

imię, nazwisko, nr indeksu: _____

Ekonometria — egzamin
31/01/2018

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, 31/01/2018, _____

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 χ^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, ∞)	3,84
F(5, ∞)	2,21

Pytanie 1.

Wyprowadź wzór na estymator MNK w Klasycznym Modelu Regresji Liniowej z wieloma zmiennymi objaśniającymi.

UWAGA: Część osób uważa, że mnożenie macierzy jest przemienne - nie jest. Rozpisanie sumy kwadratów reszt i wyniku końcowego nie jest pełnym wyprowadzeniem.

Pytanie 2.

Dlaczego w modelu regresji liniowej nie powinno się umieszczać równocześnie stałej i wszystkich zmiennych zero - jedynkowych, związanych z poziomami zmiennej dyskretnej?

UWAGA: Część osób twierdziła, że w modelu po prawej stronie nie mogą być wyłącznie zmienne zero-jedynkowe - nie jest to prawda.

Pytanie 3.

Pokaż, że w modelu oszacowanym na podstawie losowej próby estymator MNK jest nieobciążony. Zapisz niezbędną do przeprowadzenia dowodu założenia.

UWAGA: Część osób pokazała, że estymator MNK jest nieobciążony, gdy macierz X jest nielosowa.

Pytanie 4.

Opisz korzyści i niebezpieczeństwa związane z nakładaniem ograniczeń na parametry modelu.

UWAGA: Należało rozróżnić dwie sytuacje: ograniczenia są prawdziwe i ograniczenia są fałszywe. Częstą odpowiedzią było, że korzyścią będzie zmniejszenie wariancji estymatora. Jest to prawda, wtedy i tylko wtedy gdy ograniczenia nałożone na parametry modelu są prawidłowe.

1	2	3	4	Σ

ZADANIE 1

W ostatnich latach obserwuje się zwiększony wpływ migracji na funkcjonowanie lokalnych rynków pracy. Grupa badaczy postanowiła zbadać różnice w wynagrodzeniach pracowników lokalnych i napływowych o podobnym poziomie wykształcenia i doświadczenia zawodowego. W tym celu wykorzystano proste równanie płacy postaci:

$$w_i = \beta_0 + \beta_1 D_i + \varepsilon_i$$

w_i jest poziomem wynagrodzenia i -tego pracownika; D_i to zmienna przyjmująca wartość 0 dla pracowników lokalnych oraz 1 dla napływowych, ε_i to błąd losowy o jednakowym rozkładzie.

W próbie znalazło się n_{lok} pracowników lokalnych i n_{nap} pracowników przyjezdnych. Niech \bar{w}_{lok} oznacza przeciętne wynagrodzenie pracownika lokalnego, a \bar{w}_{nap} przeciętne wynagrodzenie pracownika napływowego. Pokaż, że:

1. Wartość estymatora parametru β_0 może być interpretowana jako przeciętne wynagrodzenie pracowników lokalnych,
2. Wartość estymatora parametru β_1 może być interpretowana jako przeciętna różnica w poziomie wynagrodzenia między pracownikami lokalnymi a napływowymi,
3. Przeciętne wynagrodzenie w próbie \bar{w} jest średnią ważoną wynagrodzenia pracowników lokalnych i napływowych,
4. Dodatkowo do modelu zmiennych opisujących poziom wykształcenia i doświadczenia zawodowego nie ma wpływu na wartość estymatora parametru β_1 .

1	2	Σ

ZADANIE 2

Powszechnie twierdzi się, że wykorzystując statystykę Durbina-Watsona w poprawny sposób można zweryfikować tylko występowanie autokorelacji pierwszego rzędu. Zakładając, że model regresji

$$y_t = X_t\beta + e_t$$

nie zawiera opóźnionej zmiennej zależnej, pokaż że:

1. jeżeli składnik losowy zawiera proces autoregresyjny pierwszego rzędu $e_t = \rho e_{t-1} + u_t, \forall t \text{ cov}(e_t, u_t) = 0$ to statystyka Durbina-Watsona ma prawidłową wartość,
2. jeżeli składnik losowy zawiera proces autoregresyjny drugiego (lub wyższego) rzędu, np. $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + u_t, \forall t \text{ cov}(e_t, u_t) = 0$ to wartość statystyki Durbina-Watsona będzie nieprawidłowa.

Podpowiedź

Skorzystaj z faktu, że można zapisać statystykę DW jako:

$$DW = \frac{\sum_{t=2}^T (e_t^2 - 2e_t e_{t-1} + e_{t-1}^2)}{\sum_{t=1}^T e_t^2}$$

dla dużego T możemy pominąć pojedyncze obserwacje otrzymując:

$$DW \approx 2 - 2 \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=1}^T e_{t-1}^2} \tag{1}$$

1	2	3	4	5	Σ

ZADANIE 3

Na podstawie danych dotyczących wynagrodzeń pracowników sektora budowlanego zatrudnionych w małych i średnich przedsiębiorstwach sektora prywatnego w województwie podlaskim oszacowano równanie płacy typu Mincera. W tym równaniu logarytm miesięcznego wynagrodzenia w złotych (l_{wyn}) tłumaczono wiekiem pracownika i kwadratem wieku, stażem pracy pracownika i jego kwadratem oraz poziomem wykształcenia (wyższe (bazowy), policealne, średnie zawodowe, średnie ogólne, zawodowe, gimnazjalne i niższe).

Uzyskano następujące wyniki:

Source	SS	df	MS	Number of obs =	476
-----+-----				F(9, 466)	= 19.08
Model	26,9873	9	2,9986	Prob > F	= 0.0000
Residual	73,2383	466	0,1572	R-squared	= 0.2693
-----+-----				Adj R-squared	= 0.2552
Total	100,2256	475	0,2110	Root MSE	= .39644

l_{wyn}	Coef.	Std. Err.	t	P> t
wiek	-0,0025	0,0176	-0,14	0,89
wiek2	0,0001	0,0002	0,37	0,71
staz	0,0276	0,0081	3,40	0,00
staz2	-0,0005	0,0002	-2,51	0,01
wyksz				
policealne	-0,2251	0,1228	-1,83	0,07
srednie zaw.	-0,4620	0,0535	-8,63	0,00
srednie og.	-0,3064	0,1078	-2,84	0,00
zawodowe	-0,5029	0,0471	-10,68	0,00
gimm. i nizsze	-0,5571	0,0906	-6,15	0,00
_cons	7,9499	0,3165	25,12	0,00

Następnie badacze oszacowali drugie równanie uzyskując wyniki:

Source	SS	df	MS	Number of obs =	476
-----+-----				F(7, 468)	= 24.28
Model	26,7021	7	3,8146	Prob > F	= 0.0000
Residual	73,5235	468	0,1571	R-squared	= 0.2664
-----+-----				Adj R-squared	= 0.2554
Total	100,2256	475	0,2110	Root MSE	= .39636

l_{wyn}	Coef.	Std. Err.	t	P> t
staz	0,0305	0,0061	5,00	0,00
staz2	-0,0005	0,0002	-3,10	0,00
wyksz				
policealne	-0,2148	0,1224	-1,75	0,08
srednie zaw.	-0,4593	0,0533	-8,62	0,00
srednie og.	-0,3053	0,1077	-2,83	0,00
zawodowe	-0,4959	0,0467	-10,61	0,00
gimm. i nizsze	-0,5385	0,0895	-6,02	0,00
_cons	7,9357	0,0525	151,26	0,00

Dla poprawnej specyfikacji przeprowadzili weryfikację założeń modelu (P oznacza wartość P):

Jarque- Bera	chi2(2)	= 30,98	P=0,00
RESET	F(3, 465)	= 2,61	P=0,05
RESET	F(5, 464)	= 2,67	P=0,02
Breusch-Pagan, wartosc dopasowana	chi2(1)	= 39,14	P=0,00
Breusch-Pagan, potegi zmiennych	chi2(7)	= 51,15	P=0,00
Breusch-Pagan, zmienna: nrjed	chi2(1)	= 3,48	P=0,06
Mean VIF		= 4,30	

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

1. Zweryfikuj przeprowadzając stosowny test, czy oszacowanie drugiego równania jest uzasadnione.
2. Zinterpretuj wpływ długości stażu pracownika na wysokość wynagrodzenia, jeżeli przeciętny staż pracownika wynosi 13,3 lat.
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 i jak można rozwiązać ten problem.

Rozwiązania zadań

Zadanie 1

1. Pierwszy sposób: korzystamy z własności MNK

$$\bar{w} |_{D_i=0} = \hat{w} |_{D_i=0} = b_0 + b_1 D_i |_{D_i=0} = b_0$$

Drugi sposób

$$E(w_i |_{D_i=0}) = E(b_0 + b_1 D_i + \varepsilon_i |_{D_i=0}) = \beta_0 + E(\varepsilon_i |_{D_i=0})$$

UWAGA: Nie jest prawdą, że $E(\varepsilon_i) = 0$. Z treści wynika, że „ ε_i to błąd losowy o jednakowym rozkładzie.” Średnia wartość błędu losowego nie jest znana.

2. Podstawiając do równania wynagrodzenia otrzymujemy

$$w_{nap} = \beta_0 + \beta_1 + \varepsilon_i$$

oraz

$$w_{lok} = \beta_0 + \varepsilon_i$$

wobec tego odejmując stronami

$$w_{nap} - w_{lok} = \beta_1$$

3. Korzystamy z własności MNK. Niech $N = n_{lok} + n_{nap}$

$$\bar{w} = \hat{w} = \beta_0 + \beta_1 D_i = \frac{1}{N} \left(\sum \beta_0 + \beta_1 D_i \right) = \frac{1}{N} (n_{lok} \beta_0 + n_{nap} (\beta_0 + \beta_1)) = \frac{n_{lok} \bar{w}_{lok} + n_{nap} \bar{w}_{nap}}{N}$$

UWAGA: Nawet, gdy model nie spełnia założeń KMRL estymator MNK dla wektora parametrów nadal jest najlepszym liniowym estymatorem, ale może być obciążony.

4. Podstawiając do równania wynagrodzenia rozszerzonego o dodatkowe zmienne otrzymujemy

$$w_{nap} = \beta_0 + \beta_1 + \gamma_1 edu + \gamma_2 exp + \varepsilon_i$$

oraz

$$w_{lok} = \beta_0 + \gamma_1 edu + \gamma_2 exp + \varepsilon_i$$

wobec tego odejmując stronami

$$w_{nap} - w_{lok} = \beta_1$$

Ponieważ pracownicy mają podobny poziom wykształcenia i doświadczenia zawodowego uwzględnienie tych zmiennych w równaniu płacy nie będzie miało wpływu na różnicę w wynagrodzeniach.

Zadanie 2

1. Proces AR(1) dany jest wzorem $e_t = \rho e_{t-1} + u_t$. Wstawiając to do (1) otrzymujemy:

$$DW = 2 - 2 \frac{\sum_{t=2}^T (\rho e_{t-1} + u_t) e_{t-1}}{\sum_{t=1}^T e_{t-1}^2}$$

Ponieważ u_t i e_t są nieskorelowane to:

$$DW = 2 - 2 \frac{\sum_{t=2}^T \rho e_{t-1}^2}{\sum_{t=1}^T e_{t-1}^2} = 2 - 2\rho$$

czyli w przypadku AR(1) wartość statystyki Durбина-Watsona jest prawidłowa.

2. Proces AR(2) dany jest wzorem $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + u_t$. Wstawiając to do (1) otrzymujemy:

$$DW = 2 - 2 \frac{\sum_{t=2}^T (\rho_1 e_{t-1} + \rho_2 e_{t-2} + u_t) e_{t-1}}{\sum_{t=1}^T e_{t-1}^2}$$

Z braku korelacji u_t i e_t mamy:

$$DW = 2 - 2 \frac{\sum_{t=2}^T (\rho_1 e_{t-1}^2 + \rho_2 e_{t-1} e_{t-2})}{\sum_{t=1}^T e_{t-1}^2}$$

$$DW = 2 - 2\rho_1 + 2\rho_2 \frac{\sum_{t=2}^T e_{t-1} e_{t-2}}{\sum_{t=1}^T e_{t-1}^2}$$

W przypadku występowania procesu AR(2) wartość statystyki DW będzie zaburzona.

Zadanie 3

1. Równanie (2) można traktować jak równanie (1) z narzuconymi ograniczeniami na parametry przy zmiennych **wiek** oraz **wiek2**

$$F = \frac{(73,5235 - 73,2383)/2}{73,2383/(476 - 10)} = \frac{0,1426}{0,1572} = 0,9071 < F(2, 500) = 3,01$$

Na poziomie istotności 5% nie ma podstaw do odrzucenia hipotezy o poprawności ograniczeń, zatem oszacowanie równania 2 jest uzasadnione.

2. W tym celu należy obliczyć wartość pochodną cząstkowej względem wieku

$$\frac{\partial \ln \text{wyn}}{\partial \text{staz}} = \beta_{\text{staz}} + 2\beta_{\text{staz}^2} \bar{\text{staz}} = 0,0305 + 2 \times -0,0005 \times 13,3 = 0,0305 - 0,0133 = 0,0172$$

Dla osoby o przeciętnym w próbie wynagrodzeniu i stażu pracy, kolejny rok stażu pracy przyczynia się do wzrostu oczekiwanego wynagrodzenia o 1,72% przy innych czynnikach stałych.

3. Co prawda wynik testu RESET: $F(5, 464) = 2,67$; $P = 0,02 < \alpha$; wskazuje, że nie jest spełnione założenie o liniowej formie modelu jednak parametry modelu szacowane są na podstawie próby o dużej liczbie obserwacji, a w takim przypadku statystyka testowa jest bardzo wrażliwa, po drugiej formie funkcyjnej równania jest uzasadniona teoretycznie.

4. Testami pomagającymi weryfikować hipersferyczność składnika losowego są testy Breusch-Pagana. Wyniki testów sugerują konieczność odrzucenia założenia o stałej wariancji $\chi^2(1) = 39,14$; $P = 0,00 < \alpha$; $\chi^2(7) = 51,15$; $P = 0,00 < \alpha$. Zatem składnik losowy modelu nie jest hipersferą.

UWAGA: Test normalności rozkładu składnika losowego nie jest testem sferyczności.

5. Konsekwencją heteroscedastyczności są nieprawidłowe wartości obliczonych w standardowy sposób statystyk t oraz F . Rozwiązaniem problemu jest wykorzystanie odpornej macierzy White'a w celu oszacowania wariancji estymatora wektora parametrów.