Egzamin z ekonometrii – wersja IiE, MSEMAT
04-02-2016

Pytania teoretyczne

1. Za pomocą jakiego testu weryfikowana jest normalność składnika losowego? Jakiemu założeniu KMRL odpowiada H_0 w tym teście? Jakie jest hipoteza alternatywna w tym teście? Jakie są konsekwencje dla własności MNK, jeśli H_0 jest fałszywe?
2. Udowodnić, że s^2 jest nieobciążonym estymatorem σ^2.
3. Opisać sposoby przybliżania zależności nieliniowej za pomocą modelu liniowego.
4. Dlaczego w modelu nie powinno się umieszczać stałej i wszystkich zmiennych zero-jedynkowych, związanych z poziomami zmiennej dyskretnej?

Zadanie 1

Badacz rozważa dwie specyfikacji regresji:

$$logY = \beta_1 + \beta_2 logX + u \quad (1)$$

$$log\frac{Y}{X} = \alpha_1 + \alpha_2 logX + u \quad (2)$$

gdzie: u – błąd losowy.

1. Określić, czy model (2) jest modelem (1) z ograniczeniami? Odpowiedź uzasadnić.
2. Badacz przyjął, iż $y = logY$, $x = logX$, $z = log\frac{Y}{X}$ i na podstawie danych zawierających N obserwacji oszacował dwa modele przy pomocy MNK uzyskując następujące wartości dopasowane:

   $$\hat{y} = b_1 + b_2 x \quad (3)$$

   oraz

   $$\hat{z} = \mathrm{a}_1 + \mathrm{a}_2 x \quad (4)$$

   Pokazać, że:

   (a) $b_2 = \mathrm{a}_2 + 1$

   (b) $b_1 = \mathrm{a}_1$

3. Sprawdzić, czy $\hat{y}_i - x_i = \hat{z}_i$?
4. Pokazać, że reszty regresji (3) są takie same jak w przypadku regresji (4).
5. Sprawdzić, czy błędy standardowe estymatorów b_2 i a_2 są takie same.
6. Określić relacje między statystyką t dla b_2 i statystyką t dla a_2.
7. Wyjaśnić, czy R^2 będzie takie samo dla regresji (3) i (4).
8. Pokazać, że suma dźwigni dla wszystkich obserwacji wynosi K, gdzie K oznacza liczbę szacowanych parametrów oraz, że wartość dźwigni jest zawsze większa od zera.

   Podpowiedź:
   Dźwignia $h_i = [P]_{ii}$, gdzie $P = X(X'X)^{-1}X'$

Rozwiązanie Zadanie 1

1. $logY = \alpha_1 + (\alpha_2 + 1)logX + u$

   Podsumowując,

   - $\beta_1 = \alpha_1$
   - $\beta_1 = \alpha_2 + 1$

2. (a) Na podstawie modelu (1):

$$logY = \beta_1 + \beta_2 logX + u$$

$$y = \beta_1 + \beta_2 x + u$$

$$b_2 = \frac{S_{yx}}{S_x^2} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Na podstawie modelu (2):

$$log\frac{Y}{X} = \alpha_1 + \alpha_2 logX + u$$

$$z = a_1 + a_2 x + u$$

$$\hat{z} = \mathrm{a}_1 + \mathrm{a}_2 x$$

$$\mathrm{a}_2 = \frac{S_{zx}}{S_x^2} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(z_i - \bar{z})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Wiemy, że $\mathrm{y} = \mathrm{logY}$, $\mathrm{x} = \mathrm{logX}$, $\mathrm{z} = \mathrm{log}\frac{\mathrm{Y}}{\mathrm{X}} = \mathrm{y} - \mathrm{x}$

$$\mathrm{a}_2 = \frac{S_{zx}}{S_x^2} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(z_i - \bar{z})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})\left([y_i - x_i] - \overline{[y - \bar{x}]}\right)}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} - \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = b_2 - 1$$

(b) $\mathrm{a}_1 = \bar{z} - \mathrm{a}_2\bar{x} = (\bar{y} - \bar{x}) - \mathrm{a}_2\bar{x} = \bar{y} - (\mathrm{a}_2 + 1)\bar{x} = \bar{y} - \mathrm{b}_2\bar{x} = \mathrm{b}_1$

3. $\hat{z}_i = \mathrm{a}_1 + \mathrm{a}_2 x_i = \mathrm{b}_1 + (\mathrm{b}_2 - 1)x_i = (\mathrm{b}_1 + \mathrm{b}_2 x_i) - x_i = \hat{y}_i - x_i$

4. Reszty $\mathrm{e_i} = \mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}$ dla regresji (3).
   Reszty $\mathrm{f_i} = \mathrm{z_i} - \hat{\mathrm{z}}_\mathrm{i}$ dla regresji (4).

   $f_i = z_i - \hat{z}_i = (y_i - x_i) - (\hat{y}_i - x_i) = y_i - \hat{y}_i = e_i$

5. $se(b_2) = \sqrt{\frac{\sum e_i^2/(N-2)}{\sum(x_i-\bar{x})^2}} = \sqrt{\frac{\sum f_i^2/(N-2)}{\sum(x_i-\bar{x})^2}} = se(\mathrm{a}_2)$

6. W celu przetestowania $H_0\colon \beta_2 = 0$

   Statystyka testowa:

$$t_{b_2} = \frac{b_2}{se(b_2)} = \frac{\mathrm{a}_2 + 1}{se(\mathrm{a}_2)}$$

   Wiemy, że $\beta_2 = \alpha_2 + 1$

   W takim razie H_0: $H_0\colon \alpha_2 = -1$

   Testy są równoważne, ponieważ powyższe dwa testy w przypadku braku podstaw do odrzucenia hipotezy zerowej ograniczają model do następującego modelu:

   logY w zależności tylko od stałej i zaburzenia losowego.

7. R^2 nie będą równe, ponieważ R^2 określa jaka część zmienności zmiennej objaśnianej jest wyjaśniona przez model. W dwóch modelach mamy różne zmienne zależne.
8. Korzystamy z definicji śladu jako sumy elementów diagonalnych macierzy oraz własności śladu:

$$\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{h_i} = \mathrm{tr}(\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}') = \mathrm{tr}\big((\mathrm{X}'\mathrm{X})^{-1}{}_{\mathrm{K}\times\mathrm{K}}(\mathrm{X}'\mathrm{X})_{\mathrm{K}\times\mathrm{K}}\big) = \mathrm{tr}(\mathrm{I}_{\mathrm{K}\times\mathrm{K}}) = \mathrm{K}$$

   Macierz P jest dodatnio określona, ponieważ macierz X’X jest dodatnio określone.

   W konsekwencji $(\mathrm{X}'\mathrm{X})^{-1}$ jest dodatnio określone.

Natomiast $\mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'$ musi być dodatnio określone, ponieważ jeśli $\mathrm{v}'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{v} > 0$ dla każdego $\mathrm{v} \neq 0$, to $\mathrm{vX}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{v} = \mathrm{v}^{*\prime}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{v}^{*} > 0$ dla $\mathrm{v}^{*} = \mathrm{vX} \neq 0$.

$$h_i = [P]_{ii} = \delta_i' P \delta_i > 0, skoro\ \delta_i \neq 0$$

Zadanie 2

Na podstawie pewnej bazy danych oszacowano wysokość rocznych dochodów w złotówkach (*ln_dochod* - logarytm rocznych dochodów). Zmiennymi niezależnymi są płeć (*plec*: 0 – mężczyzna, 1 – kobieta); wiek w latach(*wiek)*; wiek podniesiony do kwadratu (*wiek_2*).
Hipotezy testować na poziomie istotności 0,05. Odpowiedzi uzasadnić podając *p-value*.

```
Source    |       SS       df       MS                 Number of obs = 1595
----------+------------------------------              F( 3, 1591) = 303.00
Model     | 514.228436      3   171.409479               Prob > F = 0.0000
Residual  | 900.053861   1591   .565715815               R-squared = 0.3636
----------+------------------------------              Adj R-squared = 0.3624
Total     | 1414.2823    1594   .887253637               Root MSE = .75214
-------------------------------------------------------------------------------
ln_dochod |     Coef.     Std. Err.      t      P>|t|   [95% Conf. Interval]
----------+--------------------------------------------------------------------
   plec   | -.8483245     .0379257    -22.37    0.000   -.9227141   -.773935
   wiek   |  .1687537     .0091773     18.39    0.000    .1507529    .1867546
   wiek_2 | -.0017685     .0001077    -16.41    0.000   -.0019798   -.0015572
  _cons   | 6.792847      .1894588     35.85    0.000    6.421232    7.164462
```

1. Zinterpretować wartość współczynnika determinacji liniowej.
2. Czy zmienne objaśniające są łącznie nieistotne?
3. Zinterpretować oszacowanie parametru przy zmiennej *plec* i zmiennej *wiek*.
4. Proszę omówić w jaki sposób można przetestować hipotezę, iż największe zarobki są uzyskiwane przez osoby w wieku 48 lat (zapisać hipotezę zerową i hipotezę alternatywną, podać postać modelu z ograniczeniami).
5. Przetestować hipotezę z podpunktu 4 na poziomie istotności 0,05. Potrzebny do tego będzie jeden z poniższych modeli. Wskazać odpowiedni model, podać statystykę testową oraz wartość krytyczną, podjąć decyzję co do hipotezy.

Model 1

Zmienna $x1$ jest opisana następującym wzorem: $x1 = wiek\_2 - 96 * wiek$

```
Source     | SS              df          MS                Number of obs = 1595
----------+----------------------------------             F( 2, 1592) = 454.59
Model      | 514.092349       2     257.046174               Prob > F = 0.0000
Residual   | 900.189948    1592    .565445947               R-squared = 0.3635
----------+----------------------------------           Adj R-squared = 0.3627
Total      | 1414.2823     1594    .887253637                Root MSE = .75196
------------------------------------------------------------------------------
ln_dochod | Coef.         Std. Err.       t     P>|t|      [95% Conf. Interval]
---------+--------------------------------------------------------------------
plec       | -.8472704    .0378557      -22.38   0.000    -.9215227   -.7730181
x1         | -.0017353    .0000837      -20.72   0.000    -.0018995    -.001571
_cons      | 6.818258     .1821937       37.42   0.000    6.460893     7.175623
------------------------------------------------------------------------------
```

Model 2

Zmienna $x2$ jest opisana następującym wzorem: $x2 = 96 * wiek\_2 - wiek$

```
Source       | SS            df          MS                Number of obs = 1595
------------+----------------------------------             F( 2, 1592) = 235.55
Model        | 322.941452       2    161.470726               Prob > F = 0.0000
Residual     | 1091.34084    1592    .685515606              R-squared = 0.2283
------------+---------------------------------           Adj R-squared = 0.2274
Total        | 1414.2823     1594    .887253637                Root MSE = .82796
------------------------------------------------------------------------------
ln_dochod    | Coef.    Std. Err.       t       P>|t|         [95% Conf. Interval]
----------+-------------------------------------------------------------------
plec       | -.8150885   .0417013   -19.55      0.000     -.8968836    -.7332933
x2         | 1.89e-06    2.18e-07     8.68      0.000      1.46e-06     2.32e-06
_cons      | 10.15963    .0536086   189.51      0.000      10.05448     10.26478
------------------------------------------------------------------------------
```

Do testowania hipotez mogą przydać się następujące kwantyle:

$F_{0,95}(1,1591) = 3,8473103; F_{0,95}(2,1591) = 3,0013801;$
$F_{0,95}(1,1592) = 3,8473066; F_{0,95}(2,1592) = 3,0013765;$

Rozwiązanie Zadanie 2

1. 36,36% zmienności zmiennej zależnej zostało wyjaśnione za pomocą zmiennych niezależnych.

2. Test na łączną nieistotność regresji: $F = 303.00, p - value = 0,0000 < 0,05$ odrzucamy hipotezę zerową o łącznej nieistotności zmiennych objaśniających.
3. Kobiety mają średnio o 56,82% (100%*[exp(-0.84)-1]) niższe dochody niż mężczyźni.

   Semielastyczność dla wieku:

$$\frac{\partial E(ln\_dochod)}{\partial wiek} = \beta_{wiek} + 2 * \beta_{wiek\_2} * wiek$$

   Interpretacja oszacowań parametrów przy zmiennych dotyczących wieku nie jest możliwa w prosty sposób. Zależność między wartością oczekiwaną dochodu a wiekiem nie jest liniowa! Możemy obliczyć wierzchołek paraboli:

$$x_w = -\frac{\beta_{wiek}}{2 * \beta_{wiek\_2}} = -\frac{0,1687537}{2 * (-0,0017685)} \approx 47,71$$

   Ponadto ramiona paraboli są skierowane do dołu ($\beta_{wiek\_2} < 0$), więc w punkcie 47,71 jest przyjmowane maksimum. Czyli dochody osób poniżej 47,71 roku życia rosną, ale coraz wolniej, powyżej 47,71 lat dochód zaczyna spadać.

4. Największe zarobki będą uzyskiwane przez osoby w wieku 48 lat jeśli funkcja dla wieku będzie przyjmowała w tym punkcie maksimum. Oznacza to, że pochodna cząstkowa dla wieku musi w tym punkcie przyjmować wartość zero.

   Hipoteza zerowa: $\beta_{wiek} + 2 * \beta_{wiek\_2} * 48 = 0$
   Przekształcając uzyskujemy: $\beta_{wiek} = -96\beta_{wiek\_2}$

   Hipoteza alternatywna: $\beta_{wiek} \neq -96\beta_{wiek\_2}$

   Postać modelu z ograniczeniami:

$$\ln\_dochod_{\mathrm{i}} = \beta_0 + \beta_{plec} plec_i - 96\beta_{wiek\_2} wiek_i + \beta_{wiek\_2} wiek\_2_i + \varepsilon_i$$

   Przekształcając uzyskujemy:

$$\ln\_dochod_{\mathrm{i}} = \beta_0 + \beta_{plec} plec_i + \beta_{wiek_2}(wiek\_2_i - 96 wiek_i) + \varepsilon_i$$

5. Wartość statystki testowej:

$$F = \frac{(S_R - S)/g}{S/(N-K)} = \sim F(g, N-K)$$

gdzie: $S_R -$ suma kwadratów reszt modelu z ograniczeniami
$S -$ suma kwadratów rest zmodel bez ograniczeń
$g -$ liczba ograniczeń

Jeśli statystyka testowa jest większa od wartości krytycznej $F\sim(g, N-K)$, to odrzucamy hipotezę zerową.

$$F = \frac{(900{,}19 - 900{,}05)/1}{900{,}05/(1595-4)} = 0{,}2475$$

Uwaga: korzystamy z modelu 1.

Wartość krytyczna z której należy skorzystać do zweryfikowania hipotezy to $F_{0.95}(1{,}1591) = 3{,}8473103$. Ponieważ wartość statystki testowej jest mniejsza od wartości krytycznej to nie ma podstaw do odrzucenia hipotezy zerowej. Wobec tego największe zarobki są uzyskiwane przez osoby w wieku 48 lat.

Zadanie 3

Na podstawie pewnej bazy danych oszacowano liczbę przejechanych kilometrów podczas wyjazdów wakacyjnych gospodarstw domowych (*kilometry*). Zmienny niezależnymi są dochód w złotówkach (*doch*); średni wiek dorosłych członków gospodarstwa domowego w latach (*wiek*); fakt posiadania dzieci (*dzieci*: 0 – w gospodarstwie domowym nie ma dzieci; 1– w gospodarstwie domowym są dzieci). Model szacowano na próbie liczącej 200 obserwacji. Przyjąć założenie, iż w modelu nie występuje autokorelacja.
Hipotezy testować na poziomie istotności 0,05. Odpowiedzi uzasadnić podając *p-value*.

$$kilometry_i = \beta_1 + \beta_2 doch_{\,i} + \beta_3 wiek_i + \beta_4 dzieci_i + \varepsilon_i \quad (1)$$

kilometry	Coef.	Std. Err.	t	p-value
doch	14.20	1.80	7.89	0.022
wiek	15.74	3.76	4.19	0.000
dzieci	-81.83	27.13	-3.02	0.003
_cons	-391.55	169.78	-2.31	0.000

1. Badacz skorzystał z następujących informacji:

$Oznaczmy$: $kilometry_i = Y_i$

$$\widehat{Y_\iota} = -391{,}55 + 14{,}20doch_i + 15{,}74wiek_i - 81{,}83dzieci_i \qquad (2)$$
$$(1{,}80) \qquad (3{,}76) \qquad (27{,}13)$$

$RSS_R = 185{,}66 \;\; N = 200$

$$\widehat{Y_\iota} = -350{,}52 + 10{,}10doch_i + 18{,}65wiek_i - 75{,}43dzieci_i + 1{,}2\hat{Y}_i^2 - 2{,}5\hat{Y}_i^3 + 7{,}6\hat{Y}_i^4 \qquad (3)$$
$$(1{,}50) \qquad (4{,}49) \qquad (26{,}01) \qquad (12{,}1) \quad (24{,}2) \quad 140{,}3)$$

$RSS_U = 138{,}41 \;\; N = 200$

i obliczył następującą statystykę:

$$F_{(M;\,N-K)} = \frac{(RSS_R - RSS_{U)}/M}{RSS_U/(N-K)} = \frac{(185{,}66 - 138{,}41)/3}{138{,}41/(200-4)} = 22{,}30$$

Jaki test przeprowadził badacz? Podać jego nazwę, hipotezę zerową oraz zweryfikować ją.

Do testowania hipotezy mogą przydać się następujące kwantyle:

$F_{0,95}(3{,}196) = 2{,}85$

2. Istnieje podejrzenie, że zależność między liczbę przejechanych kilometrów podczas wyjazdów wakacyjnych gospodarstw domowych a zmiennymi objaśniającymi zależy od obecności dzieci w gospodarstwie domowym. Wiemy, że suma kwadratów reszt dla modelu oszacowanego na próbie gospodarstw domowych nie posiadających dzieci wynosi 50,22 natomiast na próbie zawierającej gospodarstwa posiadające dzieci 80,03. Proszę przetestować hipotezę o stabilności parametrów w obu podpróbkach.

   Informacje dodatkowe:

   a. model na całej próbie wersja 1

$$\widehat{Y}_i = -391{,}55 + 14{,}20 doch_i + 15{,}74 wiek_i - 81{,}83 dzieci_i$$

$$(1{,}80) \qquad (3{,}76) \qquad (27{,}13)$$

$$RSS_R = 185{,}66 \quad N = 200$$

   b. model na całej próbie wersja 2

$$\widehat{Y}_i = -321{,}25 + 12{,}10 doch_i + 11{,}34 wiek_i$$

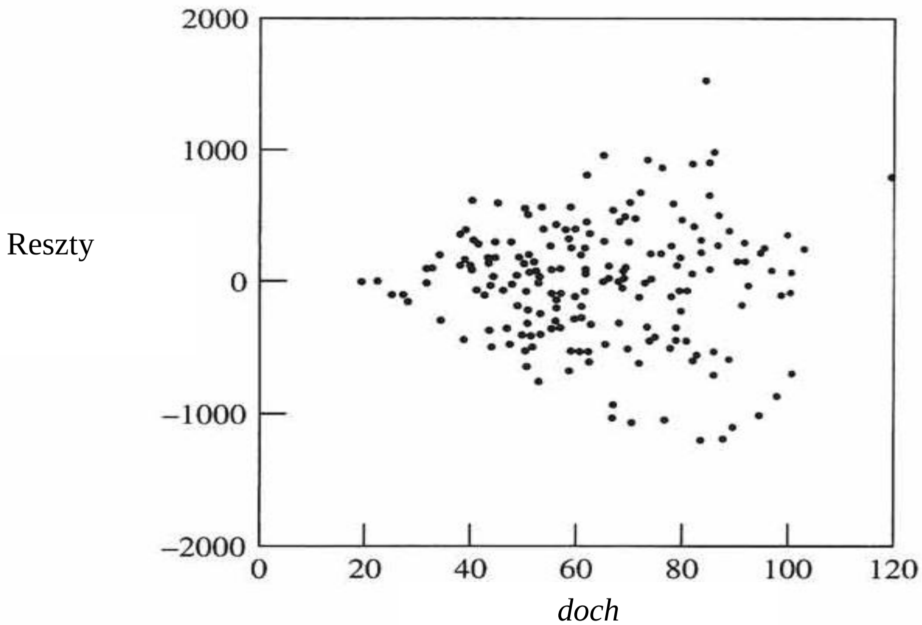
$$(1{,}65) \qquad (2{,}53) \qquad (22{,}14)$$

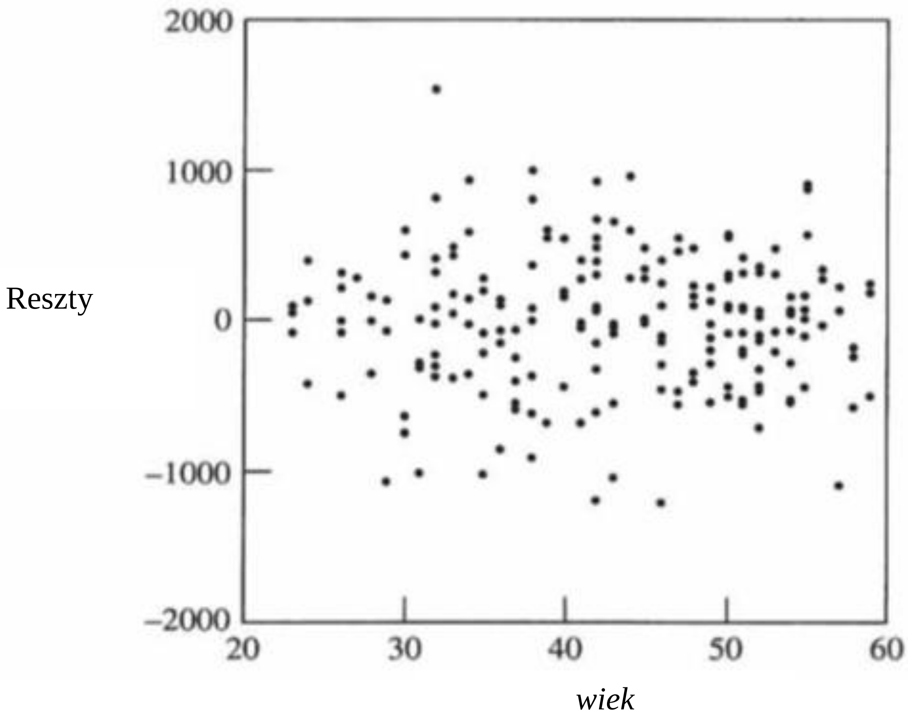
$$RSS_R = 180{,}34 \quad N = 200$$

   Do testowania hipotez mogą przydać się następujące kwantyle:

$F_{0,95}(4{,}192) = 2{,}41869$;   ; $F_{0,95}(4{,}194) = 2{,}418202$; $F_{0,95}(4{,}200) = 2{,}416800$;
$F_{0,95}(3{,}192) = 2{,}651640$; $F_{0,95}(3{,}194) = 2{,}651153$; $F_{0,95}(3{,}200) = 2{,}649752$

3. Badacz narysował poniższe wykresy reszt w zależności od zmiennych niezależnych a) *doch*; b) *wiek*:





Na podstawie powyższych wykresów określić jaki problem może wystąpić w modelu? Odpowiedź uzasadnić.

4. W kolejnym kroku badacz przeprowadził testy na heteroskedastyczność. Z poniższej listy wybrać odpowiednie testy i zinterpretować ich wyniki.

Test Breuscha-Pagana:	p-value = 0,000
Test Harveya-Godfreya:	p-value = 0,002
Test White'a:	p-value = 0,004
Test Levina i Lina:	p-value = 0,000
Test Ima, Pesarana i Shina	p-value = 0,125
Test Cichocki-Nehrebecka	p-value = 0,009
Test Maddali i Wu:	p-value = 0,000
Test Kao:	p-value = 0,001
Test McCoskey'a:	p-value = 0,040
Test Pedroniego:	p-value = 0,111
Test Larssona:	p-value = 0,066

5. Badacz oszacował model (1) za pomocą:
   a) Metody Najmniejszych Kwadratów;
   b) Metody Najmniejszych Kwadratów z użyciem odpornej macierzy White'a;
   c) Uogólnionej Metody Najmniejszy Kwadratów z założeniem, iż $\sigma_i^2 = \sigma^2 doch_i$

Metoda Najmniejszych Kwadratów

kilometry	Coef.	Std. Err.	t	p-value
doch	14.20	1.80	7.89	0.000
wiek	15.74	3.76	4.19	0.000
dzieci	-81.83	27.13	-3.02	0.003
_cons	-391.55	169.78	-2.31	0.022

Metoda Najmniejszych Kwadratów z odporną macierzą White'a

kilometry	Coef.	Std. Err.	t	p-value
doch	14.20	1.94	7.32	0.000
wiek	15.74	3.97	3.97	0.000
dzieci	-81.83	29.15	-2.81	0.006
_cons	-391.55	142.65	-2.74	0.007

Uogólniona Metoda Najmniejszych Kwadratów

kilometry	Coef.	Std. Err.	t	p-value
doch	13.95	1.48	9.42	0.001
wiek	16.72	3.02	5.53	0.000

```
dzieci    |   -76.81         21.85            -3.52       0.000
_cons     |  -425.00        121.44            -3.50       0.001
-----------------------------------------------------------------------------
```

Na podstawie wcześniejszych punktów określić w oparciu o które z powyższych wyników regresji (pkt.5) można przeprowadzić wnioskowanie statystyczne? Odpowiedź uzasadnić.

6. Badacz stanął przed wyborem:
   - jeśli w modelu (1) błędy losowe są sferyczne, to nie robi nic;
   - jeśli błędy są niesferyczne, to badacz stosuje Uogólnioną Metodę Najmniejszych Kwadratów z założeniem, iż $\sigma_i^2 = \sigma^2 doch_i$.

Czy błędy losowe w modelu (1) są niesferyczne? Odpowiedź uzasadnić.

Jeśli błędy są niesferyczne, to podać postać macierzy (L) przy pomocy której badacz przekształcił model MNK do modelu w którym błędy losowe są sferyczne.

Rozwiązanie Zadanie 3

1. Badacz przeprowadził test RESET. Hipoteza zerowa: $\hat{Y}_i^2 = \hat{Y}_i^3 = \hat{Y}_i^4 = 0$ (lub forma funkcyjna modelu (1) jest liniowa).
   Wartość statystyki: $F_{(3;\,196)} = 22{,}30$ a wartość krytyczna $F_{0,95}(3{,}196) = 2{,}85$.
   Wartość statystyki testowej jest większa od wartości krytycznej więc należy odrzucić hipotezę zerową. Oznacza to, że forma funkcyjna modelu (1) nie jest liniowa.

2. Statystyka testowa dla testu stabilności parametrów ma postać:

$$F = \frac{(S_R - \sum_j S_j)\,/[K(m-1)]}{\sum_j S_j\,/(N - mK)} = \sim F(K(m-1), N - mK)$$

gdzie: S_R − suma kwadratów reszt modelu na całej próbie
S_j − suma kwadratów reszt modelu dla j − tej podpróbki
m − liczba podpróbek

$$F = \frac{(180{,}34 - 50{,}22 - 80{,}03)/[3*(2-1)]}{(50{,}22 + 80{,}03)/(200 - 2*3)} = 24{,}8674$$

Uwaga: korzystamy z modelu na całej próbie w wersji 2.

Wartość krytyczna z której należy skorzystać do zweryfikowania hipotezy to

$F_{0.95}(3{,}194) = 2{,}651153$. Ponieważ wartość statystki testowej jest większa od wartości krytycznej, to należy odrzucić hipotezę zerową. Oznacza to, że parametry modelu nie są równe w podpróbie gospodarstw domowych nie posiadających dzieci i w podpróbie zawierającej gospodarstwa posiadające dzieci.

3.
   (a) Na podstawie wykresu reszt w zależności od zmiennej *doch* można podejrzewać, iż w modelu występuje heteroskedastyczność składnika losowego: wariancja reszt rośnie wraz ze wzrostem zmiennej *doch*.
   (b) W przypadku wykresu reszt w zależności od zmiennej *wiek* reszty zachowują się losowo. Wykres ten nie wskazuje na występowanie problemu heteroskedastyczności.

4. Test Breuscha-Pagana:
   i. hipoteza zerowa: homoskedastyczność składnika losowego.
   ii. p-value=0,000<0,05 odrzucamy hipotezę zerową o homoskedastyczności.

   Test White'a:
   i. hipoteza zerowa: homoskedastyczność składnika losowego.
   ii. p-value=0,004<0,05 odrzucamy hipotezę zerową o homoskedastyczności.

6. Z tego powodu, iż w modelu występuje problem heteroskedastyczności na co wskazuje wykres reszt w zależności od zmiennej *doch* (pkt.3) oraz wyniki testów diagnostycznych (pkt.4) wnioskowanie statystyczne należy oprzeć na modelu oszacowanym za pomocą Metody Najmniejszych Kwadratów z użyciem odpornej macierzy White'a lub Uogólnionej Metody Najmniejszy Kwadratów.

7. Tak, błędy losowe w modelu (1) są niesferyczne, ponieważ w modelu tym występuje problem heteroskedastyczności na co wskazuje wykres reszt w zależności od zmiennej *doch* (pkt.3) oraz wyniki testów diagnostycznych (pkt.4)

   Macierz L ma następująca postać: $L = \begin{bmatrix} \frac{1}{\sqrt{doch_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{doch_n}} \end{bmatrix}$