

imię, nazwisko, nr indeksu: \_\_\_\_\_

**Ekonometria — egzamin**  
**06/03/2019**

1. Egzamin trwa 90 minut.
2. Rozwiązywanie zadań należy rozpocząć po ogłoszeniu początku egzaminu a skończyć wraz z ogłoszeniem końca egzaminu. Złamanie tej zasady skutkuje usunięciem z sali i unieważnieniem pracy.
3. W trakcie egzaminu wolno używać jedynie długopisu o innym kolorze atramentu niż czerwony oraz kalkulatora.
4. Przed przystąpieniem do pisania egzaminu należy podpisać **wszystkie kartki** pracy (na dole w przewidzianym miejscu). Złożenie podpisu pod regulaminem oznacza jego akceptację. Do egzaminu mogą przystąpić osoby, które akceptują regulamin.
5. W razie braku podpisu lub numeru zadania na kartce, kartka nie zostanie oceniona. Nie będą też oceniane rozwiązania wpisane na kartkach innych, niż te rozdawane przez prowadzących.
6. Rozwiązanie każdego zadania należy zapisać na kartce z tymże zadaniem, ewentualnie na czystych kartkach znajdujących się na końcu arkusza egzaminacyjnego 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 czterech 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.
13. Warunkiem uzyskania oceny pozytywnej jest zdobycie conajmniej 50 % punktów z części teoretycznej egzaminu oraz min. 40 % punktów z części zadaniowej.

Warszawa, 06/03/2019, \_\_\_\_\_

*podpis*

Powodzenia :-)

Wartości krytyczne testu Durбина - Watsona.

n	$d_L$	$d_U$
50	1,50	1,59
100	1,65	1,69
200	1,76	1,78

Wartości krytyczne rozkładu  $\chi^2$ .

liczba stopni swobody	$\alpha = 0,95$	$\alpha = 0,99$
1	3,84	6,64
2	5,99	9,21
3	7,81	11,35
4	9,49	13,28
5	11,07	15,09
6	12,59	16,81
7	14,07	18,48
8	15,51	20,09
9	16,91	21,67
10	18,31	23,21

Wartości krytyczne rozkładu  $F$ .

liczba stopni swobody	$\alpha = 0,95$
F(1,20)	4,35
F(2,20)	3,49
F(3,20)	3,10
F(20,1)	248,02
F(20,2)	19,45
F(20,3)	8,66
F(1,500)	3,86
F(2,500)	3,01
F(3,500)	2,62
F(1, $\infty$ )	3,84
F(5, $\infty$ )	2,21

1	2	$\Sigma$

**Pytanie 1.**

Wyprowadź estymator wektora parametrów modelu regresji liniowej wielu zmiennych Metody Najmniejszych Kwadratów.

**Pytanie 2.**

Wyprowadź nieobciążony estymator wariancji składnika losowego w modelu regresji liniowej.

3	4	$\Sigma$

**Pytanie 3.**

Wyjaśnij co to jest homoscedastyczność składnika losowego i opisz wybrany sposób weryfikacji założenia o homoscedastyczności składnika losowego (podaj postać  $H_0$ ,  $H_A$ , statystyki testowej i opisz procedurę weryfikacji)

**Pytanie 4.**

Opisz korzyści i niebezpieczeństwa związane z dodawaniem zmiennych objaśniających do równania regresji i usuwaniem zmiennych objaśniających z równania regresji.

1	2	3	$\Sigma$

ZADANIE 1

Ekonometryk szacując funkcję wydatków na żywność stworzył osobne modele dla obszaru tzw. *Niemiec Zachodnich* oraz *Niemiec Wschodnich (dawnego NRD)*. Dla Niemiec Zachodnich posiadał próbę liczącą  $n_1 = 600$  obserwacji i uzyskał współczynnik dopasowania  $R_A^2 = 0.7$ , dla Niemiec Wschodnich odpowiednio  $n_2 = 400$  oraz  $R_B^2 = 0.4$ . Wiadomo również, że wariancja wydatków żywnościowych w zachodniej części Niemiec jest 3 razy większa niż we wschodniej części

Po przeprowadzeniu obliczeń zorientował się że przypadkowo pominął jedną obserwację dla Niemiec Wschodnich. Pasuje ona do prostej regresji, po jej dołączeniu całkowita suma kwadratów rośnie o 1%.

1. W jaki sposób dodanie obserwacji wpłynie na wielkość statystyk  $R^2$  oraz  $\bar{R}^2$
2. Oblicz prawidłowe wartości  $R^2$  oraz  $\bar{R}^2$  dla Wschodniej części Niemiec.
3. Jaką wartość będzie miał współczynnik  $R^2$  oraz  $\bar{R}^2$ , jeżeli funkcja wydatków na żywność zostanie oszacowana na podstawie modelu:

$$W = \alpha_A + \beta_A Y + \alpha_A D + \beta_B Y D + \xi$$

gdzie zmienna wskazująca  $D$  wyróżnia osoby z obszaru dawnej NRD (wschodniej części Niemiec). Dla uproszczenia obliczeń zignoruj pominiętą obserwację.

4. Następnie ekonometryk połączył obie próby i oszacował funkcję wydatków na żywność dla całych Niemiec. W jaki sposób można obliczyć  $R^2$  w tym modelu ?
5. W jaki sposób dysponując wynikami czterech powyższych modeli można sprawdzić czy krańcowa skłonność do wydatków na żywność w Niemczech jest identyczna?

1	2	3	$\Sigma$

## ZADANIE 2

Badacz w celu zmniejszenia uciążliwości współliniowości zmiennych objaśniających w modelu liniowym który spełnia wszystkie założenia KMRL do elementów diagonalnych macierzy  $X'X$  dodał pewną stałą  $c > 0$ , uzyskując model regresji grzbietowej (ang. *ridge regression*).

1. Czy estymator wektora parametrów w modelu regresji grzbietowej jest estymatorem liniowym?
2. Czy estymator wektora parametrów w modelu regresji grzbietowej jest estymatorem nieobciążonym?
3. Czy estymator wektora parametrów w modelu regresji grzbietowej ma nie większą wariancję niż wariancja tradycyjnego estymatora MNK?

Odpowiedzi uzasadnij.

1	2	3	4	5	$\Sigma$

ZADANIE 3 Na podstawie danych pochodzących z Badania Struktury Wynagrodzeń według zawodów w 2016 roku badacze postanowili sprawdzić czy płace sprzedawców sklepowych są ustalane zgodnie z teorią kapitału ludzkiego. W tym celu oszacowano rozszerzone równanie płacy typu Mincera. W tym równaniu logarytm miesięcznego wynagrodzenia brutto w złotych ( $l\_wyn$ ) tłumaczono stażem pracy pracownika i jego kwadratem oraz poziomem wykształcenia (zmienna wskazująca *wyzsze* równa 1 oznacza pracownika o wykształceniu wyższym, zmienna wskazująca *srednie* równa 1 oznacza pracownika o wykształceniu średnim, poziom bazowy wykształcenie podstawowe) i zmienną wskazującą czy pracuje w wymiarze pełnego etatu ( $pelny=1$ ). Uzyskano następujące wyniki: Przeciętna wartość zmiennej  $staz2$  wynosi 330

Number of obs = 298 R-squared = 0.0587  
 F(5, 292) = 3.64 Adj R-squared = 0.0426  
 Prob > F = 0.0033 Root MSE = .26327

$l\_wyn$	Coef.	Std. Err.	t	P> t
$staz$	0,0027	0,0049	0,54	0,587
$staz2$	-0,0001	0,0001	-0,67	0,501
$wyzsze$	0,1938	0,0552	3,51	0,001
$srednie$	-0,0139	0,0345	-0,40	0,687
$pelny$	0,0415	0,0657	0,63	0,528
$\_cons$	7,6711	0,0777	98,69	0,000

RESET test F(5, 288) = 0,51 P = 0,7721  
 Breusch-Pagan chi2(1) wartosc dopasowana = 20,69 P = 0,0000  
 Breusch-Pagan chi2(5) potegi zmiennych = 54,09 P = 0,0000  
 Jarque-Bera chi(2) = 70,70 P = 0,0000  
 Mean VIF = 5,17

gdzie P oznacza wartość P.

Odpowiedz na poniższe pytania, uzasadniając swoje odpowiedzi stosownymi obliczeniami lub wartościami statystyk. Przyjmij poziom istotności 5%.

1. Ile wynosi (semi)elastyczność wynagrodzenia względem stażu pracy?
2. Czy uzyskane wyniki oszacowań parametrów są zgodne z teorią kapitału ludzkiego?.
3. Czy formę funkcyjną modelu można uznać za poprawną?
4. Czy spełnione jest założenie o (hiper)sferyczności składnika losowego?
5. Jakie konsekwencje niesie niespełnienie założeń KMRL w tym modelu i w jaki sposób można rozwiązać ten problem.

## Rozwiązania zadań

### Zadanie 1

1. Ponieważ  $R^2 = 1 - \frac{RSS}{TSS}$ , a reszta dla pominiętej obserwacji wynosi 0, uwzględnienie dodatkowej obserwacji nie wpływa na wartość  $RSS$  jest stałe, wartość  $TSS$  zwiększa się, więc zarówno wartość statystyki  $R^2$  jak i statystyki  $\bar{R}^2$  zwiększa się.

$$\bar{R}^2 = 1 - \underbrace{\frac{n-1}{n-k} \frac{RSS}{TSS}}_{\substack{\text{maleje} \quad \text{maleje} \\ \text{rosnie}}}$$

2.  $R^2 = 1 - \frac{6}{10.1} = 0.4059$ ,  $\bar{R}^2 = 1 - \frac{n-1}{n-2}(1 - R^2) = 1 - \frac{399}{398}(0.5941) = 0.4044$ .
3. Szacujemy osobne modele za pomocą jednego równania, wobec tego:  $TSS_C = TSS_A + TSS_B = 4TSS_B$ ;  $RSS_A = 0.3TSS_A = 1.2TSS_B$ ;  $RSS_B = 0.6TSS_B$ ;  $RSS_C = RSS_A + RSS_B = (1.2 + 0.6)TSS_B$ ;  $R_C^2 = 1 - \frac{1.8}{4} = 0.45$ .
4. Nie da się obliczyć, ponieważ nie jest znana całkowita suma kwadratów oraz resztowa suma kwadratów tak utworzonego modelu.
5. Mając wyniki regresji z punktu 3 należy zdefiniować macierz ograniczeń  $R = [0, 1, 0, -1]$  i obliczyć statystykę Walda, ze wzoru:

$$(Rb - q)'[\sigma^2 R(X'X)^{-1} R']^{-1} (Rb - q)$$

i porównać ją z wartością krytyczną z rozkładu  $\chi^2(2)$ , lub  $F(2, 996)$

### Zadanie 2

Estymator regresji grzbietowej ma postać  $b_g = (X'X + c\mathbb{I})^{-1} X'y$

1. Estymator można zapisać jako  $b_g = Cy$ ; gdzie  $C = (X'X + c\mathbb{I})^{-1} X'$ . Zatem estymator regresji grzbietowej  $b_g$  jest estymatorem liniowym
- 2.

$$\begin{aligned} \mathbb{E}(b_g) &= \mathbb{E}[(X'X + c\mathbb{I})^{-1} X'y] = (X'X + c\mathbb{I})^{-1} X' \mathbb{E}(X\beta + \varepsilon) = \\ &= (X'X + c\mathbb{I})^{-1} X'X\beta \neq \beta \end{aligned}$$

Zatem estymator wektora parametrów w modelu regresji grzbietowej jest estymatorem obciążonym

3. Korzystając z twierdzenia o faktoryzacji macierzy macierz wariancji wektora parametrów jest dodatnio określona i można przedstawić jako

$$\sigma^2(X'X)^{-1} = \sigma^2(C\Lambda C^{-1})^{-1}$$

Z kolei wariancję estymatora regresji grzbietowej  $b_g$  jako

$$\sigma^2(C\Lambda C^{-1} + c\mathbb{I})^{-1}$$

Skoro  $c > 0$  to  $var(b_g) < var(b)$ . Zatem wariancja estymatora parametrów regresji grzbietowej jest nie większa niż wariancja estymatora MNK.

### Zadanie 3



1.

$$\frac{\partial l_{wyn}}{\partial staz} = \beta_{staz} + 2\beta_{staz^2} \bar{staz} = 0,0027 - 0,0002 \cdot 330 = -0,0633$$

2. Oszacowania parametrów przy stażu pracy są nieistotne statystycznie, co wskazuje, że dane o wynagrodzeniach sprzedawców nie potwierdzają teorii akumulacji kapitału ludzkiego.
3. Na podstawie wyniku testu RESET = 0,51,  $p$ -value = 0,77 brak jest podstaw do odrzucenia hipotezy o poprawności formy funkcyjnej.
4. Wyniki dwóch testów Bruesha -Pagana i  $p$ -value statystyki testowej równe 0 wskazują na odrzucenie hipotezy o homoscedastyczności składnika losowego, zatem nie jest on sferyczny.
5. Konsekwencją heteroscedastyczności są nieprawidłowe wartości obliczonych w standardowy sposób statystyk  $t$  oraz  $F$ . Rozwiązaniem problemu jest wykorzystanie macierzy odpornej w celu precyzyjnego oszacowania wariancji estymatora wektora parametrów.