

IMIĘ I NAZWISKO (NR GRUPY):

KOŁOKWIUM Z EKONOMETRII

STYCZEŃ 2023

Czas pracy wynosi 90 min. Podpisz kartę z zadaniami oraz kartki z odpowiedziami. W zadaniach przyjmij że $\alpha = 0.05$ oraz $t^* = 1.96$. Jeśli masz pytanie, podnieś rękę. Jeśli coś jest niejasne, pytaj. Maksymalna liczba punktów wynosi 40. Możesz używać kalkulatora. Gdy przeprowadzasz test statystyczny, zawsze podaj hipotezę zerową.

1. (18 p.) Poszczególne pytania w tym zadaniu odnoszą się do Tabeli 1 przedstawiającej oszacowania, obliczone na danych CPS (*Current Population Survey*). Dane opisują 7440 pracowników (pracują cały rok na pełen etat). Najwyższym osiągnięciem dla każdego z pracowników było albo ukończenie liceum lub dyplom licencjacki. Wiek pracowników zawiera się pomiędzy 25 a 34. Zbiór danych zawiera także informacje o regionie kraju w którym dana osoba żyje, statusie matrymonialnym oraz o liczbie dzieci:

- AHE = średnia wynagrodzenie na godzinę (w dolarach z 2012 roku)
- College = zm. binarna (1 jeśli dyplom lic., 0 jeśli liceum)
- Female = zm. binarna (1 jeśli kobieta, 0 jeśli mężczyzna)
- Age = wiek (w latach)
- Ntheast = zm. binarna (1 jeśli region to Northeast, 0 w przeciwnym przypadku)
- Midwest = zm. binarna (1 jeśli region to Midwest, 0 w przeciwnym przypadku)
- South = zm. binarna (1 jeśli region to South, 0 w przeciwnym przypadku)
- West = zm. binarna (1 jeśli region to West, 0 w przeciwnym przypadku)

Tabela 1: Zarobki w zależności od wykształcenia, płci i innych charakterystyk.

	zm. zależna: AHE			zm. zależna: log(AHE)	
	(1)	(2)	(3)	(4)	(5)
College (X1)	8.31 (0.23)	8.32 (0.22)	8.34 (0.22)	0.44 (0.01)	0.40 (0.01)
Female (X2)	-3.85 (0.23)	-3.81 (0.22)	-3.80 (0.22)	-0.19 (0.01)	-0.24 (0.17)
Age (X3)		0.51 (0.04)	0.52 (0.04)	0.10 (0.04)	0.10 (0.04)
Age2 (X4)				-0.001 (-0.0008)	-0.001 (-0.0008)
Female×College (X5)					0.09 (0.02)
Northeast (X6)			0.18 (0.36)		
Midwest (X7)			-1.23 (0.31)		
South (X8)			-0.43 (0.30)		
stała	17.02 (0.17)	1.87 (1.18)	2.05 (1.18)	-0.79 (0.67)	0.80 (0.67)
R^2	0.162	0.180	0.182	0.197	0.198
n	7440	7440	7440	7440	7440

W nawiasach () podano odchylenia standardowe.

- (a) Policz skorygowany R^2 dla regresji (1). Wyjaśnij różnice między zwykłym i skorygowanym R^2 .

Odpowiedz, używając wyników z kolumny (1):

- (b) Czy pracownicy z dyplomami licencjackimi zarabiają więcej niż pracownicy z ukończonym liceum? O ile więcej? Czy ta różnica jest statystycznie istotna?

(c) Czy mężczyźni zarabiają więcej niż kobiety? O ile? Czy ta różnica jest istotna statystycznie?

Odpowiedz, używając oszacowań z kolumny (2):

(d) Czy wiek jest istotną determinantą zarobków? Zbuduj 95% przedział ufności i skomentuj.

Odpowiedz, używając oszacowań z kolumny (3):

(e) Czy różnice międzyregionalne są łącznie istotne? Przetestuj odpowiednią hipotezę i odpowiedz. Wartość krytyczna z odpowiedniego rozkładu dla $\alpha = 0.05$ wynosi 0.485844.

(f) Dlaczego zmienna *West* została pominięta w regresji? Co by się stało, gdyby została umieszczona w równaniu?

Odpowiedz, używając oszacowań z kolumny (4):

(g) Jeśli wiek wzrasta z 25 do 26 przy innych czynnikach niezmiennych, jak zmienia się zarobki? A jak w przypadku gdy wiek wzrasta z 33 do 34? Użyj odpowiedniego przybliżenia.

(h) Przyjmij, że wpływ wieku na zarobki może być różny dla absolwentów liceum i dla absolwentów uniwersytetu. Zmodyfikuj równanie regresji w taki sposób, aby uchwyciło te różnice.

Odpowiedz, używając oszacowań z kolumny (5):

(i) Co mierzy współczynnik przy interakcji? Zinterpretuj go.

2. (12 p.) W tym zadaniu, pytania odnoszą się do oszacowań luki płacowej wysoko wykwalifikowanych imigrantów, uzyskanych na danych z amerykańskiego badania *Survey of Income and Program Participation (SIPP)*, z lat 2008-2013:

- *ln_wage* - logarytm zarobków
- *immigr* - zm. binarna (1 jeśli imigrant, 0 jeśli nie)
- *age* - wiek (w latach)
- *age2* - wiek do kwadratu
- *female* - zm. binarna (1 jeśli kobieta, 0 jeśli mężczyzna)
- *black* - zm. binarna (1 jeśli rasa czarna, 0 w przeciwnym przypadku)
- *asian* - zm. binarna (1 jeśli rasa żółta, 0 w przeciwnym przypadku)
- *cognitive* - zm. binarna (1 jeśli pracownik umysłowy, 0 jeśli pracownik fizyczny)

Wszystkie osoby w próbie są wysoko wykwalifikowane, tzn. mają co najmniej dyplom licencjacki lub wyższy.

Tabela 2: Zarobki w zależności od statusu imigracyjnego i innych charakterystyk

	(1) oszacowania na całej próbie					(2) oszacowania na podpróbach	
	oszacowania	bł. std.	stat. t	p-value	VIF	umysłowi	fizyczni
<i>immigr</i>	-0.09856	0.004359	-22.61	2×10^{-16}	1.406877	0.0262	-0.178
<i>age</i>	0.1112	0.001333	83.41	2×10^{-16}	89.434584	0.0944	0.09735
<i>age2</i>	-0.0001	0.000016	-77.12	2×10^{-16}	89.417841	-0.0009	-0.0011
<i>female</i>	-0.4289	0.002658	-161.39	2×10^{-16}	1.005427	-0.4077	-0.4157
<i>black</i>	-0.2053	0.005058	-40.59	2×10^{-16}	1.020609	-0.1456	-0.1319
<i>asian</i>	0.1664	0.005723	29.08	2×10^{-16}	1.415301	0.1076	0.0708
<i>stała</i>	6.151	0.02732	225.13	2×10^{-16}		6.645	6.166
SSR	224586					109858	84057
R^2	0.1019					0.1137	0.0882
<i>n</i>	358688					223083	135605
<i>Statystyki:</i>							
Statystyka testu White'a:		81.89	p-value	2.2×10^{-16}			
Statystyka testu RESET:		445.44	p-value	2.2×10^{-16}			
Statystyka testu Jarque-Bery:		499017	p-value	2.2×10^{-16}			

Zmienną zależną w oszacowanych modelach jest oczywiście *ln_wage*.

Odpowiedz na pytania korzystając z wyników z (1) części Tabeli 2

- (a) Zinterpretuj współczynnik przy zmiennej *immigr*. Wyjaśnij co to oznacza w kategoriach ekonomicznych.
 - (b) Czy współliniowość jest problemem w analizowanym modelu? Uzasadnij, opisz jej konsekwencje oraz podaj ewentualne rozwiązanie tego problemu.
 - (c) Czy w modelu występuje heteroskedastyczność? Uzasadnij. Jeśli tak, to co powinien zrobić ekonometryk? Jakie problemy powoduje heteroskedastyczność?
 - (d) Czy zastosowana forma funkcyjna jest poprawna? Uzasadnij.
 - (e) Czy składnik losowy ma rozkład normalny w rozważanym modelu? Uzasadnij. Co się dzieje gdy tak nie jest?
 - (f) W zbiorze danych zawarta jest zmienna która odróżnia pracowników umysłowych i fizycznych (cognitive). W dwóch ostatnich kolumnach, zawarto oszacowania modelu, odpowiednio dla pracowników umysłowych i fizycznych. Czy istnieje strukturalna różnica między oszacowanymi parametrami w tych dwóch regresjach? Przeprowadź odpowiedni test i odpowiedz. (Wykorzystaj również wyniki z (2) części Tabeli 2) Wartość krytyczna z rozkładu F wynosi 0.7633769.
3. (3 p.) Ekonometryk w ramach swojego projektu chce zbadać jak na stopę zabójstw (per capita) wpływają zmiany w nakładach na policję (per capita) w powiatach w Polsce. Można się spodziewać, że przekazanie większych środków policji, powinno przyczynić się do spadku odsetka zabójstw w danym powiecie. Dodatkową zmienną którą warto uwzględnić w regresji to czy gangi są obecne w danym powiecie. Tak więc nasz Ekonometryk chciałby oszacować następujące równanie:

$$zabojstwa_per_capita = \beta_0 + \beta_1 finansowanie_policji + \beta_2 gang + \varepsilon$$

Jednak nasz Ekonometryk zapomniał zebrać dane o obecności gangów w powiatach, więc w praktyce może jedynie oszacować następujące równanie:

$$zabjstwa_per_capita = \beta_0 + \beta_1 finansowanie_policji + \varepsilon$$

- (a) Ekonometryk popełnił tutaj błąd zmiennej pominiętej. Wyjaśnij krótko jakie konsekwencje niesie on dla oszacowania parametru β_1 z równania (2)?
 - (b) Jeśli Ekonometryk oszacuje równanie (2), to czy $\hat{\beta}_1$ będzie przeszacowywało czy niedoszacowywało prawdziwy parametr β_1 z równania (1)? Uzasadnij swoją odpowiedź odpowiednim rozumowaniem.
4. (7 p.) Klasyczny model regresji liniowej - podaj założenia. Opisz co mówi tw. Gaussa-Markova.