

Zadanie 1

Przeprowadzono regresję wyjaśniającą logarytm zarobków osiągniętych (l_dochod) przez zamężne kobiety za pomocą ilości lat poświęconych przez nie nauce (wyksz).

Model 1: Estymacja KMNK z wykorzystaniem 1118 obserwacji 1-1118
Zmienna zależna: l_dochod

	<i>Współczynnik</i>	<i>Błąd stand.</i>	<i>t-Student</i>	<i>wartość p</i>	
const	7,45161	0,0318137	234,2264	<0,00001	***
wyksz	0,0650681	0,00796457	16,7724	<0,00001	***

Średnia arytmetyczna zmiennej zależnej = 6,9665

Odchylenie standardowe zmiennej zależnej = 0,495513

Suma kwadratów reszt = 219,045

Błąd standardowy reszt = 0,443031

Wsp. determinacji $R^2 = 0,20133$

Skorygowany $R^2 = 0,20061$

Stopnie swobody = 1116

1. Podać interpretację oszacowanych parametrów w tym modelu oraz interpretację statystyki R^2 .
2. Jaki będzie prawdopodobny kierunek obciążenia oszacowania parametru przy zmiennej wyksz wynikły z pominięcia:
 - (a) inteligencji respondentki.
 - (b) liczby dzieci respondentki.
 - (c) wielkości miejscowości, w której zamieszkuje respondentka
 - (d) zarobków męża respondentki.

Zadanie 2

Dane dla Stanów Zjednoczonych. Opis zmiennych:

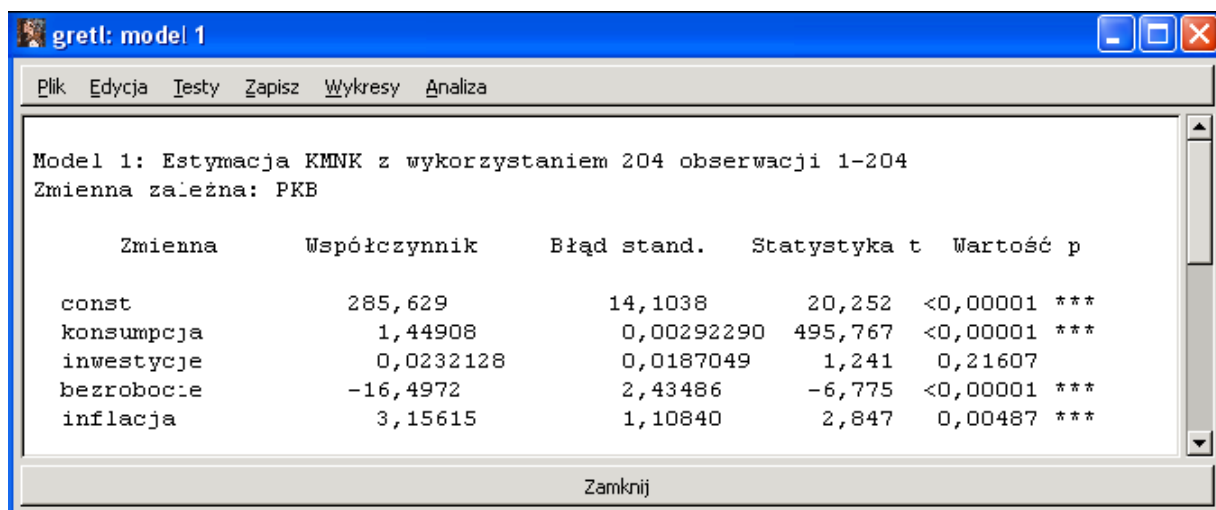
PKB – PKB wyrażone w bilionach dolarów (zmienna zależna w regresji)

konsumpcja – konsumpcja wyrażona w bilionach dolarów

inwestycje – inwestycje wyrażone w miliardach dolarów

bezrobocie – stopa bezrobocia wyrażona w procentach

inflacja – poziom inflacji wyrażony w procentach



Zmienna	Współczynnik	Błąd stand.	Statystyka t	Wartość p
const	285,629	14,1038	20,252	<0,00001 ***
konsumpcja	1,44908	0,00292290	495,767	<0,00001 ***
inwestycje	0,0232128	0,0187049	1,241	0,21607
bezrobocie	-16,4972	2,43486	-6,775	<0,00001 ***
inflacja	3,15615	1,10840	2,847	0,00487 ***

1. Które zmienne w modelu można uznać za istotne na poziomie istotności $\alpha = 5\%$?
2. Które zmienne powinno się usunąć z modelu i dlaczego?
3. Jaki wniosek związany z wpływem inwestycji na PKB wynika z tych oszacowań?
4. Jakie najprawdopodobniej będą relacje między odchyleniami standardowymi, statystykami t oraz R^2 w regresji ze wszystkimi zmiennymi i w regresji z usuniętymi zmiennymi nieistotnymi?

Zadanie 3

Analizowany jest następujący model:

$$c_i = \beta_1 + \beta_2 y_i + \beta_3 p_i + \varepsilon_i, \quad (**)$$

gdzie c_i oznacza logarytm nominalnych wydatków i -tego gospodarstwa domowego na żywność, y_i jest logarytmem jego nominalnego dochodu, a p_i - logarytmem indeksu cen żywności.

2. Jaka jest interpretacja β_2 w tym modelu?
3. W przypadku gospodarstw wiejskich poziom dochodu zależy od cen żywności. Jaki problem ekonometryczny może z tego powodu wystąpić dla tej grupy gospodarstw?
4. Jeśli miejsce zamieszkania wpływa na c_i , a estymowany jest model (*), to jakie będą własności estymatora MNK?
5. Jeśli miejsce zamieszkania nie wpływa na c_i , ale w estymowanym modelu zawarte są zmienne zero-jedynkowe związane z miejscem zamieszkania, to jakie będą własności estymatora MNK?

Zadanie 4

Opis zmiennych:

wiek - wiek w latach

waga - waga w kilogramach

tlen - ilość pobieranego tlenu w mililitrze na 1 kg masy ciała w ciągu minuty

bieg - czas biegu 1500m

puls_odp - puls po zakończeniu biegu, w czasie odpoczynku

puls_bieg - puls w trakcie biegu

puls_max - maksymalny puls w trakcie biegu.

1. Wyznaczono macierz korelacji dla zmiennych znajdujących się w zbiorze danych.
Czy widać potencjalne zagrożenia w regresji zmiennej **bieg** na zmiennych **wiek**, **waga**, **puls_bieg**, **puls_max**?

gretl: macierz korelacji

Współczynniki korelacji, wykorzystane obserwacje 1 - 31
Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,3550 dla n = 31

wiek	waga	tlen	bieg	puls_odp	
1,0000	-0,2335	-0,3046	0,1887	-0,1641	wiek
	1,0000	-0,1628	0,1435	0,0440	waga
		1,0000	-0,8622	-0,3994	tlen
			1,0000	0,4504	bieg
				1,0000	puls_odp
puls_bieg					
-0,3379	-0,4329				wiek
0,1815	0,2494				waga
-0,3980	-0,2367				tlen
0,3136	0,2261				bieg
0,3525	0,3051				puls_odp
1,0000	0,9298				puls_bieg
	1,0000				puls_max

Na uwagę zasługuje silna zależność między zmiennymi **puls_bieg** i **puls_max**. Można spodziewać się, że umieszczenie obu tych zmiennych w modelu w charakterze regresorów może wywołać problem współliniowości.

2. Przeprowadzono regresję zmiennej **bieg** na zmiennych **wiek**, **waga**, **puls_bieg**, **puls_max**. Czy pojawia się problem współliniowości?

gretl: model 6

Model 6: Estymacja KMNK z wykorzystaniem 31 obserwacji 1-31
Zmienna zależna: bieg

	współczynnik	błąd standardowy	t-Student	wartość p
const	-3,29820	6,85940	-0,4808	0,6347
wiek	0,0867860	0,0522243	1,662	0,1086
waga	0,0287540	0,0301380	0,9541	0,3488
puls_bieg	0,0908954	0,0649721	1,399	0,1736
puls_max	-0,0454626	0,0765152	-0,5942	0,5575

Średnia arytmetyczna zmiennej zależnej = 10,5861
 Odchylenie standardowe zmiennej zależnej = 1,38741
 Suma kwadratów reszt = 44,5196
 Błąd standardowy reszt = 1,30855
 Wsp. determinacji R-kwadrat = 0,22907
 Skorygowany wsp. R-kwadrat = 0,11046
 Statystyka F (4, 26) = 1,93133 (wartość p = 0,135)
 Logarytm wiarygodności = -49,5972
 Kryterium informacyjne Akaike'a (AIC) = 109,194
 Kryterium bayesowskie Schwarz (BIC) = 116,364
 Kryterium infor. Hannana-Quinna (HQC) = 111,532

Wyłączając stałą, największa wartość p jest dla zmiennej 7 (puls_max)

gretl: ocena współliniowości VIF (Variance Inflation Factor)

Ocena współliniowości VIF - czynnika powiększania wariancji

Minimalna możliwa wartość = 1.0
 Wartości > 10.0 mogą wskazywać na problem współliniowości-rozdęcia wariancji

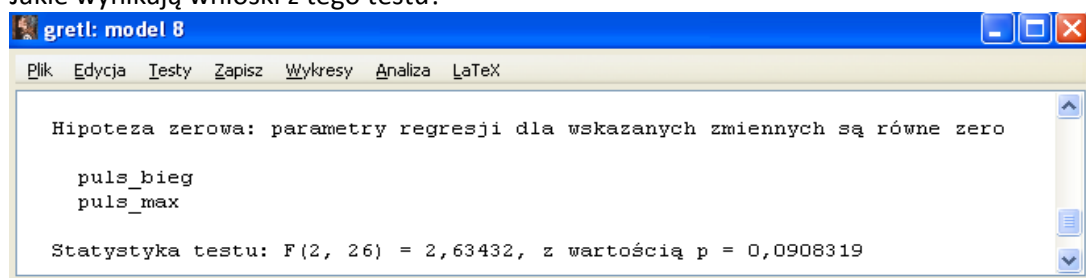
1)	wiek	1,298
2)	waga	1,104
6)	puls_bieg	7,773
7)	puls_max	8,614

$VIF(j) = 1/(1 - R(j)^2)$, gdzie $R(j)$ jest współczynnikiem korelacji wielorakiej pomiędzy zmienną 'j' a pozostałymi zmiennymi niezależnymi modelu.

Własności macierzy $X'X$:

1-norm = 2534509,6
 Wyznacznik = 4,0085415e+013
 Wskaźnik uwarunkowania macierzy CN = 1,4156921e-008

3. Przetestowano łączną nieistotność zmiennych **puls_bieg** i **puls_max** na poziomie istotności 0,1. Jakiego wniosku z tego testu?



4. Czy występują obserwacje nietypowe?

The screenshot shows the 'gretl: obserwacje dźwigniowe (leverage) i wpływowe (influence)' window. It displays a table with four columns: 'reszty u', 'leverage 0<=h<=1', 'influence u*h/(1-h)', and 'DFFITS'. The table lists 31 observations. Observation 28 is marked with an asterisk in the leverage column, indicating it is a leverage point. A note at the bottom states: '*** oznacza dźwigniową obserwację (leverage point), h(i)>2(k+1)/n'.

	reszty u	leverage 0<=h<=1	influence u*h/(1-h)	DFFITS
1	0,36866	0,114	0,047359	0,107
2	-0,56218	0,162	-0,10872	-0,206
3	-1,0428	0,179	-0,22744	-0,414
4	-1,428	0,151	-0,25324	-0,509
5	-1,1905	0,180	-0,2608	-0,476
6	0,75807	0,046	0,036249	0,130
7	1,7916	0,115	0,2336	0,544
8	1,0417	0,105	0,12244	0,291
9	2,5063	0,053	0,14078	0,502
10	-1,155	0,161	-0,22232	-0,428
11	0,10313	0,070	0,0077464	0,022
12	2,5366	0,149	0,4434	0,956
13	0,72356	0,128	0,10622	0,227
14	0,45783	0,059	0,028607	0,090
15	-0,81089	0,114	-0,10458	-0,237
16	-2,4379	0,086	-0,22816	-0,640
17	0,33458	0,069	0,024616	0,071
18	-0,49653	0,059	-0,031087	-0,097
19	-0,3091	0,168	-0,062436	-0,116
20	-0,43834	0,041	-0,018718	-0,070
21	1,0078	0,180	0,22135	0,401
22	0,67755	0,127	0,098635	0,211
23	-1,0327	0,058	-0,063856	-0,203
24	-0,41536	0,221	-0,11752	-0,190
25	0,30471	0,099	0,033326	0,081
26	1,4102	0,221	0,3989	0,664
27	-1,2445	0,178	-0,2703	-0,496
28	0,11099	0,317*	0,051568	0,070
29	-2,1884	0,143	-0,36621	-0,784
30	1,2791	0,167	0,25562	0,486
31	-0,66004	0,082	-0,058684	-0,157

*** oznacza dźwigniową obserwację (leverage point), $h(i) > 2(k+1)/n$

5. Narysowano wykresy reszt względem regresorów.
Czy słuszne jest założenie o stałości wariancji?

