

**Egzamin z ekonometrii – wersja ogólna**  
**04-02-2016**

**Pytania teoretyczne**

1. Wymienić założenia Klasycznego Modelu Regresji Liniowej (KMRL).
2. Wyprowadzić estymator MNL dla modelu z wieloma zmiennymi objaśniającymi.
3. Udowodnić, że w modelu ze stałą  $TSS=ESS+RSS$ .
4. Wyjaśnić, dlaczego  $R^2$  nie można używać do porównywania modeli.

## Zadanie 1

Metodą Najmniejszych Kwadratów na próbie liczącej 50 obserwacji oszacowano następujący model:

$$\hat{y}_i = 2 + 2x_{1i} + 3x_{2i} + x_{3i}$$

W celu sprawdzenia czy w modelu występuje współliniowość oszacowano trzy regresje pomocnicze:

- I)  $\hat{x}_{1i} = 1 - x_{2i} + 2x_{3i}, R^2 = 0,02$
- II)  $\hat{x}_{2i} = 2 + 4x_{1i} - 2x_{3i}, R^2 = 0,95$
- III)  $\hat{x}_{3i} = 0,5 + x_{1i} + 0,5x_{2i}, R^2 = 0,97$

1. Za pomocą statystyki VIF sprawdzić, czy zmienna  $x_2$  wywołuje współliniowość. Odpowiedź uzasadnić.
2. Za pomocą testu opartego na statystyce F sprawdzić czy zmienna  $x_1$  wywołuje współliniowość. Odpowiedź uzasadnić.

Wartości krytyczne:  $F_{0,95}(1,47) = 4,05$ ;  $F_{0,95}(2,47) = 3,20$ ;  $F_{0,95}(3,47) = 2,80$ ;  
 $F_{0,95}(1,50) = 4,03$ ;  $F_{0,95}(2,50) = 3,18$ ;  $F_{0,95}(3,50) = 2,79$

3. Mamy następujące regresje:

$$(*) \quad y_i \text{ na stałej, } x_{2i} \text{ i } x_{3i}$$

$$(**) \quad y_i^* \text{ na } z_i \quad \text{gdzie } y_i^* = y_i - x_{2i}, \quad z_i = x_{2i} + x_{3i}$$

Jaka zachodzi relacja między RSS w tych regresjach? Odpowiedź uzasadnić.

4. Postanowiono oszacować regresję wyjaśniającą liczbę zdobytych punktów na egzaminie z analizy matematycznej za pomocą liczby punktów uzyskanych przez studenta podczas egzaminu wstępnego na studia z matematyki. Jaki będzie prawdopodobny kierunek obciążenia parametru przy zmiennej oznaczającej liczbę punktów uzyskaną podczas egzaminu wstępnego z matematyki wynikiły z pominięcia ilorazu inteligencji studenta? Odpowiedź uzasadnić.

5. Przyjmujemy następujące oznaczenia:

- $y$  – dochód,
- $wiek$  – wiek wyrażony w latach,
- $miasto$  – zmienna przyjmująca wartość 1, jeżeli osoba mieszka w mieście oraz wartość 0 jeżeli osoba mieszka na wsi.

Analizujemy następującą regresję:

$$y_i = \beta_1 + \beta_2 miasto_i + \beta_3(1 - miasto_i) \times wiek_i + \beta_4 miasto_i \times wiek_i + \varepsilon_i$$

Jaka jest różnica w oczekiwanej zmianie dochodu przy wzroście wieku o jednostkę między osobami mieszkającymi w mieście i na wsi?

6. Badacz przeprowadził następującą regresję, gdzie  $spslops$  – zmienna ciągła mierząca prestiż zawodowy danej osoby,  $zodiak$  – zmienna dyskretna związana ze znakiem zodiaku danej osoby (1 – jeśli osoba ma dany znak zodiaku, 0 – w pozostałych przypadkach).

Source	SS	df	MS	Number of obs =	1120
Model	1833.82209	11	166.711099	F( 11, 1108) =	1.14
Residual	161789.87	1108	146.019738	Prob > F =	0.3245
				R-squared =	0.0112
				Adj R-squared =	0.0014
Total	163623.692	1119	146.223138	Root MSE =	12.084

spslops	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
zodiak					
byk	-1.309057	1.68456	-0.78	0.437	-4.614344 1.996231
bliźnięta	-.828004	1.707218	-0.49	0.628	-4.17775 2.521742
rak	-.6447555	1.716872	-0.38	0.707	-4.013443 2.723932
lew	.3678109	1.771107	0.21	0.836	-3.10729 3.842912
panna	2.339062	1.680263	1.39	0.164	-.9577947 5.635919
waga	.5977519	1.712001	0.35	0.727	-2.761377 3.956881
skorpion	-.7294914	1.721835	-0.42	0.672	-4.107916 2.648934
strzelec	2.889039	1.365184	2.11	0.030	-.574442 6.352519
koziorożec	1.732339	1.753696	0.99	0.323	-1.7086 5.173278
wodnik	.4877855	1.707218	0.29	0.775	-2.86196 3.837531
ryba	1.535559	1.726892	0.89	0.374	-1.852789 4.923907
_cons	37.84906	1.173689	32.25	0.000	35.54615 40.15196

Na podstawie uzyskanych wyników badacz stwierdza, iż znak zodiaku wpływa na prestiż zawodowy, uzasadniając to faktem, iż kategoria *strzelec* okazała się statystycznie istotną zmienną na poziomie istotności 5%. Zweryfikować powyższe stwierdzenie.

## Rozwiązanie Zadanie 1

1.  $VIF = \frac{1}{1-R_2^2} = \frac{1}{1-0,95} = 20$  (współczynnik inflacji wariancji)

Minimalna wartość jaka może przyjąć powyższa statystyka to 1. Im wyższa wartość VIF, tym silniejsza korelacja zmiennej  $x_2$  z pozostałymi zmiennymi objaśniającymi w wyjściowej regresji. Jeżeli  $VIF > 10$ , to przyjmujemy, że  $x_2$  wywołuje współliniowość w modelu wyjściowym.

$VIF = 20$ , zmienna  $x_2$  wywołuje współliniowość w modelu wyjściowym.

2. Testujemy łączną nieistotność regresji pomocniczej

$$\hat{x}_{1i} = 1 - x_{2i} + 2x_{3i}, \quad R_1^2 = 0,02$$

Wyznaczamy wartość statystyki testowej:

$$F = \frac{R_1^2/g}{(1 - R_1^2)/(N - K)} = \frac{0,02/2}{(1 - 0,02)/(50 - 3)} = 0,479592$$

Wyznaczamy wartość krytyczną:

$$F^* = F_{0,95}(2,47) = 3,20$$

Porównujemy wartość statystyki testowej z wartością krytyczną:

$F < F^*$  - brak podstaw do odrzucenia hipotezy zerowej o tym, że regresja jest łącznie nieistotna. Oznacza to, że nie istnieje liniowa zależność między zmienną  $x_1$  a zmiennymi  $x_2, x_3$  - w wyjściowym modelu nie występuje współliniowość.

3. Wprowadzamy następujące oznaczenia:  $RSS^*$  - suma kwadratów reszt dla modelu (\*),  $RSS^{**}$  - suma kwadratów reszt dla modelu (\*\*).

Sumy kwadratów reszt są rozwiązaniem następującego problemu optymalizującego:

$$\begin{aligned} RSS^* &= \min_{b_1, b_2, b_3} \sum_{i=1}^N (y_i - b_1 - b_2 x_{2i} - b_3 x_{3i})^2 \\ RSS^{**} &= \min_{\gamma} \sum_{i=1}^N (y_i - \gamma z_i)^2 = \min_{\gamma} \sum_{i=1}^N ([y_i - x_{2i}] - \gamma [x_{2i} + x_{3i}])^2 \\ &= \min_{\gamma} \sum_{i=1}^N (y_i - (\gamma + 1)x_{2i} - \gamma x_{3i})^2 \end{aligned}$$

W przypadku regresji (\*\*) minimalizujemy tę samą funkcję co dla regresji (\*), ale wprowadzamy następujące ograniczenie:

$$\begin{cases} \beta_1 = 0 \\ \beta_2 = \beta_3 + 1 \end{cases}$$

Czyli  $RSS^*$  to wartość minimalizowanej funkcji dla problemu bez ograniczeń, natomiast  $RSS^{**}$  to wartość minimalizowanej funkcji dla problemu z ograniczeniami, więc musi zachodzić

$$RSS^* \leq RSS^{**}$$

4. Jeżeli prawdziwy jest model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

natomiast estymujemy model

$$y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i$$

to obciążenie parametru przy zmiennej  $x_1$  wyraża się wzorem:

$$E(b_1) - \beta_1 = \beta_2 \frac{S_{x_2}}{S_{x_1}} \rho_{x_1, x_2}$$

$\beta_2 > 0$  i  $\rho_{x_1, x_2} > 0 \Rightarrow$  dodatnie obciążenie (inteligencja dodatnio wpływa na wynik egzaminu z analizy matematycznej i jest dodatnio skorelowana z uzyskana liczba punktów na egzaminie wstępnym).

5. Parametr  $\beta_3$  interpretujemy jako oczekiwaną zmianę dochodu przy wzroście wieku o jednostkę dla osób mieszkających na wsi.

Parametr  $\beta_4$  interpretujemy jako oczekiwaną zmianę dochodu przy wzroście wieku o jednostkę dla osób mieszkających w mieście.

Czyli różnica w oczekiwanej zmianie dochodu przy wzroście wieku o jednostkę między osobami mieszkającymi w mieście a na wsi wynosi  $\beta_4 - \beta_3$ .

6. W celu stwierdzenia, iż znak zodiaku wpływa na prestiż zawodowy powinniśmy przeanalizować test na łączną nieistotność kategorii zmiennej zodiak.

$$\begin{aligned} F(11, 1108) &= 1.14 \\ \text{Prob} > F &= 0.3245 \end{aligned}$$

$P$  – value = 0,3245  $>$   $\alpha = 0,05 \Rightarrow$  nie ma podstaw do odrzucenia hipotezy zerowej, o łącznej nieistotności kategorii zmiennej zodiak.

## Zadanie 2

Na podstawie pewnej bazy danych oszacowano wysokość rocznych dochodów w złotych ( $\ln\_dochod$  - 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		
Model	514.228436	3	171.409479	F( 3, 1591) = 303.00		
Residual	900.053861	1591	.565715815	Prob > F = 0.0000		
				R-squared = 0.3636		
				Adj R-squared = 0.3624		
				Root MSE = .75214		
Total	1414.2823	1594	.887253637			

$\ln\_dochod$	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  $x_1$  jest opisana następującym wzorem:  $x_1 = \text{wiek}_2 - 96 * \text{wiek}$

Source	SS	df	MS	Number of obs = 1595		
Model	514.092349	2	257.046174	F( 2, 1592) = 454.59		
Residual	900.189948	1592	.565445947	Prob > F = 0.0000		
				R-squared = 0.3635		
				Adj R-squared = 0.3627		
				Root MSE = .75196		
Total	1414.2823	1594	.887253637			

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	-.0015171
_cons	6.818258	.1821937	37.42	0.000	6.460893	7.175623

## Model 2

Zmienna  $x_2$  jest opisana następującym wzorem:  $x_2 = 96 * \text{wiek}_2 - \text{wiek}$

Source	SS	df	MS	Number of obs = 1595		
Model	322.941452	2	161.470726	F( 2, 1592) = 235.55		
Residual	1091.34084	1592	.685515606	Prob > F = 0.0000		
				R-squared = 0.2283		
				Adj R-squared = 0.2274		
				Root MSE = .82796		
Total	1414.2823	1594	.887253637			

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

- 36,36% zmienności zmiennej zależnej zostało wyjaśnione za pomocą zmiennych niezależnych.
- 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 już 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:

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

Przekształcając uzyskujemy:

$$\ln\_dochod_i = \beta_0 + \beta_{plec}plec_i + \beta_{wiek\_2}(wiek\_2_i - 96wiek_i) + \varepsilon_i$$

5. Wartość statystyki 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*). Zmiennymi niezależnymi są dochód w złotych (*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*.

$$\text{kilometry}_i = \beta_1 + \beta_2 \text{doch}_i + \beta_3 \text{wiek}_i + \beta_4 \text{dzieci}_i + \varepsilon_i \quad (1)$$

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:  $\text{kilometry}_i = Y_i$

$$\hat{Y}_i = -391,55 + 14,20\text{doch}_i + 15,74\text{wiek}_i - 81,83\text{dzieci}_i \quad (2)$$

(1,80)            (3,76)            (27,13)

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

$$\hat{Y}_i = -350,52 + 10,10\text{doch}_i + 18,65\text{wiek}_i - 75,43\text{dzieci}_i + 1,2\hat{Y}_i^2 - 2,5\hat{Y}_i^3 + 7,6\hat{Y}_i^4 \quad (3)$$

(1,50)            (4,49)            (26,01)            (12,1)            (24,2)            140,3)

$$RSS_U = 138,41 \quad N = 200$$

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 liczbą 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*

$$\hat{Y}_i = -391,55 + 14,20\text{doch}_i + 15,74\text{wiek}_i - 81,83\text{dzieci}_i$$

(1,80)            (3,76)            (27,13)

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

*b. model na całej próbie wersja 2*

$$\hat{Y}_i = -321,25 + 12,10\text{doch}_i + 11,34\text{wiek}_i$$

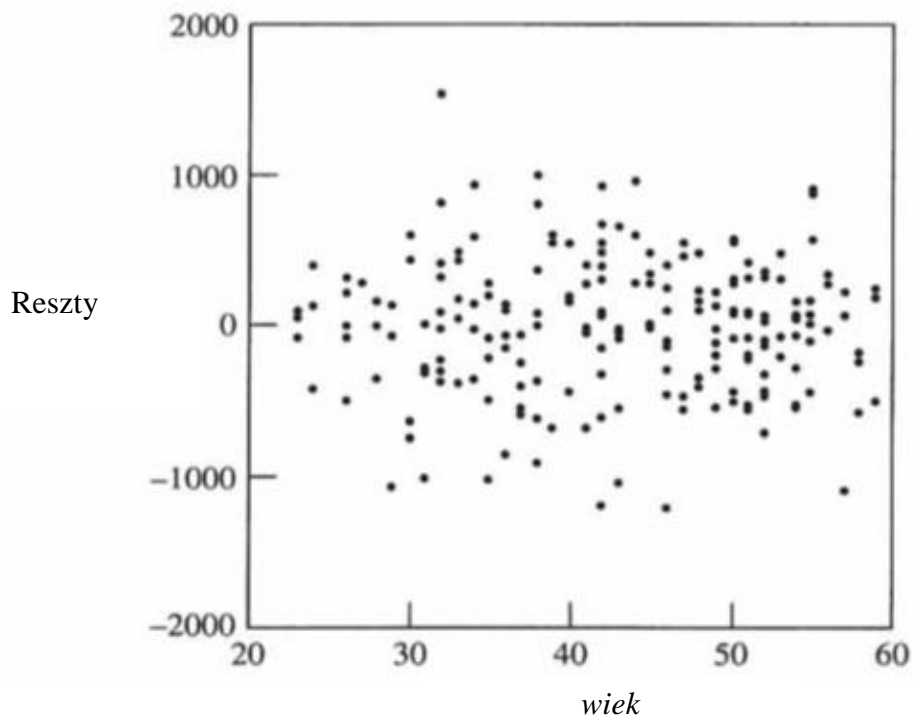
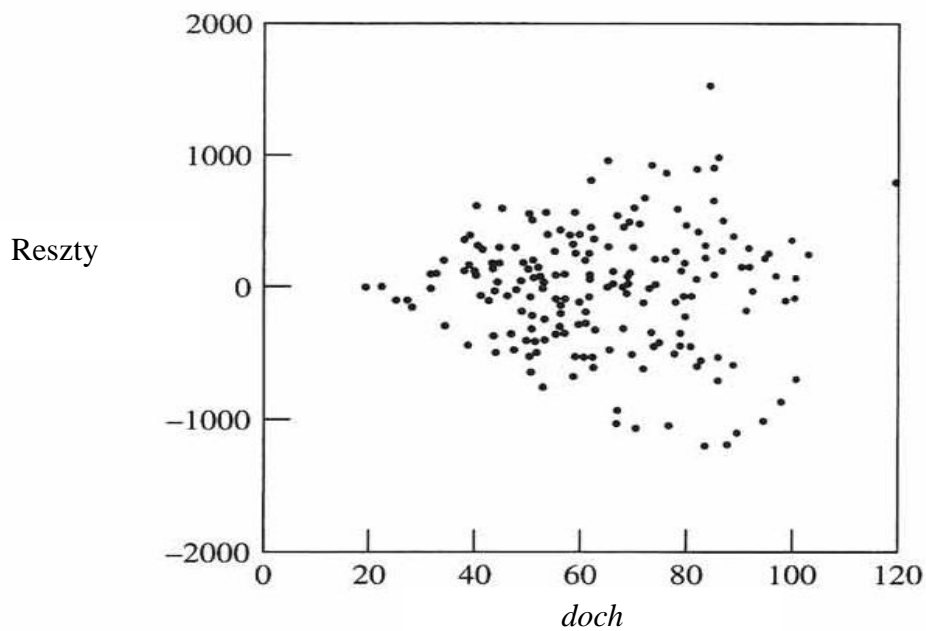
(1,65)            (2,53)            (22,14)

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

Do testowania hipotez mogą przydać się następujące kwantyle:

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

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-Godfrey:	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:
- Metody Najmniejszych Kwadratów;
  - Metody Najmniejszych Kwadratów z użyciem odpornej macierzy White'a;
  - 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  $\sigma_i^2 = \sigma^2 \text{doch}_i$ .

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 następującą postać: 
$$L = \begin{bmatrix} \frac{1}{\sqrt{doch_1}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{doch_n}} \end{bmatrix}$$