Egzamin z ekonometrii – wersja IiE, MSEMAT
08-02-2017

Pytania teoretyczne

1. W jaki sposób przeprowadzamy test Chowa?
2. Pokazać, że s^2 jest nieobciążonym estymatorem σ^2.
3. Udowodnić, że w modelu ze stałą TSS=ESS+RSS.
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              Number of obs =    1085
-------------+------------------------------           F(  8,  1076) =   12.18
       Model | 5346.40487      8  668.300609           Prob > F      =  0.0000
    Residual | 59033.7979   1076  54.8641244           R-squared     =  0.0830
-------------+------------------------------           Adj R-squared =  0.0762
       Total | 64380.2028   1084  59.391331            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
 wyzszexang  |  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
------------------------------------------------------------------------------
```

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

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

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 miedzy 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 osiągane dla osoby w wieku 44 lat

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

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

5. Postać analizowanego modelu:

$$bezrobocie_i = \beta_0 + \beta_1 wiek_i + \beta_2\, wiek_i^2 + \beta_3 srednie_i + \beta_4 wyzsze_i + \beta_5 angielski_i + \beta_6 srednieXang_i + \beta_7 wyzszeXang_i + \beta_8 plec_i + \varepsilon_i$$

Efekt cząstkowy dla wieku:

$$\frac{\partial \mathrm{E}(\mathrm{bezrobocie_i})}{\partial \mathrm{wiek_i}} = \beta_1 + \beta_2 \mathrm{wiek_i}$$

Oszacowanie efektu cząstkowego na podstawie modelu dla osoby trzydziestoletniej wynosi:

$b_1 + 2\, b_2\, \mathrm{wiek_i} = 0{,}6579234 + 2 \cdot (0{,}0074501) \cdot 30 \approx 0{,}21$ - 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 jeżyka 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 maja 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, wiec nie ma sensu wprowadzać jej do regresji.

8. Zmienna *edu_1* przyjmuje wartość 1 dla osób, które maja wykształcenie średnie lub wyższe, więc będzie dokładnie współliniowa ze zmiennymi zerojedynkowymi *srednie* i *wyzsze*. Dla każdej obserwacji w próbie będzie zachodzić:

$$\mathrm{edu\_1_i} = \mathrm{srednie_i} + \mathrm{wyzsze_i}.$$

   Jeżeli jedna z kolumn macierzy X jest kombinacją liniową pozostałych, to macierz X' X nie ma pełnego rzędu kolumnowego, wiec 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 miedzy zmiennymi dotyczącymi wieku a zmienna płeć:

$$bezrobocie_i = \beta_0 + \beta_1 wiek_i + \beta_2\, wiek_i^2 + \beta_3 plec_i X wiek_i + \beta_4\, plec_i X wiek_i^2 + \beta_5 srednie_i + \beta_6 wyzsze_i + \beta_7 angielski_i + \beta_8 srednie X ang_i + \beta_9 wyzsze X ang_i + \beta_{10} plec_i + \varepsilon_i$$

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.

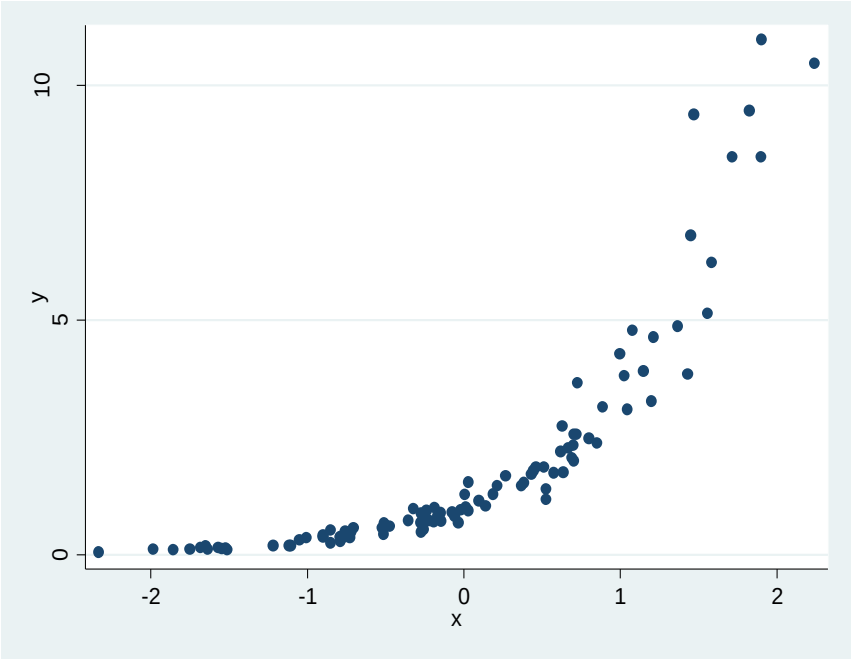
reg y x

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

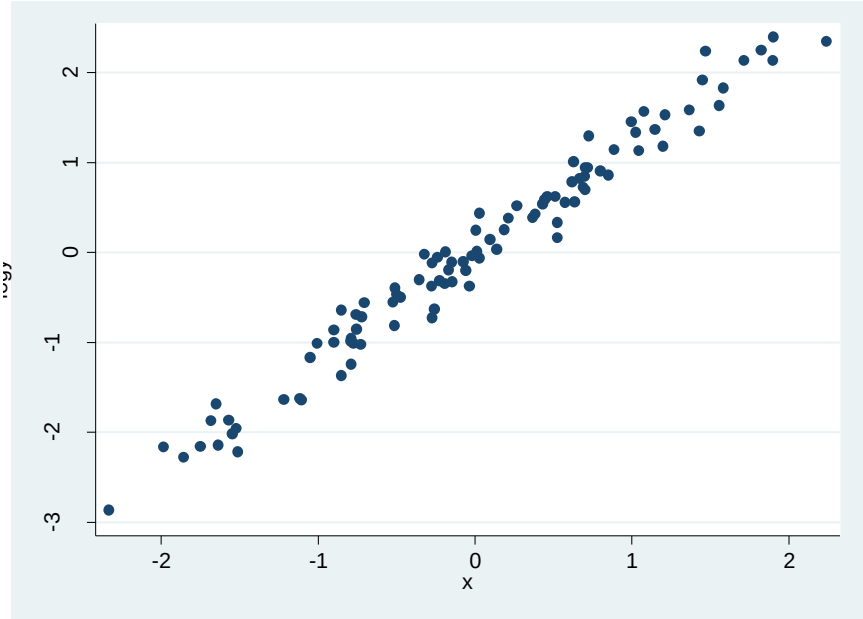
estat hettest

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

------------------------------------------------------------------------------
        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

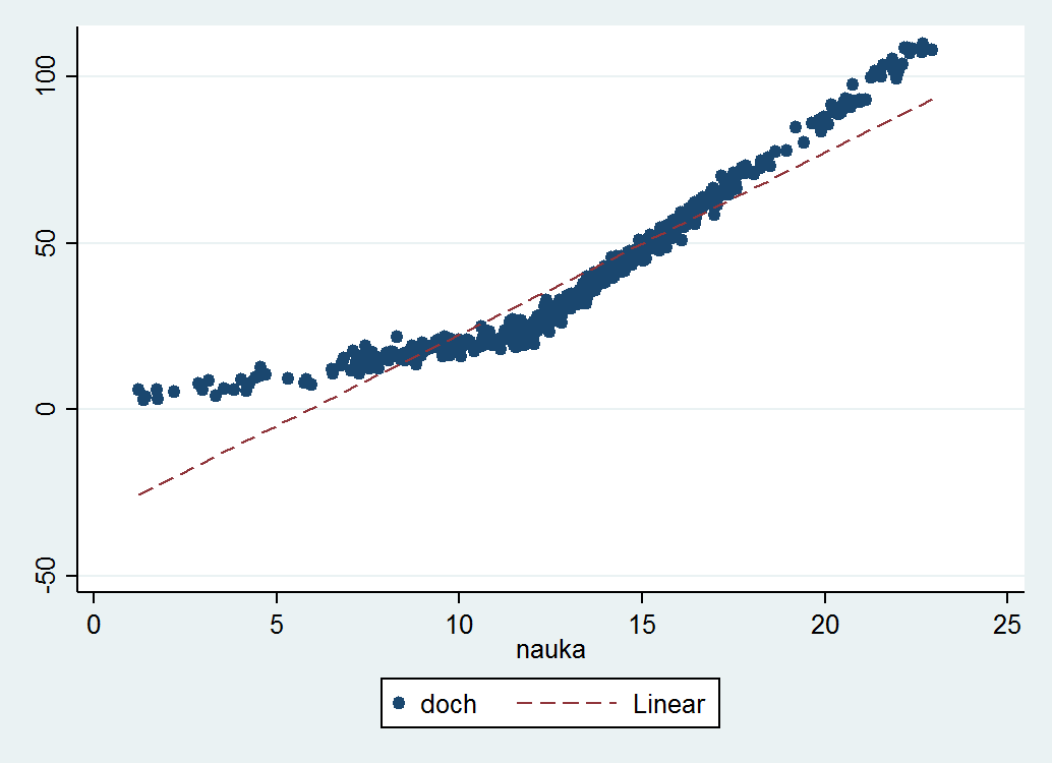
Badacz przeprowadził regresję dochodu (*doch)* na liczbie lat nauki (*nauka*) uzyskując poniższe wyniki i wykres.

reg doch nauka

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

------------------------------------------------------------------------------
        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;

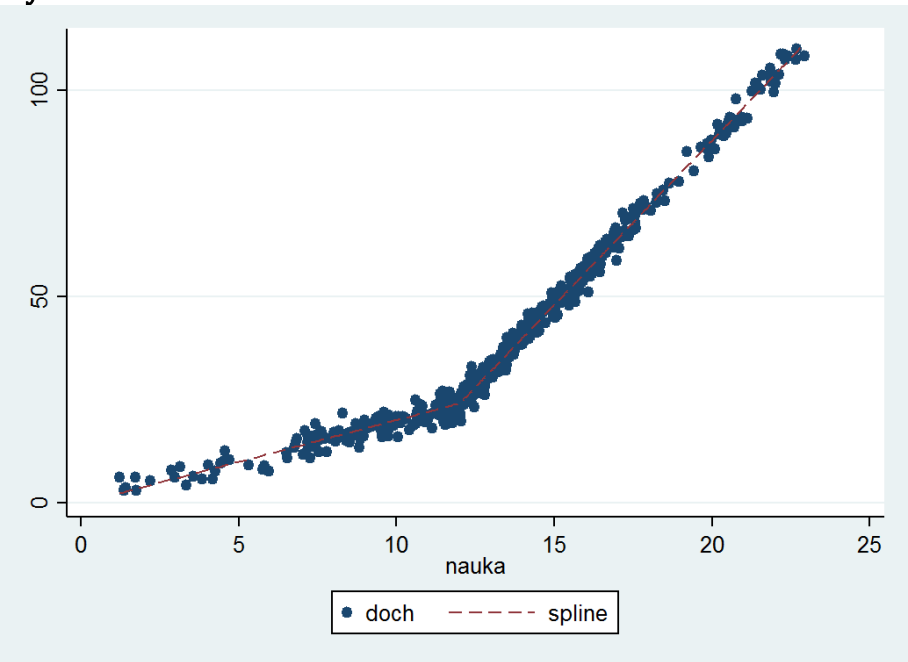
i uzyskał poniższe wyniki i wykres.

reg doch nauka_nizsze nauka_wyzsze

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

------------------------------------------------------------------------------
        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. Wyjaśniać 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.

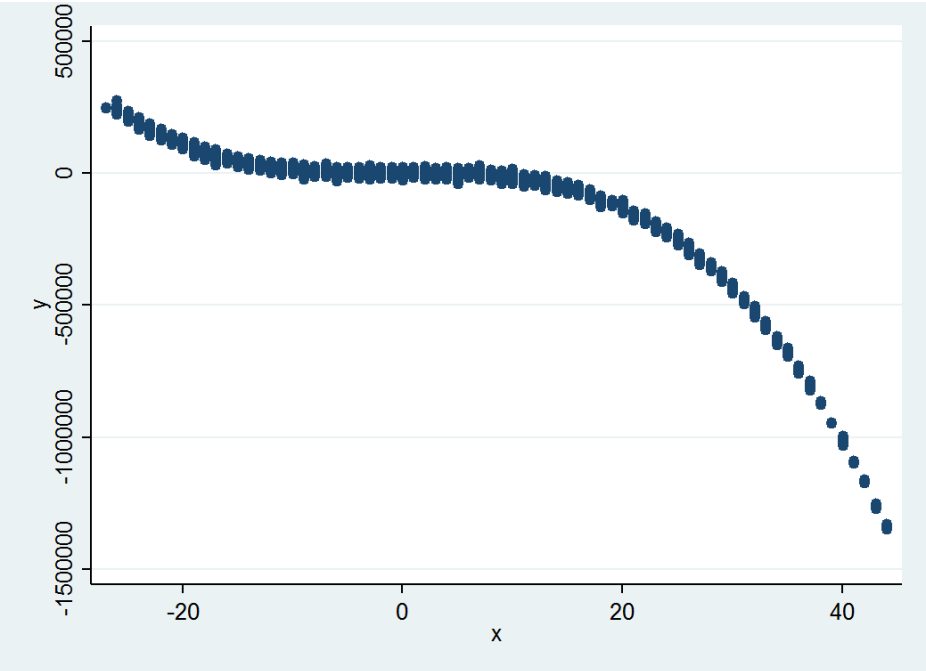
reg y x

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



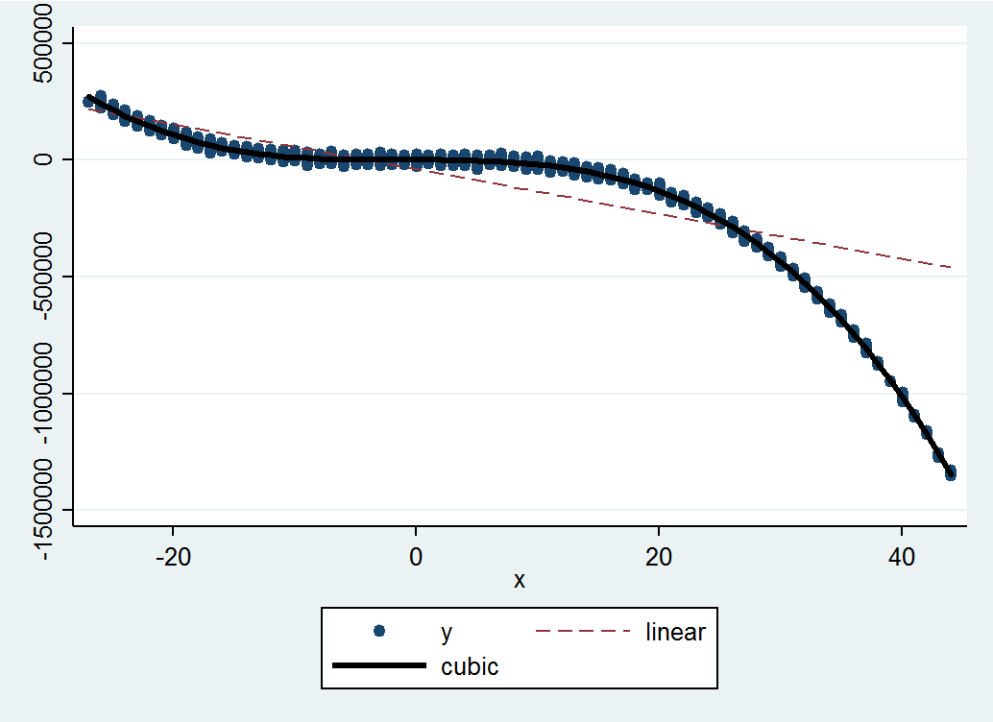
reg y x x^2 x^3

```
      Source |       SS       df       MS              Number of obs =    2293
-------------+------------------------------           F(  3,  2289) =       .
       Model |  8.7979e+13     3  2.9326e+13           Prob > F      =  0.0000
    Residual |  2.2843e+11  2289  99793436.9           R-squared     =  0.9974
-------------+------------------------------           Adj R-squared =  0.9974
       Total |  8.8208e+13  2292  3.8485e+10           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 sześcianu. 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 sześcianu x. Również na podstawie wykresu drugiego w tym problemie badacz stwierdza, że forma funkcyjna modelu z uwzględnieniem kwadratu i sześcianu x jest poprawna.

Zadanie 3

Dany jest model regresji liniowej ze stałą i jedną zmienną:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ (1) \quad i = 1, \dots, N$$

dla którego spełnione są założenia KMRL. Niech $\hat{\beta}_1$ będzie nieobciążonym estymatorem β_1 w modelu (1).
Niech $\tilde{\beta}_1$ będzie estymatorem β_1 uzyskanym przy założeniu, że stała w modelu (1) jest równa zero.

1. Wyprowadzić estymator $\tilde{\beta}_1$.
2. Pokazać, że $\tilde{\beta}_1$ jest nieobciążonym estymatorem β_1 w przypadku, gdy stała w modelu (1) jest równa zero. Czy istnieją inne przypadki kiedy estymator $\tilde{\beta}_1$ jest nieobciążony?
3. Wyprowadzić wariancję estymatora $\tilde{\beta}_1$.
4. Pokazać, że $Var(\tilde{\beta}_1) \leq Var(\hat{\beta}_1)$

   Podpowiedź: $\sum_{i=1}^{N} x_i^2 \geq \sum_{i=1}^{N} (x_i - \bar{x})^2$

Rozwiązanie Zadanie 3

1.

$$\tilde{\beta}_1 = \left(\sum_{i=1}^{N} x_i y_i\right) / \left(\sum_{i=1}^{N} {x_i}^2\right)$$

2.

$$\tilde{\beta}_1 = \frac{(\sum_{i=1}^{N} x_i y_i)}{(\sum_{i=1}^{N} {x_i}^2)} = \frac{(\sum_{i=1}^{N} x_i(\beta_0 + \beta_1 x_i + \varepsilon_i))}{(\sum_{i=1}^{N} {x_i}^2)} = \frac{(\beta_0 \sum_{i=1}^{N} x_i + \beta_1 \sum_{i=1}^{N} {x_i}^2 + \sum_{i=1}^{N} x_i \varepsilon_i)}{(\sum_{i=1}^{N} {x_i}^2)}$$
$$= \frac{\beta_0(\sum_{i=1}^{N} x_i)}{(\sum_{i=1}^{N} {x_i}^2)} + \beta_1 + \left(\sum_{i=1}^{N} x_i \varepsilon_i\right) / \left(\sum_{i=1}^{N} {x_i}^2\right)$$

$$E(\tilde{\beta}_1) = \frac{\beta_0(\sum_{i=1}^{N} x_i)}{(\sum_{i=1}^{N} {x_i}^2)} + \beta_1$$

Jeżeli stała w modelu (1) jest równa zero, to

$$E(\tilde{\beta}_1) = \beta_1$$

Czy istnieją inne przypadki kiedy $\tilde{\beta}_1$ jest nieobciążony?
Tak: jeżeli $(\sum_{i=1}^{N} x_i) = 0$

3.

$$Var(\tilde{\beta}_1) = Var\left[\frac{(\sum_{i=1}^{N} x_i y_i)}{(\sum_{i=1}^{N} {x_i}^2)}\right] = \frac{1}{(\sum_{i=1}^{N} {x_i}^2)^2} Var\left[\left(\sum_{i=1}^{N} x_i y_i\right)\right]$$
$$= \frac{1}{(\sum_{i=1}^{N} {x_i}^2)^2} \sum_{i=1}^{N} {x_i}^2 \, Var[y_i] = \frac{1}{(\sum_{i=1}^{N} {x_i}^2)^2} \sum_{i=1}^{N} {x_i}^2 \, Var[\beta_1 x_i + \varepsilon_i] =$$
$$= \frac{1}{(\sum_{i=1}^{N} {x_i}^2)^2} \sum_{i=1}^{N} {x_i}^2 \, Var[\varepsilon_i] = \frac{1}{(\sum_{i=1}^{N} {x_i}^2)^2} \sum_{i=1}^{N} {x_i}^2 \, \sigma^2 = \sigma^2 / \left(\sum_{i=1}^{N} {x_i}^2\right)$$

4.

$$Var(\tilde{\beta}_1) = \sigma^2 / \left(\sum_{i=1}^{N} {x_i}^2\right)$$

$$Var(\hat{\beta}_1) = \sigma^2 / \left(\sum_{i=1}^{N} (x_i - \bar{x})^2\right)$$

Korzystając, iż $\sum_{i=1}^{N} x_i^2 \geq \sum_{i=1}^{N} (x_i - \bar{x})^2$

$$Var(\tilde{\beta}_1) \leq Var(\hat{\beta}_1)$$