta*) oraz ciągłej liczba lat pracy (*lata_pracy*). Dodatkowo w modelu użyto interakcji nauki stosowanej z płcią (*stos*kobieta*) oraz interakcji lat pracy z nauką stosowaną (*stos*lata*).

Call:

```
lm(formula = I(log(salary)) ~ professor + habilitacja + applied +
    yrs.service + appservice + as.factor(sex) + femapp, data = dane)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	11.159210	0.046650	239.212	< 2e-16 ***
profesor	0.465250	0.030878	15.067	< 2e-16 ***
habilitacja	0.157287	0.032794	4.796	2.31e-06 ***
stosowana	0.069865	0.033105	2.110	0.0355 *
lata_pracy	-0.002632	0.001095	-2.403	0.0167 *
stos*lata	0.002599	0.001448	1.795	0.0735 .
kobieta	-0.102013	0.045992	2.218	0.0271 *
stos*kobieta	0.098876	0.062100	1.592	0.1121

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1805 on 389 degrees of freedom

Multiple R-squared: 0.5273, Adjusted R-squared: 0.5188

F-statistic: 61.99 on 7 and 389 DF, p-value: < 2.2e-16

Ramsey RESET F(3, 386) = 2.87	Prob > F = 0.0365
White's test chi2(21) = 77.60	Prob > chi2 = 0.0000
Jarque-Berra chi2(2) = 8.27	Prob > chi2 = 0.0160
Breusch-Godfrey chi2(1) = 2.712	Prob > chi2 = 0.0996
Chow F(8, 389) = 0.185	Prob > F = 0.993

Hipotezy weryfikuj na poziomie istotności $\alpha = 10\%$.

- (4%) Czy zmienne w modelu są łącznie istotne?
- (4%) Zinterpretuj wartość współczynnika determinacji.
- (4%) Które zmienne są statystycznie istotne? Uzasadnij.
- (4%) Czy istnieje luka płacowa między kobietami i mężczyznami? Jeśli tak, to ile wynosi?
- (4%) Zinterpretować parametr dla zmiennej *profesor*.
- (5%) Określić wpływ lat pracy na płacę.
- (10%) Czy model spełnia założenia KMRL? Określić które spełnia/niespełnia.
- (5%) Jakie są konsekwencje spełnienia/niespełnienia założeń KMRL dla wnioskowania z tego modelu?

Zadanie 2 (40%)

Oszacowano modele wydatków gospodarstw domowych na transport (zmienna zlogarytmowana). Zmiennymi objaśnianymi były: $\ln CENYTR$ - logarytm cen usług transportowych, $w11$ - wydatki na hotele i restauracje (niezlogarytmowana), $\ln DOCHOD$ - logarytm dochodu gosp. dom. oraz zmienne zerojedynkowe określające klasę miejscowości w której mieści się gosp. dom. $klm_ \#$ (1 dla miejscowości powyżej 500 tys. mieszkańców; 2 dla miejscowości [200 tys., 500 tys.); 3 dla miejscowości [100 tys., 200 tys.); 4 dla miejscowości [20 tys., 100 tys.); 5 dla miast do 20 tys. mieszk.; 6 dla wsi).

	model_1	model_2	model_3	model_4	model_5
$\ln CENYTR$	2.935	2.865	1.614		
$w11$	-0.000				0.000
$\ln DOCHOD$	0.679**	0.676***	0.733***	0.730***	
$_{Iklm_2}$	0.226	0.228			
$_{Iklm_3}$	-0.025	-0.038			
$_{Iklm_4}$	-0.093	-0.089			
$_{Iklm_5}$	0.171	0.172			
$_{Iklm_6}$	0.062	0.064			
$_{cons}$	-0.075	-0.056	-0.415	-0.372	5.470***
N	57	57	57	57	57
F	2.485	2.896	9.975	19.885	0.007
R ²	0.293	0.293	0.270	0.266	0.0001
RSS	20.821	20.828	21.501	21.626	29.441

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Przyjmij 5% poziom istotności. $F_{0.05}(5, 49) = 2.404$, $F_{0.05}(5, 50) = 2.400$

- (10%) Przetestować hipotezę, że klasa miejscowości nie wpływa na wydatki na transport.
- (10%) Wyznaczyć i skomentować wartość skorygowanego R^2 dla modelu (5).
- (10%) Wyznaczyć dla jakich wartości N , K oraz R^2 skorygowany współczynnik determinacji liniowej będzie przyjmował wartości ujemne.
- (10%) Zaproponować postać funkcyjną modelu dla wydatków na transport, która będzie uwzględniała jednocześnie następujące warunki:
 - zależność od odległości między miejscowością pracy a zameldowania ma przyjmować maksimum dla 75 km;
 - mają zależeć od posiadania samochodu;
 - mają zależeć od średniej cen paliw, ale tylko dla osób posiadających samochód;
 - wpływ dodatkowego dziecka ma być równy połowie wpływu dodatkowego dorosłego.

Zadanie 3 (20%)

Ekonometryk miał za zadanie przeanalizować zależność między popytem na panele słoneczne (The term solar panel is used colloquially for a photo-voltaic (PV) module.¹). Zdecydował, że zmienną objaśnianą będzie ich cena w dolarach za 1 wat generowanej energii. Zmienną objaśniającą był popyt na panele słoneczne wyrażony w MW energii generowanej z tych paneli. Ekonometryk rozważał kilka form funkcyjnych.

$$cena = 127.1 - 0.0005popyt + e \quad (\text{Wykres 1})$$

$$cena = 336.55 - 73.11 \ln(popyt) + e \quad (\text{Wykres 2})$$

$$\ln(cena) = 2.88 - 0.36 \ln(popyt) + e \quad (\text{Wykres 3})$$

$$cena = 14.16 + 1102.5 \frac{1}{\sqrt{x}} + e \quad (\text{Wykres 4})$$

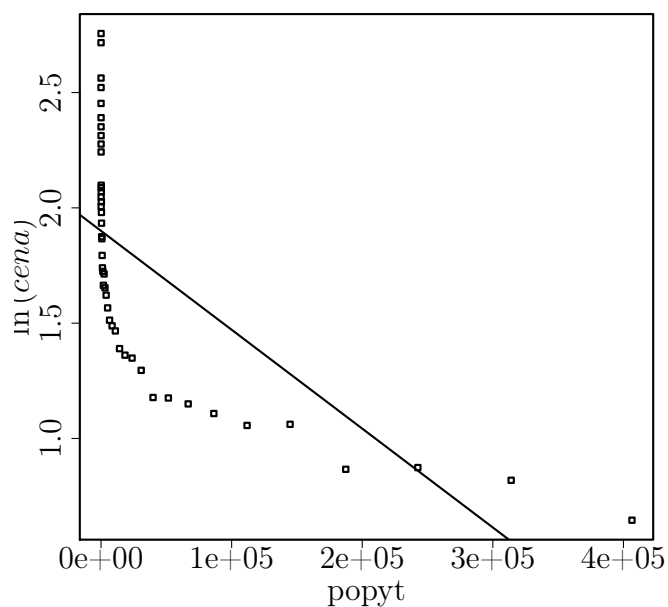
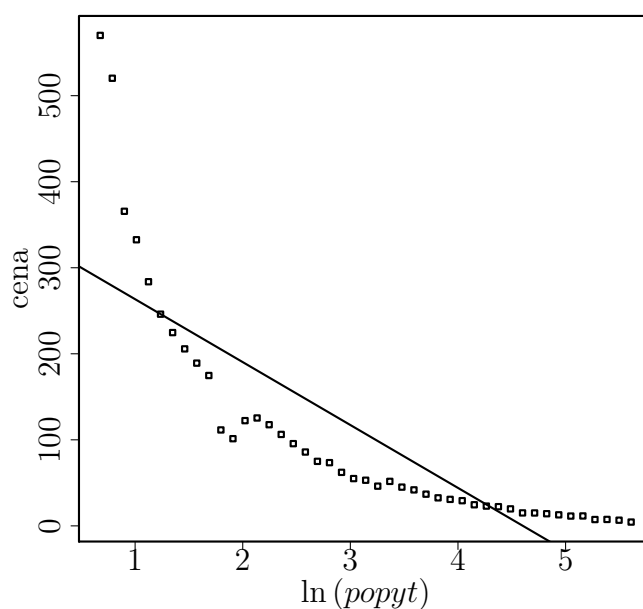
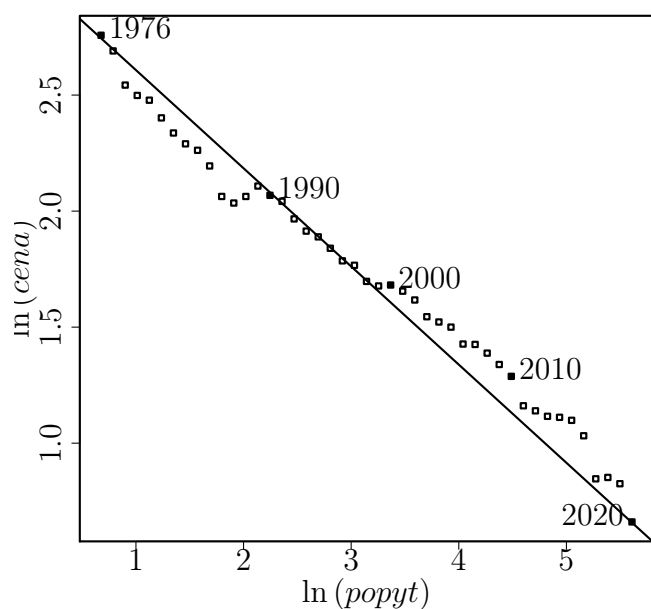
Tablica poniżej przedstawia statystyki dla tych modeli regresji. Wykresy danych i regresji przedstawione są na kolejnej stronie.

Tabela 1: Statystyki modeli regresji z wykresów (1)-(4)

	Wykres 1	Wykres 2	Wykres 3	Wykres 4
R^2	11.72%	67.24%	98.57%	97.67%
$\bar{R}^2$	9.67%	66.48%	98.54%	97.62%
σ_e	124.8	76.02	0.065	20.26

- (4%) Który model należy uznać za najlepszy? Uzasadnij.
- (3%) Który z modeli jest najlepszy do interpretacji? Uzasadnij.
- (3%) Jak zinterpretować wpływ dodatkowej jednostki popytu na cenę w modelu z wykresu 3?
- (7%) Jak zinterpretować wpływ dodatkowej jednostki popytu na cenę w modelu z wykresu 4? (dla popytu na poziomie 100000)
- (3%) Czy spełnione będzie założenie o nielosowości zmiennej objaśniającej? Uzasadnij.

¹https://en.wikipedia.org/wiki/Solar_panel

Wykres 1. $cena = 127.1 - 0.0005popyt + e$ Wykres 2. $cena = 336.55 - 73.11 \ln(popyt) + e$ Wykres 3. $\ln(cena) = 2.88 - 0.36 \ln(popyt) + e$ Wykres 4. $cena = 14.16 + 1102.5 \frac{1}{\sqrt{x}} + e$ 