

Egzamin z Ekonometrii dla IiE i MSEMAt - 03.03.2020

Regulamin:

1. Przed przystąpieniem do pisania egzaminu należy podpisać wszystkie kartki arkusza kolokwium (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 końcu 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, 03-03-2020,

*Podpis***Pytania teoretyczne**

Pytanie 1 Za pomocą jakiego testu weryfikowana jest normalność składnika losowego? Jakiemu założeniu KMRL odpowiada H_0 w tym teście? Jaka jest hipoteza alternatywna w tym teście? Jakie są konsekwencje dla własności MNK, jeśli H_0 jest fałszywe?

Pytanie 2 Co to jest obserwacja nietypowa? Kiedy obserwację nietypową można uznać za błędną?

Pytanie 3 Co to znaczy, że estymator jest nieobciążony i dlaczego jest to ważne?

Pytanie 4 Udowodnić twierdzenie Gaussa-Markowa.

Zadanie 1 (35%)

Nagroda Pulitzera (ang. Pulitzer Prize) jest coroczną amerykańską nagrodą przyznaną za wybitne dokonania w dziedzinie dziennikarstwa, literatury pięknej i muzyki.¹ Nagroda Pulitzera to w Stanach jedno z najbardziej prestiżowych wyróżnień.² Nate Silver z blogu FiveThirtyEight³ przeanalizował zależność między liczbą zdobywców Nagrody Pulitzera a nakładami gazet, w których pracują zdobywcy.⁴ Państwa wykładowca postanowił wykorzystać dane⁵ udostępnione przez autora ww. artykułu i sprawdzić przedstawione tezy.

Zbiór danych zawiera informacje o nakładach 50 największych gazet w Stanach Zjednoczonych wraz z liczbą laureatów i nominowanych do Nagrody Pulitzera pracujących w każdym czasopiśmie. Tabela zamieszczona poniżej przedstawia oszacowania 4 modeli. W modelach (1)-(3) zmienną objaśnianą była *change0413*, czyli procentowa zmiana nakładu gazety 2013 rok w porównaniu do 2004 roku. W modelu (4) zmienną objaśnianą jest *circulation* - nakład gazety (liczba sztuk). Zmienna objaśniająca *pulitzer9014* zawiera liczbę laureatów i nominowanych do Nagrody Pulitzera pracujących w każdym czasopiśmie w latach 1990-2014. Zmienna objaśniająca *pulitzer9003* zawiera liczbę laureatów i nominowanych do Nagrody Pulitzera pracujących w każdym czasopiśmie w latach 1990-2003.

W większości analizowanych gazet obserwowane były spadki nakładu w okresie 2004-2013. Wydaje się, że dostępność informacji w Internecie może przekładać się na zmniejszenie popytu na prasę. Jednak, jak zauważa Nate Silver, na sprzedaż ma wpływ wiele czynników i mierzenie tylko i wyłącznie nakładu jest niewystarczające.

Dependent variable:				
	change04-13 (1)	change04-13 (2)	change04-13 (3)	circulation (4)
pulitzers9014	0.004* (0.001)	0.001 (0.002)		9250.601*** (2287.120)
pulitzers9003			0.006* (0.003)	
Constant	-0.350*** (0.043)	-0.322*** (0.040)	-0.344*** (0.044)	195899.000** (65565.780)
Observations	50	49	50	50
R2	0.124	0.004	0.087	0.254
Adjusted R2	0.106	-0.017	0.068	0.239
Residual Std. Error	0.256	0.237	0.261	394492.200
F Statistic	6.802*	0.212	4.555*	16.359***

Note: *p<0.05; **p<0.01; ***p<0.001

¹https://pl.wikipedia.org/wiki/Nagroda_Pulitzera

²<https://www.gandalf.com.pl/blog/literackie-pulitzery-ostatnich-lat/>

³<https://fivethirtyeight.com/>

⁴<https://fivethirtyeight.com/features/do-pulitzers-help-newspapers-keep-readers/>

⁵<https://github.com/fivethirtyeight/data/tree/master/pulitzer>

Przyjmij poziom istotności $\alpha = 5\%$.

1. (5%) Zinterpretuj parametr przy zmiennej *pulitzer9014* w modelu (1).
 2. (5%) Zinterpretuj parametr przy zmiennej *pulitzer9014* w modelu (2).
 3. (5%) Zinterpretuj parametr przy zmiennej *pulitzer9014* w modelu (4).
 4. (5%) New York Times jest czasopismem z największą liczbą nagród i nominowanych (117). Można podejrzewać, że New York Times jest outlierem. Model (2) został oszacowany z pominięciem obserwacji dla New York Timesa. Czy obserwacja dla tej gazety jest przykładem *bad leverage outlier*? (*Bad leverage outlier* to punkt którego uwzględnienie wpływa mocno na zmianę nachylenia funkcji regresji.)
 5. (5%) W artykule postawiono hipotezę, że większa liczba laureatów i nominowanych do Pulitzera powinna przekładać się na wyższy nakład. Jak w świetle uzyskanych oszacowań należałoby zweryfikować tę hipotezę?
 6. (5%) W modelu (3) zmienna objaśniająca informuje o liczbie Pulitzerów w latach 1990-2003. Czy wykorzystanie tej informacji zmienia wyniki? Uzasadnij.
 7. (5%) Czy w modelu może występować endogeniczność? Uzasadnij.
-
1. Dodatkowa nagroda Pulitzera dla dziennikarza przełoży się na wzrost nakładu gazety w której pracuje o 0.004 punktu procentowego.
 2. Dodatkowa nagroda Pulitzera dla dziennikarza nie przełoży się na wzrost nakładu gazety w której pracuje, ponieważ zmienna *Pulitzers9014* jest nieistotna w modelu (2).
 3. Dodatkowa nagroda Pulitzera dla dziennikarza przełoży się na wzrost nakładu gazety w której pracuje o 9250 sztuk.
 4. Tak, obserwacja dla New York Timesa jest *bad leverage outlierem*. W modelu (2), w którym pominięto tę obserwację, zmienna *Pulitzers9014* jest statystycznie nieistotna. Oznacza to, że z dodatniego nachylenia funkcji regresji otrzymujemy zerowe (statystycznie nieistotne) nachylenie. Wynika z tego, że uwzględnienie tej obserwacji zmienia nachylenie funkcji regresji.
 5. Hipotezę o pozytywnym wpływie liczby laureatów i nominowanych do nagrody Pulitzera na nakład należy zweryfikować pozytywnie. Pomimo braku wpływu na dynamikę (model bez *bad leverage outlier* zawiera nieistotną zmienną *pulitzers9014*), istnieje zależność między wielkością zmiennej *pulitzers9014* a nakładem w modelu (4). Można to rozumieć tak jak Nate Silver, który niecytowanym tu fragmencie swojego wpisu zauważa, że na dynamikę (ogólnie negatywną) ma wpływ wiele czynników (dostępność informacji w Internecie), jednak gazety zatrudniające dziennikarzy laureatów i nominowanych do nagrody Pulitzera mają wyższe nakłady.
 6. W modelu (3) wykorzystano zmienną *pulitzers9003*, ponieważ nagrody i nominacje po 2003 roku nie mogły mieć wpływu na sprzedaż w roku bazowym porównania - 2003. Wykorzystanie tej zmiennej nie zmienia wniosków. Sam parametr został oszacowany na wyższym poziomie (0.006 względem 0.004), ale wartość 0.004 znajduje się w przedziale ufności oszacowania z modelu (3).
 7. Endogeniczność jest związana z tym, że gazety z wyższymi nakładami mogą sobie pozwolić na zatrudnianie dziennikarzy laureatów i nominowanych do nagrody Pulitzera. Występuje tutaj zatem relacja zwrotna: liczba pulitzerów wpływa na nakład, a nakład na liczbę pulitzerów.

Zadanie 2 (40%)

Analitik miał za zadanie przeanalizować czy wydatki na obronność Brazylii są współmierne do wydatków krajów sąsiednich. Analitik zdecydował ograniczyć próbę analizowanych krajów do wszystkich państw Ameryki Południowej. W pierwszym kroku postanowił zbudować model ekonometryczny. Zmienną objaśnianą była zmienna *ArmedForces*, która kodowała liczebność wojska w tysiącach żołnierzy. Zmienną objaśniającą stanowiła zmienna *Population*, czyli liczba mieszkańców w milionach.

Call:

```
lm(formula = ArmedForces ~ Population)
```

Residuals:

Min	1Q	Median	3Q	Max
-51.133	-25.377	-18.049	8.215	120.405

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	58.2882	26.0611	2.237	0.06666 .
Population	1.7083	0.3609	4.733	0.00321 **

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

Residual standard error: 56.8 on 6 degrees of freedom

Multiple R-squared: 0.7887, Adjusted R-squared: 0.7535

F-statistic: 22.4 on 1 and 6 DF, p-value: 0.003215

RESET test

data: model1

RESET = 1.0287, df1 = 2, df2 = 4, p-value = 0.4361

studentized Breusch-Pagan test

data: model1

BP = 0.010393, df = 1, p-value = 0.9188

Jarque Bera Test

data: model1\$residuals

X-squared = 4.5231, df = 2, p-value = 0.1042

Przyjmij poziom istotności $\alpha = 5\%$.

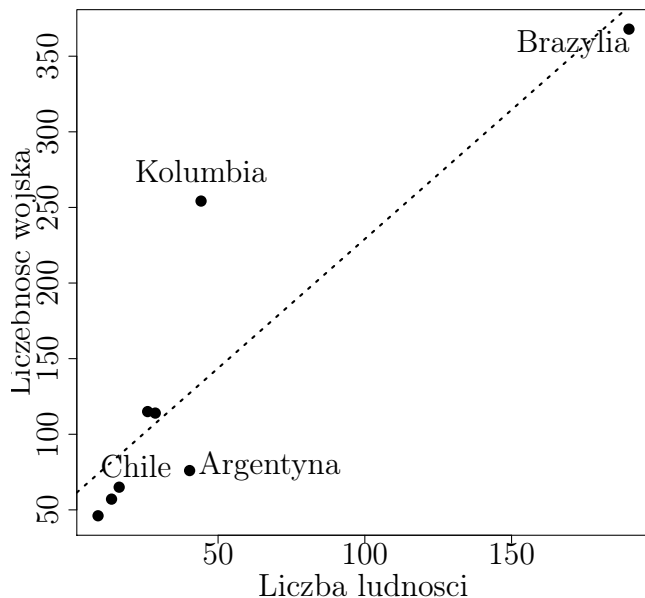
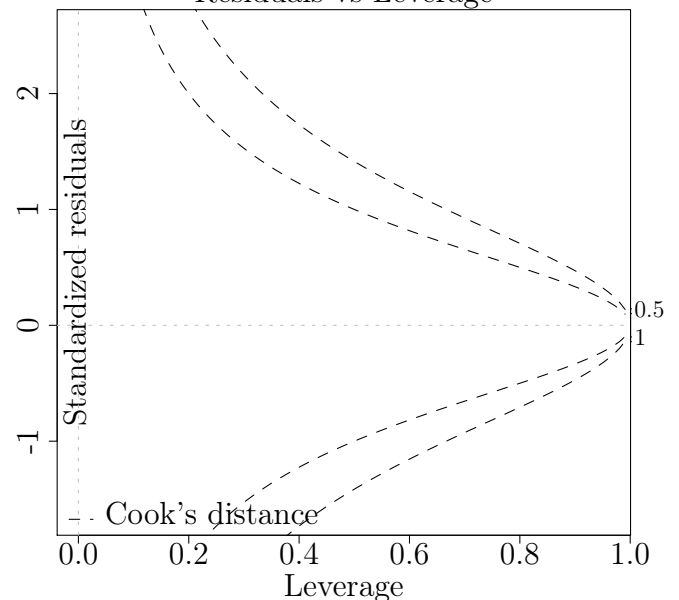
- (5%) Zinterpretuj wartość współczynnika determinacji liniowej.
- (5%) Zinterpretuj oszacowanie parametru przy zmiennej *Population*.
- (5%) Czy spełnione jest założenie o liniowości formy funkcyjnej w modelu? Uzasadnij.
- (5%) Czy spełnione jest założenie o homoskedastyczności składnika losowego? Uzasadnij.
- (5%) Czy spełnione jest założenie o normalności rozkładu składnika losowego? Uzasadnij.
- (5%) Czy spełnione są założenia KMRL? Uzasadnij.
- (5%) Czy można znaleźć nieobciążony i lepszy niż estymator MNK w tym przypadku? Uzasadnij.
- (5%) Podaj i uzasadnij o jaką zmienną można by rozszerzyć model oszacowany w tym zadaniu.

- a) $R^2 = 78.87\%$. Model wyjaśnia 78.87% zmienności wielkości sił zbrojnych.
- b) Dodatkowy milion mieszkańców związany jest ze wzrostem liczby żołnierzy o około 1708.
- c) Tak, statystyka w teście RESET Ramseya wynosi 1.0287 i odpowiadające jej p-value równe jest 43.61%. Oznacza to, że nie możemy odrzucić H_0 o liniowości formy funkcyjnej.
- d) Tak, to założenie jest spełnione. Na podstawie statystyki testowej w teście Breuscha-Pagana równej 0.010393 i odpowiadającego jej p-value=91.88% nie mam podstaw do odrzucenia H_0 o homoskedastyczności składnika losowego.
- e) Tak, to założenie jest spełnione. Statystyka testowa wynosi 4.5231, i odpowiadające jej p-value równe jest 10.42%. Oznacza to brak podstaw do odrzucenia H_0 o normalności składnika losowego.
- f) Tak, spełnione są założenia KMRL. Pokazaliśmy w punktach (c)-(d), że spełnione są założenia twierdzenia Gaussa-Markowa. Dodatkowe założenie o normalności rozkładu składnika losowego też jest spełnione.
- g) Przy spełnionych założeniach tw. G.-M. oraz dodatkowym o normalności rozkładu składnika losowego, estymator MNK jest najlepszym nieobciążonym estymatorem w klasie wszystkich estymatorów nieobciążonych.
- h) Produkt Krajowy Brutto. Im wyższy, tym łatwiej jest przeznaczyć więcej środków na obronność. Dostęp do morza/oceanu. Dostęp oznacza konieczność posiadania dodatkowego rodzaju sił zbrojnych (marynarki wojennej).

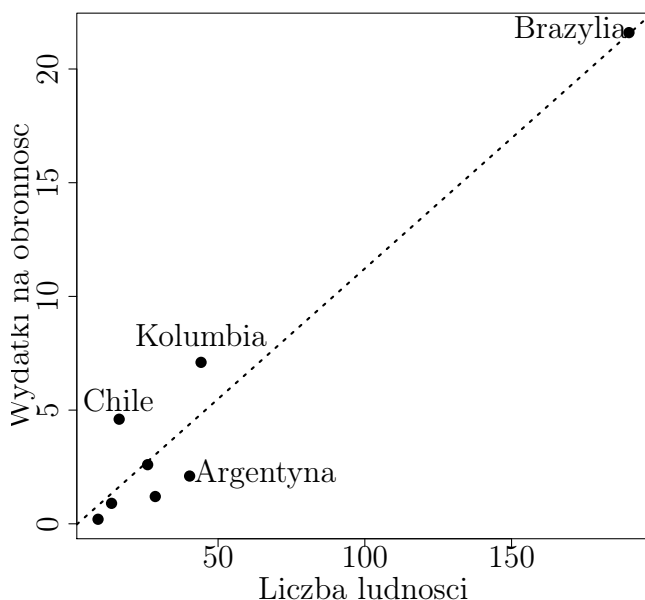
Zadanie 3 (25%)

Ten sam analityk, co w zadaniu 2, który miał za zadanie przeanalizować czy wydatki na obronność Brazylii, postanowił w drugim kroku przeanalizować występowanie obserwacji nietypowych i odstających. W tym celu sporządził wykres liczebności wojska (w tysiącach) w zależności od liczby ludności mieszkańców (w milionach). Wykres 1 przedstawia zależność między liczbą ludności krajów a liczebnością wojska.

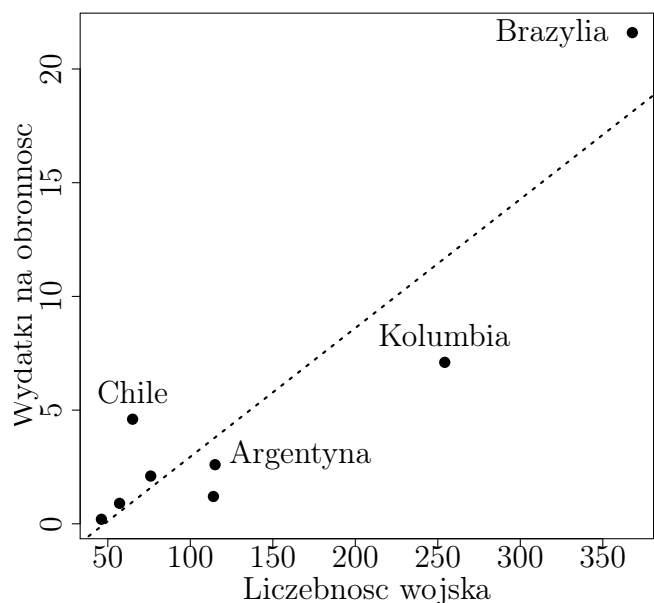
Wykres 1

Wykres 2
Residuals vs Leverage

Wykres 3



Wykres 4



- (5%) Czym można wytłumaczyć fakt, że Kolumbia ma nieproporcjonalnie dużą liczebność wojska na tle krajów regionu?
- (10%) Na podstawie wykresu 1 oceń jakim rodzajem obserwacji jest ta dla Kolumbii. Zaznacz jej miejsce na wykres 2 "Residuals vs Leverage". Uzasadnij swój wybór.
- (10%) Wykres 3 prezentuje dane oraz zależność między liczbą ludności kraju a wydatkami na obronność. Wykres 4 przedstawia wydatki na obronność w zależności od liczebności wojska. Ce-

lem analityka jest sprawdzenie czy wydatki na obronność Brazylii są współmierne do wydatków krajów sąsiednich. Brazylia, co jest widoczne na wykresie 1, ma najwyższe nominalne wydatki na obronność. Czy w wartościach względnych też tak jest? Uzasadnij.

- a) W Kolumbii rząd musi anagażować siły zbrojen w wojnie z organizacjami przestępczymi i paramilitarnymi produkującymi i handlującymi narkotykami. Oznacza to, że wojsko w Kolumbii, oprócz standardowych zadań, prowadzi jeszcze wojnę z tymi organizacjami, co skutkuje wyższą liczebnością wojska na tle krajów regionu.
- b) Pod względem charakterystyk x-owych (liczba ludności) Kolumbia znajduje się blisko średniej, zatem będzie charakteryzować się niską dźwignią (bliską zera). Reszta natomiast będzie duża. Przy definicji reszty $e_i = y_i - \hat{y}_i$ dla Kolumbii otrzymujemy dodatnią resztę, więc należy ją umieścić blisko górnego lewego rogu, na lewo od przerywanych linii.
- c) Wydatki na obronność względem liczby ludności w Brazylii są zgodne z linią regresji. Chile i Kolumbia wydają więcej *per capita* niż Brazylia. Wydatki na obronność względem liczebności wojska plasują Brazylię ponad linią trendu, podobnie jak Chile. Wydaje się, że wydatki na żołnierza będą wyższe w Chile niż w Brazylii (w przybliżeniu 5/70 vs. 22/370). Podsumowując, wydatki względne wyglądają korzystniej tylko dla Chile.