

Egzamin z Ekonometrii dla IiE i MSEMAt - 03.02.2020

Regulamin:

1. Przed przystąpieniem do pisania egzaminu należy podpisać wszystkie kartki arkusza egzaminu (na górze w przewidzianym miejscu). Złożenie podpisu pod regulaminem oznacza jego akceptację. Do egzaminu mogą przystąpić osoby, które akceptują regulamin.
2. W razie braku podpisu, kartka nie zostanie oceniona. Nie będą też oceniane rozwiązania wpisane na kartkach innych, niż te rozdawane przez prowadzących.
3. Egzamin trwa 90 minut.
4. Rozwiązywanie zadań należy rozpocząć po ogłoszeniu początku egzaminu zakończyć wraz z ogłoszeniem końca egzaminu. Złamanie tej zasady skutkuje usunięciem z sali i unieważnieniem egzaminu.
5. W trakcie egzaminu wolno używać jedynie długopisu o innym kolorze atramentu niż czerwony oraz kalkulatora.
6. Rozwiązanie każdego zadania należy zapisać na kartce z tymże zadaniem, ewentualnie na czystych kartkach znajdujących się na początku egzaminu lub na dodatkowych kartkach uzyskanych od prowadzących egzamin.
7. Na jednej kartce może znajdować się rozwiązanie tylko jednego zadania. Oceniane jest rozwiązanie tylko tego zadania, którego numer widnieje na kartce.
8. Egzamin składa się z 4 pytań teoretycznych i **3** zadań.
9. Posiadanie przy sobie wszelkich materiałów drukowanych (w tym książek) oraz innych np. wykonanych własnoręcznie materiałów zostanie uznane za ściąganie.
10. Rozmowy z innymi zdającymi będą traktowane identycznie jak ściąganie.
11. Każda zauważona próba ściągania skutkuje usunięciem z egzaminu.
12. Wszystkie pytania należy kierować bezpośrednio do osób pilnujących.

Regulamin zrozumiałam(em)

Warszawa, 03-02-2020,

*Podpis***Pytania teoretyczne****Pytanie 1** Podać treść twierdzenia Gaussa-Markowa i wyjaśnić jego znaczenie.**Pytanie 2** Jakich statystyk używamy do wykrywania obserwacji nietypowych i błędnych?**Pytanie 3** Wyjaśnić, dlaczego R^2 nie można używać do porównywania modeli.**Pytanie 4** Co to znaczy, że estymator MNK jest estymatorem zgodnym?

Zadanie 1 (45%)

Na podstawie danych o poszczególnych stanach oszacowano model dla logarytmu liczby ofiar śmiertelnych wypadków drogowych ($\ln TrffcFrt$). Zmienna *Jail* oznacza bezwzględny wyrok więzienia po złapaniu kierowcy w stanie nietrzeźwości w danym stanie, *ComServ* bezwzględny wyrok prac społecznych, natomiast *Jail*ComServ* jest interakcją zmiennych *Jail* i *ComServ*. Zmienna *BeerTax* to poziom opodatkowania alkoholu w danym stanie, *MinLegDrkAge* oznacza minimalny dopuszczalny wiek osoby, która może spożyć alkohol, *AvgMiles* to przeciętna liczba mil przejechanych przez kierowcę w danym stanie USA, *PerInc* oznacza przeciętny roczny dochód per capita w tysiącach dolarów.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.813479	0.260537	3.122	0.001954 **
BeerTax	0.060673	0.025977	2.336	0.020114 *
MinLegDrkAge	0.001127	0.012543	0.090	0.928455
Jail	0.098895	0.034220	2.890	0.004110 **
ComServ	0.190454	0.055350	3.441	0.000655 ***
AvgMiles	0.066662	0.007550	8.829	< 2e-16 ***
PerInc	-0.054997	0.005571	-9.872	< 2e-16 ***
Jail*ComServ	-0.125010	0.069022	-1.811	0.071028 .

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

Residual standard error: 0.1988 on 328 degrees of freedom

Multiple R-squared: 0.492, Adjusted R-squared: 0.4812

F-statistic: 45.39 on 7 and 328 DF, p-value: < 2.2e-16

Ramsey RESET F(12, 316) = 13.89 Prob > F = 0.0000

White's test chi2(29) = 250.93 Prob > chi2 = 0.0000

Jarque-Berra chi2(2) = 48.85 Prob > chi2 = 0.0000

Breusch-Godfrey chi2(1) = 153.86 Prob > chi2 = 0.0000

Chow F(8, 328) = 5.44 Prob > F = 0.0000

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

- (5%) Czy zmienne w modelu są łącznie istotne? Uzasadnij.
- (5%) Zinterpretuj wartość współczynnika determinacji.
- (5%) Jak podwyższenie minimalnego dopuszczalnego wieku spożywania alkoholu przełoży się na liczbę wypadków drogowych? Uzasadnij.
- (5%) Jak wzrost średniego dochodu per capita o 1000\$ wpłynie liczbę ofiar śmiertelnych wypadków? Uzasadnij.
- (5%) Określić wpływ obligatoryjnego więzienia na liczbę ofiar śmiertelnych wypadków. Uzasadnij.
- (10%) Czy model spełnia założenia KMRL? Określić które spełnia/nie spełnia. Uzasadnij.
- (5%) Jakie są konsekwencje spełnienia/niespełnienia założeń KMRL dla wnioskowania z tego modelu?
- (5%) Interpretując wyniki modelu odpowiedz i uzasadnij na czym powinna polegać efektywna polityka zmniejszania liczby ofiar śmiertelnych wypadków drogowych?

Rozwiązanie zadania 1

- a) Tak. Statystyka $F = 45.39$ i odpowiadające jej $p - value = 0$, więc odrzucamy H_0 o łącznej nieistotności zmiennych w modelu.
- b) $R^2 = 0.492$. Model wyjaśnia 49,2% zmienności logarytmu liczby ofiar śmiertelnych wypadków drogowych.
- c) Podwyższenie minimalnego dopuszczalnego wieku spożywania alkoholu nie przełoży się na zmienią objaśnianą, ponieważ $MinLegDrkAge$ jest nieistotna statystycznie ($p - value = 92.8\%$).
- d) Wzrost przeciętnego dochodu rocznego p.c. o 1000 dolarów związany jest ze zmniejszeniem liczby ofiar o około 5.5%.
- e) Interpretacja parametru przy obligatoryjnym więzieniu zależy od tego czy w stanie zasądzana jest kara prac społecznych.
- Dla stanów w których $ComServ = 0$: Obligatoryjne więzienie związane jest ze zwiększoną liczbą ofiar o około 9.9%.
 - Dla stanów w których $ComServ = 1$: Obligatoryjne więzienie związane jest ze zmniejszoną liczbą ofiar o około 2.6% lub $[\exp(0.098895 - 0.125010) - 1] * 100\%$.
- f) Założenia KMRL
- Nie jest spełnione założenie o liniowości formy funkcyjnej
 - test RESET: statystyka testowa $F(12, 316) = 13.89$ i $p - value = 0\%$, więc odrzucamy H_0 o liniowości formy funkcyjnej
 - test Chowa: statystyka testowa $F(8, 328) = 5.44$ i $p - value = 0\%$, więc odrzucamy H_0 o stabilności parametrów
 - Nie jest spełnione założenie o homoskedastyczności składnika losowego
 - test White'a: statystyka testowa $chi2(29) = 250.93$ i $p - value = 0\%$, więc odrzucamy H_0 o homoskedastyczności ε
 - Nie jest spełnione dodatkowe założenie o normalności rozkładu składnika losowego
 - test Jarque-Bera: statystyka testowa $chi2(2) = 48.85$ i $p - value = 0\%$, więc odrzucamy H_0 o normalności rozkładu ε
 - Gdyby dane miały wymiar czasowy, to powinniśmy jeszcze sprawdzić założenie o braku autokorelacji.
- g) Niespełnione założenie o liniowości formy funkcyjnej przekłada się na obciążenie estymatorów przez co uzyskane parametry nie przybliżają prawdziwych wartości. Model należy poprawić lub z niego nie korzystać.

Drugim problemem jest heteroskedastyczność ε , która powoduje obciążenie oszacowania macierzy wariancji kowariancji. Gdyby forma zależności była liniowa (pozytywne wyniki testów RESET i Chowa), to powinniśmy użyć macierzy odpornej lub SUMNK aby móc korzystać ze statystyk t i p -value.

Niespełnione dodatkowe założenie o normalności rozkładu ε w małych próbach uniemożliwia korzystanie ze statystyk t i F oraz odpowiadającym im p -value we wnioskowaniu. Nasza próba liczy ponad 300 obserwacji, więc niespełnienie tego założenia nie jest problematyczne.

h) Analizując oszacowania parametrów należałoby uznać, że efektywna polityka zmniejszania liczby ofiar śmiertelnych wypadków drogowych powinna polegać na:

- i. zwiększaniu dochodów mieszkańców
- ii. zmniejszeniu liczby przejeżdżanych kilometrów
- iii. zmniejszenia opodatkowania piwa
- iv. niekorzystania z narzędzi prawnych jakimi są obligatoryjne więzienie i prace społeczne.

Realnie analizując uzyskane wyniki należy zauważyć, że oszacowania nie są zgodne z intuicją. Należałoby się zastanowić nad dwiema kwestiami:

- (a) Występowanie endogeniczności zmiennych w modelu. Być może w stanach w których obserwowana jest duża liczba wypadków stosuje się bardziej dotkliwe kary. Innymi słowy, może występować relacja zwrotna między y a x -ami.
- (b) Oszacowanie modelu dla współczynnika ofiar wypadków względem np. liczby przejechanych kilometrów. Tam gdzie ruch jest większy powinniśmy obserwować średnio więcej wypadków.

Zadanie 2 (35%)

Oszacowano modele popytu na piwo gospodarstw domowych. Zmienną objaśnianą był logarytm ilości. Zmiennymi objaśniającymi były: $\ln PB$ - logarytm ceny piwa, $\ln PL$ - logarytm średniej ceny mocniejszych alkoholi, $\ln PR$ - logarytm miary cen pozostałych dóbr i usług, $\ln INC$ - logarytm dochodu, natomiast $\ln INC^2$ to kwadrat logarytmu dochodu. Parametry, których wartość odchylenia standardowego wynosi ($\cdot$), zostały nie oszacowane, ale zostało nałożone ograniczenie o danej wartości. $F_{0.05}(1, 20) = 4.351$, $F_{0.05}(1, 24) = 4.26$, $F_{0.05}(1, 25) = 4.24$, $F_{0.05}(1, 26) = 4.22$

	model_1	model_2	model_3	model_4	model_5	model_6
$\ln PB$	-1.020*** (0.24)	-1.088*** (0.26)	-1.045*** (0.19)	-1.000 (.)	-0.403 (0.27)	
$\ln PL$	-0.583 (0.56)	-0.708 (0.59)	-0.628 (0.49)	-0.604 (0.49)	0.553 (0.67)	-1.632* (0.65)
$\ln PR$	0.210* (0.08)	0.211* (0.08)	0.205* (0.07)	0.211* (0.08)	0.331** (0.10)	0.288** (0.10)
$\ln INC$	0.923* (0.42)	12.329 (15.52)	1.000 (.)	0.903* (0.34)	-1.000 (.)	-0.067 (0.44)
$\ln INC^2$		-0.541 (0.74)				
_cons	-3.243 (3.74)	-63.007 (81.35)	-3.919*** (0.85)	-3.016 (2.58)	13.609*** (1.16)	8.113* (3.39)
N	30	30	30	30	30	30
R ²	0.825	0.829	0.952	0.896	0.421	0.698
RSS	0.090	0.088	0.090	0.347	0.167	0.155

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

W nawiasach podane są wartości odchyłeń standardowych oszacowań parametrów.

- (5%) Zinterpretuj oszacowanie parametru przy cenie piwa w modelu 1.
- (5%) Zinterpretuj oszacowanie parametru przy dochodzie w modelu 1.
- (10%) Dla modelu (1) przetestować hipotezę, że popyt na piwo jest dokładnie proporcjonalny do dochodu.
- (5%) Czy cena piwa jest ortogonalna do ceny mocniejszych alkoholi i do dochodu? Uzasadnij.
- (10%) Dla modelu (1) wyznaczyć statystykę testową oraz p-value dla hipotezy, że parametr przy cenie piwa jest równy (-1). Tablica dystrybucyjności rozkładu t-Studenta z 25 stopniami swobody podana jest poniżej. **Podpowiedź:** Rozkład t-Studenta jest rozkładem symetrycznym.

x	0	0.1	0.2	0.3	0.4	0.5	1	1.5	2	3
F_t	0.5	0.54	0.58	0.62	0.65	0.69	0.84	0.93	0.97	0.99

Rozwiązanie zadania 2

- a) Przy 1% wzroście ceny piwa będziemy obserwowali spadek popytu na piwo o 1.02%.
- b) Przy 1% wzroście dochodu będziemy obserwowali wzrost popytu na piwo o 0.923%.
- c) Należy porównać modele (1) i (5). Weryfikujemy hipotezę $H_0 : \beta_{\ln PB} = -1$. Wówczas $g = 1$, $N = 30$, $K = 5$.

$$F_{test} = \frac{(S_R - S_U)/g}{S_U/(N - K)} = \frac{(0.347 - 0.090)/1}{0.090/(30 - 5)} = 0.11$$

Obszar krytyczny $K = (F_{1,25}, +\infty) = (4.24, +\infty)$. Statystyka testowa F_{test} nie wpada do obszaru krytycznego, więc nie możemy odrzucić H_0 .

- d) Cena piwa nie jest ortogonalna do ceny mocniejszych alkoholi i dochodu. Uzasadnienie ekonomiczne: ceny piwa i mocniejszych alkoholi powinny być skorelowane oraz alkohol jako dobro normalne. Uzasadnienie ekonometryczne: usuwając tę zmienną z modelu (1) otrzymujemy model (6), w którym oszacowania parametrów istotnie się zmieniają. Gdyby cena piwa była ortogonalna, to oszacowania przy pozostałych zmiennych nie powinny się bardzo zmienić.
- e) Testujemy $H_0 : \beta_{\ln PB} = -1$ przy pomocy statystyki t

$$t_{test} = \frac{\hat{\beta} - \beta_{\ln PB}}{Std.Err(\hat{\beta})} = \frac{-1.020 - (-1)}{0.24} = \frac{-0.02}{0.24} = -0.8(3)$$

Odpowiadające jej p-value

$$p - value = 2 [1 - F_t(|t_{test}|)] \approx 2(1 - 0.54) = 0.92$$

P-value równe 92% interpretujemy jako brak podstaw do odrzucenia H_0 .

Zadanie 3 (20%)

Ekonometryk rozważał model regresji dla ilości produkowanej pszenicy (w kg) w farmach w Kanadzie w zależności od liczby roboczogodzin pracowników

$$pszenica_i = \beta_1 roboczogodziny_i + \varepsilon_i. \quad (1)$$

- a) (10%) Uzasadnić brak stałej w modelu. Innymi słowy, dlaczego požądane jest, żeby funkcja regresji przechodziła przez punkt (0; 0) w tym przypadku?
- b) (10%) Jak zmieni się oszacowanie parametru β_1 przy zmianie jednostek produkcji pszenicy na tony i jednocześnie zmianie roboczogodzin na roboczodni? Przyjąć 8-godzinny dzień pracy.
- a) Przy założeniu, że praca jest potrzebna do produkcji pszenicy, to przy zerowym nakładzie pracy otrzymamy zerową produkcję, dlatego stała powinna być równa zero.
- b) Definiujemy nowe zmienne:

$$\begin{aligned} pszenica^* &= pszenica/1000 \\ roboczogodziny^\dagger &= roboczogodziny/8. \end{aligned}$$

$$\begin{aligned} \mathbb{X} &= \begin{bmatrix} roboczogodziny_1 \\ roboczogodziny_2 \\ roboczogodziny_3 \\ \vdots \\ roboczogodziny_N \end{bmatrix} & \mathbb{X}^\dagger &= \begin{bmatrix} roboczogodziny_1^\dagger \\ roboczogodziny_2^\dagger \\ roboczogodziny_3^\dagger \\ \vdots \\ roboczogodziny_N^\dagger \end{bmatrix} = \begin{bmatrix} roboczogodziny_1/8 \\ roboczogodziny_2/8 \\ roboczogodziny_3/8 \\ \vdots \\ roboczogodziny_N/8 \end{bmatrix} \\ \mathbf{y} &= \begin{bmatrix} pszenica_1 \\ pszenica_2 \\ pszenica_3 \\ \vdots \\ pszenica_N \end{bmatrix} & \mathbf{y}^* &= \begin{bmatrix} pszenica_1^* \\ pszenica_2^* \\ pszenica_3^* \\ \vdots \\ pszenica_N^* \end{bmatrix} = \begin{bmatrix} pszenica_1/1000 \\ pszenica_2/1000 \\ pszenica_3/1000 \\ \vdots \\ pszenica_N/1000 \end{bmatrix} \end{aligned}$$

Oznaczmy przez $\tilde{\beta}$ oszacowanie parametru β_1 w modelu z przekształconymi zmiennymi. Wówczas, wykorzystując podstawowy wzór MNK otrzymujemy

$$\begin{aligned} \tilde{\beta} &= \left(\mathbb{X}^{\dagger'} \mathbb{X}^\dagger \right)^{-1} \mathbb{X}^{\dagger'} \mathbf{y}^* = \left((\mathbb{X}/8)' (\mathbb{X}/8) \right)^{-1} (\mathbb{X}/8)' \mathbf{y}/1000 = \\ &= \left(\frac{1}{64} \mathbb{X}' \mathbb{X} \right)^{-1} \frac{1}{8000} \mathbb{X}' \mathbf{y} = \frac{64}{8000} (\mathbb{X}' \mathbb{X})^{-1} \mathbb{X}' \mathbf{y} = \frac{1}{125} \hat{\beta} \end{aligned} \quad (2)$$