

**Egzamin z ekonometrii – wersja ogólna**  
**08-02-2017**

**Pytania teoretyczne**

1. Za pomocą którego testu testujemy stabilność parametrów? Jakiemu założeniu KMRL odpowiada  $H_0$  w tym teście? Jaka jest hipoteza alternatywna w tym teście?
2. Wyprowadzić estymator MNK dla modelu z wieloma zmiennymi objaśniającymi.
3. Wyjaśnić różnice między parametrami i oszacowaniami parametrów oraz między odchyleniami losowymi i resztami.
4. Dlaczego zmienną dyskretną rozkodowywujemy na zmienne zerojedynkowe?

## Zadanie 1

Na podstawie pewnej bazy danych analizowano determinanty długości trwania bezrobocia w miesiącach (*Bezrobocie*). Zmiennymi objaśniającymi są wiek wyrażony w latach (*wiek*), wiek podniesiony do kwadratu (*wiek\_2*), wykształcenie średnie (*srednie*: 1- jeśli dla osoby najwyższym ukończonym poziomem wykształcenia jest wykształcenie średnie, 0- w pozostałych przypadkach), wykształcenie wyższe (*wyzsze*: 1- jeśli dla osoby najwyższym ukończonym poziomem wykształcenia jest wykształcenie wyższe, 0- pozostałych przypadkach), znajomość angielskiego (*angielski*: 1- jeśli osoba zna biegle język angielski w mowie i piśmie, 0- pozostałych przypadkach), interakcja między zmienną *srednie* a *angielski*, interakcja między zmienną *wyzsze* a *angielski*, płeć (*plec*: 0 – mężczyzna, 1- kobieta).

Hipotezy testować na poziomie istotności 0,05. Odpowiedzi uzasadnić podając p-value.

Source	SS	df	MS			
Model	5346.40487	8	668.300609	Number of obs =	1085	
Residual	59033.7979	1076	54.8641244	F( 8, 1076) =	12.18	
Total	64380.2028	1084	59.391331	Prob > F =	0.0000	
				R-squared =	0.0830	
				Adj R-squared =	0.0762	
				Root MSE =	7.407	

Bezrobocie	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiek	.6579234	.1261968	5.21	0.000	.4103038	.905543
wiek_2	-.0074501	.0016	-4.66	0.000	-.0105895	-.0043106
srednie	-1.546159	.6146765	-2.52	0.012	-2.752259	-.3400584
wyzsze	-5.118414	1.437908	-3.56	0.000	-7.939837	-2.296992
angielski	-2.578042	2.470105	-1.04	0.297	-7.424812	2.268727
sredniexang	2.167486	2.542764	0.85	0.394	-2.821852	7.156825
wyszszexang	2.119629	2.988318	0.71	0.478	-3.743963	7.983221
plec	1.88982	.4529972	4.17	0.000	1.000962	2.778678
_cons	1.811145	2.341711	0.77	0.439	-2.783692	6.405983

- Oceń, które zmienne są istotne.
- Czy zmienne objaśniające są łącznie istotne?
- Zinterpretować wartość współczynnika determinacji.
- Jaka zależność między wiekiem respondenta a długością trwania bezrobocia wynika z modelu? Osoby w jakim wieku przebywają najdłużej na bezrobociu?
- Wyznaczyć efekt cząstkowy dla wieku. Jaka jest oczekiwana zmiana długości trwania bezrobocia przy wzroście wieku o 1 rok dla trzydziestoletniego respondenta?
- Dokonać interpretacji oszacowań parametrów przy zmiennych, nawet wtedy gdy zmienne są nieistotne:
  - srednie*
  - sredniexang*

7. Postanowiono wprowadzić do analizowanej regresji zmienną *edu*, która oznacza liczbę lat poświęconych na naukę. Jaki problem pojawi się w zmodyfikowanym modelu? Odpowiedź uzasadnić.
8. Niech zmienna *edu\_1* przyjmuje wartość 1 dla osób, które mają wykształcenie średnie lub wyższe. Jeśli zmienną tę wprowadzimy do modelu to jaki pojawi się problem? Odpowiedź uzasadnić.
9. Przypuszcza się, że wpływ wieku na długość trwania bezrobocia zależy od płci respondenta. Podać postać modelu, który będzie odpowiadał temu założeniu.

### Rozwiązanie Zadanie 1

1. Istotne zmienne, to te dla których *p – value* jest mniejsze od przyjętego poziomu istotności wynoszącego 0,05. Czyli istotne zmienne to:

*wiek* ( $t = 5,21, p - value = 0,000$ )  
*wiek\_2* ( $t = -4,66, p - value = 0,000$ )  
*srednie* ( $t = -2,52, p - value = 0,012$ )  
*wyzsze* ( $t = -3,56, p - value = 0,000$ )  
*plec* ( $t = 4,17, p - value = 0,000$ )

2. Test na łączną istotność regresji:  $F = 12,18, p - value = 0,000 < 0,05$ , odrzucamy hipotezę zerową o łącznej nieistotności regresji.
3. 8.3% zmienności długości trwania bezrobocia zostało wyjaśnionych za pomocą zmiennych niezależnych.
4. Zależność między długością trwania bezrobocia a wiekiem jest kwadratowa:

$$\widehat{\text{bezrobocie}} = -0,0074501\text{wiek}^2 + 0,6579234\text{wiek}$$

Jest to parabola z ramionami skierowanymi do dołu, czyli wraz ze wzrostem wieku długość trwania bezrobocia rośnie, ale coraz wolniej. Maksimum funkcji jest osiągnięte dla osoby w wieku 44 lat

$$\frac{-b}{2a} = \frac{-0,6579234}{2 * (-0,0074501)} \approx 44,15$$

Dla osób powyżej 44 lat zależność między wiekiem a długością trwania bezrobocia jest ujemna – długość trwania bezrobocia maleje coraz szybciej wraz ze wzrostem wieku.

5. Postać analizowanego modelu:

$$\begin{aligned} \text{bezrobocie}_i &= \beta_0 + \beta_1 \text{wiek}_i + \beta_2 \text{wiek}_i^2 + \beta_3 \text{srednie}_i + \beta_4 \text{wyzsze}_i \\ &+ \beta_5 \text{angielski}_i + \beta_6 \text{srednie} \times \text{ang}_i + \beta_7 \text{wyzsze} \times \text{ang}_i + \beta_8 \text{plec}_i \\ &+ \varepsilon_i \end{aligned}$$

Efekt cząstkowy dla wieku:

$$\frac{\partial E(\text{bezrobocie}_i)}{\partial \text{wiek}_i} = \beta_1 + \beta_2 \text{wiek}_i$$

Oszacowanie efektu cząstkowego na podstawie modelu dla osoby trzydziestoletniej wynosi:

$\beta_1 + 2 \beta_2 \text{wiek}_i = 0,6579234 + 2 \cdot (0,0074501) \cdot 30 \approx 0,21$  - oczekiwany wzrost długości trwania bezrobocia przy wzroście wieku o 1 rok dla trzydziestoletniego respondenta wynosi 0,21 miesiąca.

6. a) Osoby z wykształceniem średnim, które nie znają języka angielskiego w porównaniu z osobami z wykształceniem podstawowym nieznającymi języka angielskiego o 1,55 miesiąca krótszy czas przebywania na bezrobociu.
- b) Osoby z wykształceniem średnim, które znają język angielski w porównaniu z osobami z wykształceniem średnim nieznającymi języka angielskiego mają o 0,41 miesiąca krótszy czas przebywania na bezrobociu ( $-2,578042 + 2,167486 = -0,410556$ ).
7. Zmienna *edu* oznaczająca wykształcenie mierzone za pomocą liczby lat nauki będzie silnie skorelowana ze zmiennymi zerojedynkowymi dotyczącymi poziomu osiągniętego wykształcenia, więc wystąpi (najprawdopodobniej) problem współliniowości. Zmienna *edu* nie wnosi nic nowego, więc nie ma sensu wprowadzać jej do regresji.
8. Zmienna *edu\_1* przyjmuje wartość 1 dla osób, które mają wykształcenie średnie lub wyższe, więc będzie dokładnie współliniowa ze zmiennymi zerojedynkowymi *srednie* i *wyzsze*. Dla każdej obserwacji w próbie będzie zachodzić:

$$\text{edu}_1_i = \text{srednie}_i + \text{wyzsze}_i.$$

Jeżeli jedna z kolumn macierzy X jest kombinacją liniową pozostałych, to macierz  $X'X$  nie ma pełnego rzędu kolumnowego, więc nie istnieje macierz odwrotna do  $X'X$  - nie można wyznaczyć estymatora MNK dla takiego modelu. Należy usunąć jedną ze zmiennych wywołujących problem dokładnej współliniowości.

9. Należy wprowadzić do modelu interakcje między zmiennymi dotyczącymi wieku a zmienna płeć:

*bezrobocie<sub>i</sub>*

$$\begin{aligned} &= \beta_0 + \beta_1 \text{wiek}_i + \beta_2 \text{wiek}_i^2 + \beta_3 \text{plec}_i \text{X} \text{wiek}_i + \beta_4 \text{plec}_i \text{X} \text{wiek}_i^2 + \beta_5 \text{srednie}_i \\ &+ \beta_6 \text{wyzsze}_i + \beta_7 \text{angielski}_i + \beta_8 \text{srednieXang}_i + \beta_9 \text{wyzszeXang}_i \\ &+ \beta_{10} \text{plec}_i + \varepsilon_i \end{aligned}$$

## Zadanie 2

Udzielić odpowiedzi odnośnie następujących problemów:

### Problem 1

Badacz dysponując pewnym zbiorem danych wykorzystał program Stata w celu zbadania występowania pewnego problemu. Badacz stwierdził, że problem ten występuje, a następnie podjął strategię w celu rozwiązania tego problemu. Poniżej znajdują się wyniki poszczególnych działań badacza.

Przyjęty przez badacza poziom istotności to 0,05.

**reg y x**

Source	SS	df	MS			
Model	387.15257	1	387.15257	Number of obs =	100	
Residual	185.17243	98	1.88951459	F( 1, 98) =	204.90	
Total	572.325	99	5.78106061	Prob > F =	0.0000	
				R-squared =	0.6765	
				Adj R-squared =	0.6732	
				Root MSE =	1.3746	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	1.977532	.1381521	14.31	0.000	1.703373	2.25169
_cons	1.94407	.1374596	14.14	0.000	1.671286	2.216854

**estat hettest**

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

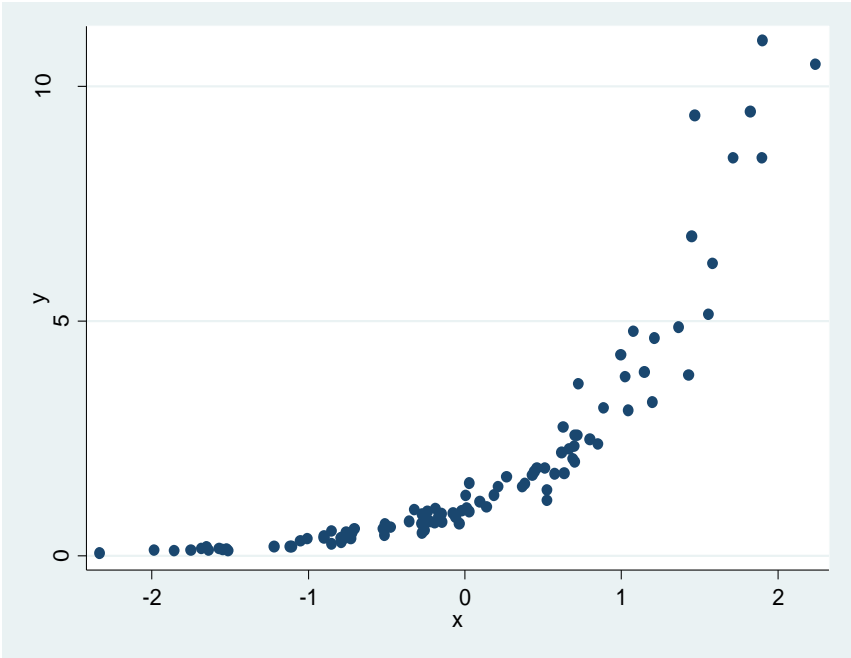
Ho: Constant variance

Variables: fitted values of y

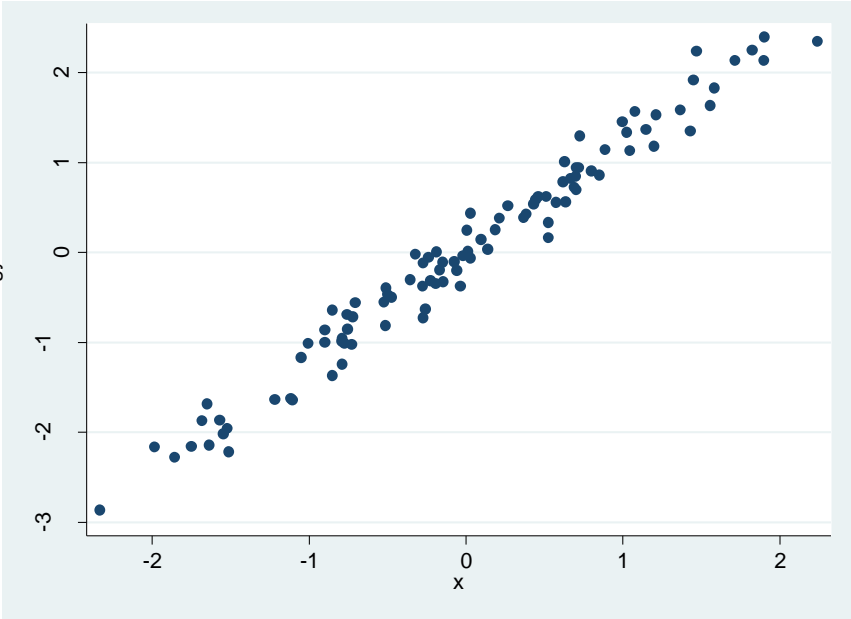
chi2(1) = 32.27

Prob > chi2 = 0.0000

Wykres rozrzutu pomiędzy y a x



Wykres rozrzutu pomiędzy logarytmem y a x



**reg logy x**

Source	SS	df	MS			
Model	142.56	1	142.56	Number of obs =	100	
Residual	3.96000021	98	.040408165	F( 1, 98) =	3528.00	
				Prob > F =	0.0000	
				R-squared =	0.9730	
				Adj R-squared =	0.9727	
				Root MSE =	.20102	
Total	146.520001	99	1.48000001			

logy	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	1.2	.0202031	59.40	0.000	1.159908	1.240092
_cons	2.14e-09	.0201018	0.00	1.000	-.0398913	.0398913

**estat hettest**

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of logy

chi2(1) = 0.02

Prob > chi2 = 0.8781

1. Wyjaśnić jaki problem badacz testował i na jakiej podstawie stwierdził, że on występuje.
2. Uzasadnić rozwiązanie powyższego problemu, które badacz wybrał.
3. Zaproponować alternatywne rozwiązanie problemu, które nie zostało wykorzystane przez badacza.



## Problem 2

Badacz przeprowadził regresję dochodu (*doch*) na liczbie lat nauki (*nauka*) uzyskując poniższe wyniki i wykres.

**reg doch nauka**

Source	SS	df	MS			
Model	253767.956	1	253767.956	Number of obs =	500	
Residual	31421.5231	498	63.0954279	F( 1, 498) =	4021.97	
				Prob > F =	0.0000	
				R-squared =	0.8898	
				Adj R-squared =	0.8896	
				Root MSE =	7.9433	
Total	285189.479	499	571.522002			

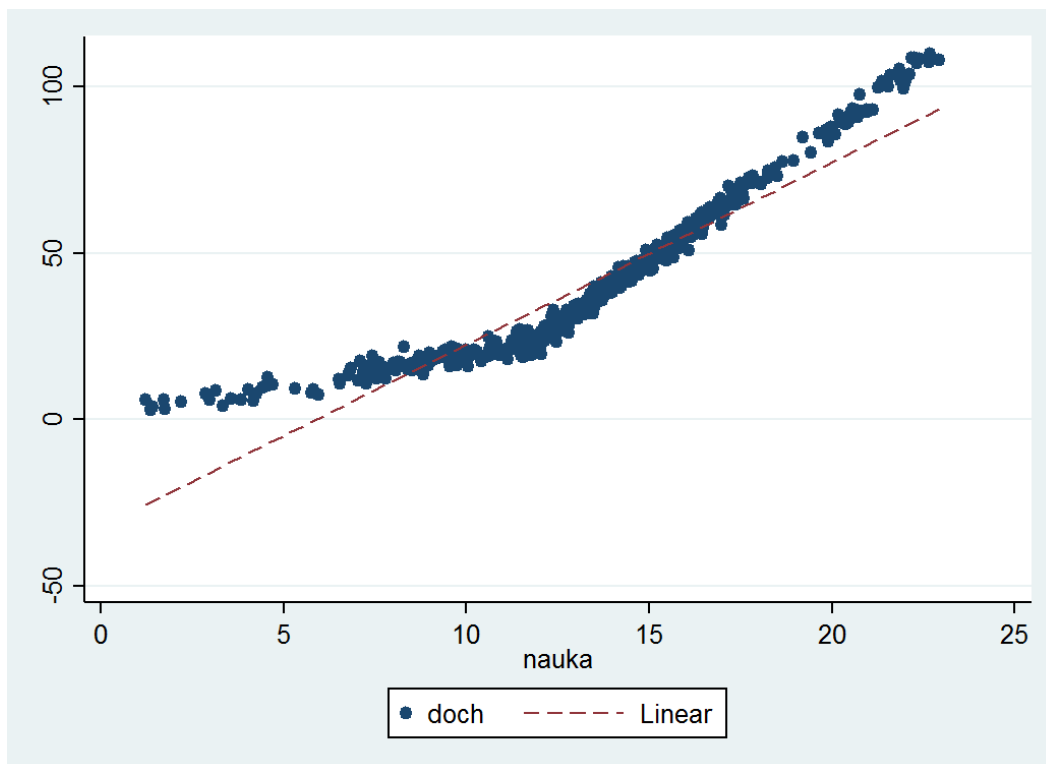
  

doch	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
nauka	5.472857	.0862968	63.42	0.000	5.303306	5.642407
_cons	-32.12951	1.194932	-26.89	0.000	-34.47724	-29.78178

**predict Linear**

(option **xb** assumed; fitted values)

## Wykres



Następnie badacz utworzył nowe zmienne:

- a) *nauka\_nizsze* przyjmująca wartości zmiennej *nauka* jeśli liczba lat nauki była niższa od 12 lat bądź im równa i wartość 12 jeśli liczba lat nauki była większa od 12 lat;
- b) *nauka\_wyzsze* przyjmująca wartości zmiennej 0 jeśli liczba lat nauki była niższa od 12 lat bądź im równa i wartość będąca różnicą między liczbą lat nauki a 12 jeśli liczba lat nauki była większa od 12 lat;

i uzyskał poniższe wyniki i wykres.

```
reg doch nauka_nizsze nauka_wyzsze
```

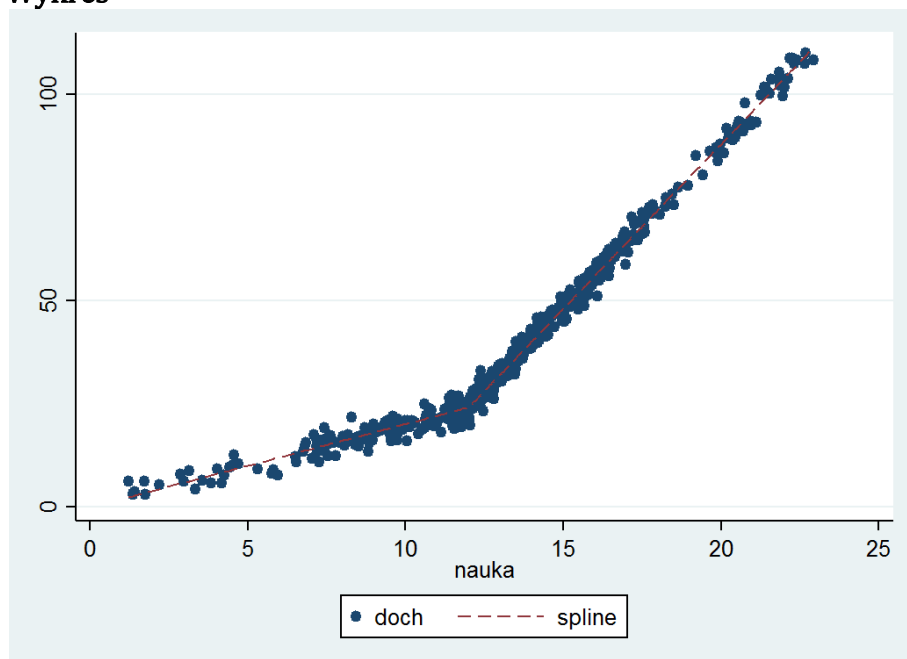
Source	SS	df	MS			
Model	283338.491	2	141669.245	Number of obs =	500	
Residual	1850.98836	497	3.72432266	F( 2, 497) =	38038.93	
				Prob > F =	0.0000	
				R-squared =	0.9935	
				Adj R-squared =	0.9935	
				Root MSE =	1.9299	
Total	285189.479	499	571.522002			

doch	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
nauka_nizsze	2.015082	.044107	45.69	0.000	1.928423	2.101742
nauka_wyzsze	7.976738	.0350599	227.52	0.000	7.907854	8.045622
_cons	-.0942312	.4621001	-0.20	0.838	-1.002142	.8136793

```
predict spline  
(option xb assumed; fitted values)
```

## Wykres



4. Wyjaśnić jaki jest cel utworzenia przez badacza nowych zmiennych i przeprowadzenia regresji dochodu na tych zmiennych. Dlaczego badacz po uzyskaniu drugiego wykresu nie podjął dalszych działań?

### Problem 3

Badacz dysponując pewnym zbiorem danych wykorzystał program Stata w celu zbadania występowania pewnego problemu. Badacz stwierdził, że problem ten występuje, a następnie podjął strategię w celu rozwiązania tego problemu. Poniżej znajdują się wyniki poszczególnych działań badacza.

Przyjęty przez badacza poziom istotności to 0,05.

**Uwaga:**  $x^2$  to kwadrat zmiennej  $x$ .  $x^3$  to sześcian zmiennej  $x$ .

**reg y x**

Source	SS	df	MS			
Model	5.9161e+13	1	5.9161e+13	Number of obs =	2293	
Residual	2.9047e+13	2291	1.2679e+10	F( 1, 2291) =	4666.20	
Total	8.8208e+13	2292	3.8485e+10	Prob > F =	0.0000	
				R-squared =	0.6707	
				Adj R-squared =	0.6706	
				Root MSE =	1.1e+05	

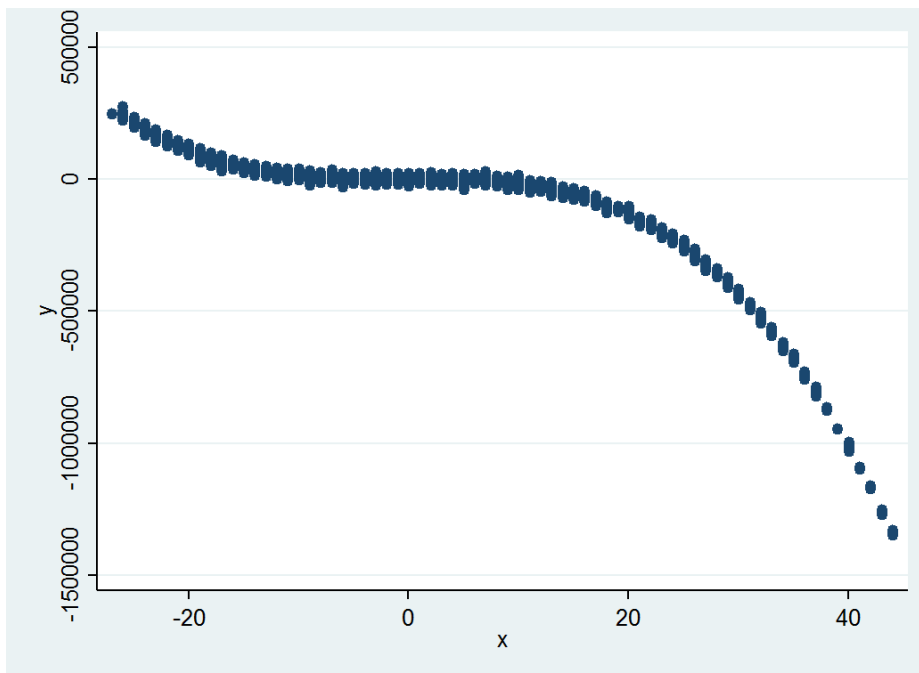
  

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	-9575.092	140.1721	-68.31	0.000	-9849.97	-9300.215
_cons	-37867.18	2351.439	-16.10	0.000	-42478.35	-33256.01

**ovtest**

Ramsey RESET test using powers of the fitted values of y  
Ho: model has no omitted variables  
F(3, 2288) = 96238.53  
Prob > F = 0.0000

**Wykres rozrzutu pomiędzy y a x**



**reg y x x^2 x^3**

Source	SS	df	MS			
Model	8.7979e+13	3	2.9326e+13	Number of obs =	2293	
Residual	2.2843e+11	2289	99793436.9	F( 3, 2289) =	.	
Total	8.8208e+13	2292	3.8485e+10	Prob > F =	0.0000	
				R-squared =	0.9974	
				Adj R-squared =	0.9974	
				Root MSE =	9989.7	

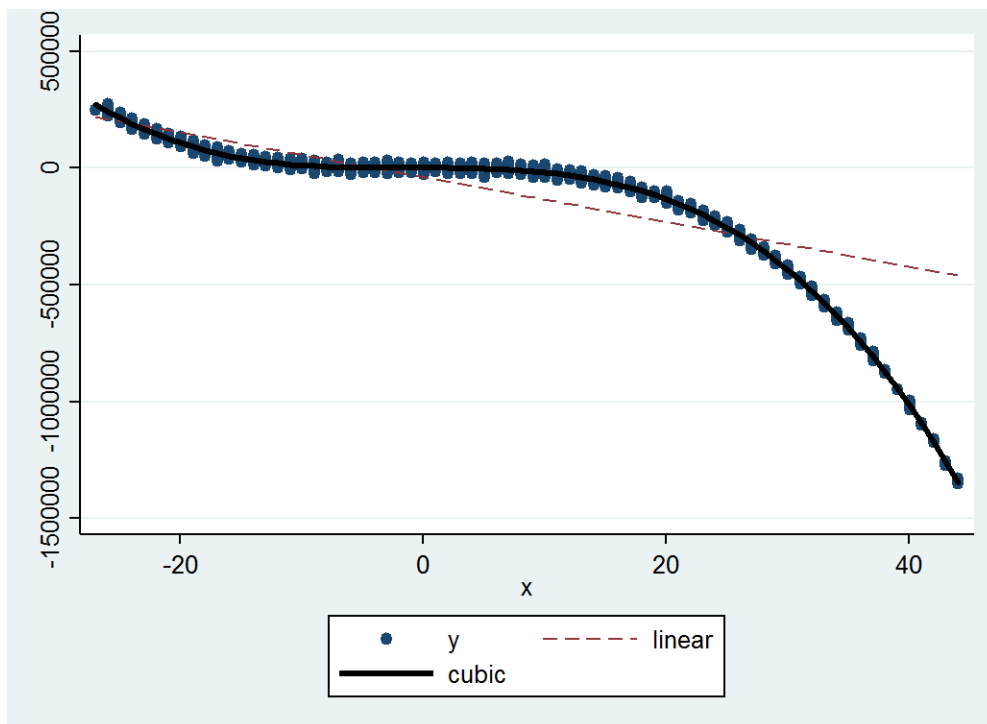
  

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x	-5.618982	24.07808	-0.23	0.815	-52.83611	41.59815
x^2	-30.70687	.9647104	-31.83	0.000	-32.59867	-28.81507
x^3	-15.02577	.0391987	-383.32	0.000	-15.10264	-14.9489
_cons	249.4734	309.3394	0.81	0.420	-357.1414	856.0882

**ovtest**

Ramsey RESET test using powers of the fitted values of y  
 Ho: model has no omitted variables  
 F(3, 2286) = 0.14  
 Prob > F = 0.9390

**Wykres**



5. Wyjaśnić jaki problem badacz testował i na jakiej podstawie stwierdził, że on występuje.
6. Uzasadnić rozwiązanie powyższego problemu, które badacz wybrał.

## Rozwiązanie Zadanie 2

1. Na podstawie testu Breuscha-Pagana ( $chi(2) = 32,27, p - value = 0,000 < 0,05$ ) badacz odrzucił hipotezę zerową o homoskedastyczności. Również na podstawie wykresu rozrzutu pomiędzy  $y$  a  $x$  badacz stwierdza, iż zależność między  $y$  a  $x$  nie jest liniowa, co implikuje nie spełnienie założenia dot. formy funkcyjnej modelu.
2. W celu poprawienia formy funkcyjnej w modelu badacz zlogarytmował zmienną  $y$  a następnie przeprowadził regresję  $\log y$  na  $x$ . Na podstawie wykresu rozrzutu pomiędzy  $\log y$  a  $x$  badacz stwierdza, że związek jest już bardziej liniowy. W przypadku testu Breuscha-Pagana brak jest podstaw do odrzucenia hipotezy zerowej o homoskedastyczności ( $chi(2) = 0,02, p - value = 0,8781 > 0,05$ ).
3. Badacz oczywiście mógł zastosować Uogólnioną Metodę Najmniejszy Kwadratów (UMNK) lub odporne estymatory macierzy wariancji kowariancji  $b$ . Natomiast badacz nie użył powyższych metod, bo problemem nie była heteroskedastyczność, ale błędna specyfikacja modelu.
4. Badacz wprowadził nowe zmienne i przeprowadził na nich regresję dochodu aby przybliżyć nieliniową zależność między dochodem a liczbą lat nauki, (którą zauważył na podstawie pierwszego wykresu w tym problemie) za pomocą zależności liniowej, a konkretnie krzywej łamanej. Badacz zakończył działania bo wykres drugi w tym problemie wskazuje, iż udało mu się przybliżyć tą nieliniową zależność za pomocą krzywej łamanej i wobec tego nie ma konieczności podejmowania dalszych działań.
5. Badacz testował poprawność formy funkcyjnej modelu. Na podstawie testu RESET ( $F = 96238,53, p - value = 0,0000 < 0,05$ ) badacz odrzucił hipotezę zerową o poprawnej formie funkcyjnej modelu. Również na podstawie wykresu rozrzutu pomiędzy  $y$  a  $x$  badacz stwierdza, że forma funkcyjna modelu nie jest poprawna: zależność między  $y$  a  $x$  nie jest liniowa.
6. W celu poprawy formy funkcyjnej modelu badacz uwzględnił w regresji zmienną  $x$  podniesioną do kwadratu i do sześciannu. W przypadku testu RESET ( $F = 0,14, p - value = 0,9390 < 0,05$ ) brak jest podstaw do odrzucenia hipotezy zerowej o poprawnej formie funkcyjnej modelu z uwzględnieniem kwadratu i sześciannu  $x$ . Również na podstawie wykresu drugiego w tym problemie badacz stwierdza, że forma funkcyjna modelu z uwzględnieniem kwadratu i sześciannu  $x$  jest poprawna.

### Zadanie 3

Na podstawie pewnej bazy danych szacowano czynniki wpływające na odczucia rodzica dotyczące szczepień ich dzieci (*szczepionka*, mierzone w skali od 0 do 100). Zmiennymi niezależnymi są rasa rodzica (*rasa*: 1- biała, 0 - inna); przekonania rodzica co do tego, że szczepienia powodują autyzm (*autyzm*, mierzone w skali od 0 do 100); wpływ lekarza na decyzje rodzica odnośnie szczepienia (*wplyw*, mierzony w skali od 0 do 100). Hipotezy testować na poziomie istotności 0,05. Odpowiedzi uzasadnić podając p-value.

```
sum szczepionka rasa autyzm wplyw
```

Variable	Obs	Mean	Std. Dev.	Min	Max
szczepionka	2000	47.3155	26.8503	1	100
rasa	2000	.885	.3191017	0	1
autyzm	2000	27.4605	17.03163	0	71
wplyw	2000	60.605	15.83222	0	100

### Regresja I

```
reg szczepionka i.rasa autyzm wplyw, vce(robust)
```

Linear regression

Number of obs = 2000  
F( 3, 1996) = 47.22  
Prob > F = 0.0000  
R-squared = 0.0648  
Root MSE = 25.985

szc	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
rasa	-4.363207	1.798159	-2.43	0.015	-7.889673	-.8367423
autyzm	-.2537355	.0352135	-7.21	0.000	-.3227946	-.1846764
wplyw	.2573966	.0386828	6.65	0.000	.1815336	.3332596
_cons	42.54512	3.013613	14.12	0.000	36.63497	48.45528

## Regresja II

reg szczepionka i.rasa autyzm wplyw

Source	SS	df	MS			
Model	93383.6877	3	31127.8959	Number of obs =	2000	
Residual	1347772.23	1996	675.236589	F( 3, 1996) =	46.10	
Total	1441155.92	1999	720.938429	Prob > F =	0.0000	
				R-squared =	0.0648	
				Adj R-squared =	0.0634	
				Root MSE =	25.985	

szczepionka	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rasa	-4.363207	1.838179	-2.37	0.018	-7.968157	-.7582576
autyzm	-.2537355	.0357901	-7.09	0.000	-.3239253	-.1835457
wplyw	.2573966	.0387209	6.65	0.000	.1814591	.3333342
_cons	42.54512	3.099787	13.73	0.000	36.46596	48.62428

## Regresja III

reg szczepionka i.rasa autyzm wplyw i.rasa\*i.rasa

Source	SS	df	MS			
Model	93383.6877	3	31127.8959	Number of obs =	2000	
Residual	1347772.23	1996	675.236589	F( 3, 1996) =	46.10	
Total	1441155.92	1999	720.938429	Prob > F =	0.0000	
				R-squared =	0.0648	
				Adj R-squared =	0.0634	
				Root MSE =	25.985	

szczepionka	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
rasa	-4.363207	1.838179	-2.37	0.018	-7.968157	-.7582576
autyzm	-.2537355	.0357901	-7.09	0.000	-.3239253	-.1835457
wplyw	.2573966	.0387209	6.65	0.000	.1814591	.3333342
_cons	42.54512	3.099787	13.73	0.000	36.46596	48.62428



### testparm i.rasa autyzm wplyw

- ( 1) 1.rasa = 0
- ( 2) autyzm = 0
- ( 3) wplyw = 0

F( 3, 1996) = 46.10  
Prob > F = 0.0000

### test wplyw = -autyzm

- ( 1) autyzm + wplyw = 0

F( 1, 1996) = 0.00  
Prob > F = 0.9514

1. Przeprowadzono regresję z VCE (Regresja I), a następnie ponownie bez VCE (Regresja II). Zauważono, że współczynniki nie zmieniły się wobec tego postanowiono nie korzystać z VCE. Czy uważasz, że taka decyzja została podjęta słusznie?  
Uwaga: VCE to odporny estymator White'a macierzy wariancji kowariancji b
2. Niektórzy badacze uważają, że przekonania rodzica co do autyzmu mają największy wpływ na ich odczucia odnośnie szczepień. Inni badacze mówią, że to wpływ lekarza jest najważniejszy. Jeszcze inni twierdzą, że obie też zmienne są równie ważne i że różnice pomiędzy nimi nie istnieją.  
Korzystając z wydruków wskazać, które z powyższych podejść jest prawdziwe.  
Odpowiedź uzasadnić.
3. Pewna osoba została poinformowana, że ważne jest, aby dodać do regresji zmienną podniesioną do kwadratu. Stąd dodała ona zmienną *rasa* podniesioną do kwadratu do regresji (Regresja III). Ku zaskoczeniu tej osoby żaden z wyników nie zmienił się. Wyjaśnić dlaczego żaden z wyników się nie zmienił.

### Rozwiązanie Zadanie 3

1. Taka decyzja została podjęta niesłusznie. W wyniku użycia odpornego estymatora White'a macierzy wariancji kowariancji b oszacowania estymatorów nie zmieniają się. Natomiast uzyskujemy w wyniku użycia VCE odporne na problem związany z heteroskedastycznością odchylenia standardowe estymatorów.
2. Prawidłowe jest podejście stwierdzające, iż zarówno przekonania rodzica co do autyzmu jak i wpływ lekarza są równie ważne a różnice między nimi nie istnieją. Wskazuje na to wyniki testu badającego czy współczynniki przy zmiennej *wplyw* jest równy współczynnikowi z minusem przy zmiennej *autyzm*. Statystyka:  $F = 0,00, p - value = 0,9514 < 0,05$  czyli brak jest podstaw do odrzucenia

hipotezy zerowej mówiącej, że te dwie są równie ważne a różnice między nimi nie istnieją.

3. W przypadku dodania do regresji zmiennej *rasa* podniesionej do kwadratu wystąpi problem dokładnej współliniowości: zmienna *rasa* podniesiona do kwadratu przyjmuje dokładnie takie same wartości jak zmienna *rasa*. W takim przypadku jedna z kolumn macierzy  $X$  jest kombinacją liniową pozostałych a macierz  $X'X$  nie ma pełnego rzędu kolumnowego, więc nie istnieje macierz odwrotna do  $X'X$  - nie można wyznaczyć estymatora MNK dla takiego modelu. Stąd należy usunąć jedną ze zmiennych wywołujących problem dokładnej współliniowości czyli zmienną *rasa* podniesioną do kwadratu.

Więc kwadrat zmiennej 0-1 nie ma sensu!