Egzamin z ekonometrii – wersja ogólna
31-01-2014

Pytania teoretyczne

1. Podać postać przekształcenia Boxa-Coxa i wyjaśnić, do czego jest stosowane w ekonometrii.
2. Porównaj zastosowania znanych ci kontrastów ze standardowym sposobem rozkodowania zmiennej dyskretnej.
3. Wyprowadzić postać macierzy wariancji kowariancji $\boldsymbol{b}$ i podać interpretację jej elementów.
4. Wyjaśnić, jakie korzyści i niebezpieczeństwa łączą się z narzucaniem ograniczeń na model.

Zadanie 1

Dany jest model:

$$y_i = \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \delta^2), \quad i = 1, \dots N$$

gdzie x_i jest nielosowe.

1. Pokazać, że estymator $\tilde{\beta} = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i}$ i estymator $\hat{\beta} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} {x_i}^2}$ są nieobciążonymi estymatorami β.
2. Obliczyć wariancję estymatora $\tilde{\beta}$ i estymatora $\hat{\beta}$ i pokazać, iż wariancja $\hat{\beta}$ jest mniejsza od wariancji $\tilde{\beta}$.

Rozwiązanie Zadanie 1

1. $E(\tilde{\beta}) = \frac{\sum_{i=1}^{N} E(\beta x_i + \varepsilon_i)}{\sum_{i=1}^{N} x_i} = \frac{\sum_{i=1}^{N} E(\beta x_i) + E(\varepsilon_i)}{\sum_{i=1}^{N} x_i} = \frac{\sum_{i=1}^{N} (\beta x_i)}{\sum_{i=1}^{N} x_i} = \frac{\beta \sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i} = \beta$

$$E(\hat{\beta}) = \frac{\sum_{i=1}^{N} E(x_i(\beta x_i + \varepsilon_i))}{\sum_{i=1}^{N} x_i^2} = \frac{\sum_{i=1}^{N} E(\beta x_i^2 + x_i \varepsilon_i)}{\sum_{i=1}^{N} x_i^2} = \frac{\sum_{i=1}^{N} E(\beta x_i^2) + E(x_i \varepsilon_i)}{\sum_{i=1}^{N} x_i^2} = \frac{\sum_{i=1}^{N} (\beta x_i^2)}{\sum_{i=1}^{N} x_i^2} =$$

$$= \frac{\beta \sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N} x_i^2} = \beta$$

2. $Var(\tilde{\beta}) = Var\left(\frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i}\right) = \frac{Var(\sum_{i=1}^{N} y_i)}{(\sum_{i=1}^{N} x_i)^2} = \frac{Var(\sum_{i=1}^{N} \beta x_i + \varepsilon_i)}{(\sum_{i=1}^{N} x_i)^2} = \frac{\sum_{i=1}^{N} Var(\varepsilon_i)}{(\sum_{i=1}^{N} x_i)^2} = \frac{N\sigma^2}{(\sum_{i=1}^{N} x_i)^2}$

$$Var(\hat{\beta}) = Var\left(\frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}\right) = \frac{Var(\sum_{i=1}^{N} x_i y_i)}{(\sum_{i=1}^{N} x_i^2)^2} = \frac{Var(\sum_{i=1}^{N} x_i(\beta x_i + \varepsilon_i))}{(\sum_{i=1}^{N} x_i^2)^2} = \frac{\sum_{i=1}^{N} x_i^2 Var(\varepsilon_i)}{(\sum_{i=1}^{N} x_i^2)^2} =$$

$$= \frac{\sum_{i=1}^{N} x_i^2 \sigma^2}{(\sum_{i=1}^{N} x_i^2)^2} = \frac{\sigma^2 \sum_{i=1}^{N} x_i^2}{(\sum_{i=1}^{N} x_i^2)^2} = \frac{\sigma^2}{(\sum_{i=1}^{N} x_i^2)}$$

$$Var(\hat{\beta}) < Var(\tilde{\beta})$$

$$\frac{\sigma^2}{(\sum_{i=1}^{N} x_i^2)} < \frac{N\sigma^2}{(\sum_{i=1}^{N} x_i)^2}$$

$$\frac{1}{(\sum_{i=1}^{N} x_i^2)} < \frac{N}{(\sum_{i=1}^{N} x_i)^2}$$

Zadanie 2

Do dyspozycji jest baza danych z następującymi zmiennymi: całkowita liczba godzin szkoleń (t*othours*), roczne wynagrodzenie w USD (*avgsal*), logarytm *avgsal* (*lavgsal*), roczne przychody ze sprzedaży w USD (*sales*), logarytm *sales* (*lsales*), zmienna dla roku 1987 (*dy1:* 1 – dla roku 1987, 0 – w przeciwnym przypadku), zmienna dla roku 1988 (*dy2:* 1 – dla roku 1988, 0 – w przeciwnym przypadku), zmienna dla roku 1989 (*dy3:* 1 – dla roku 1989, 0 – w przeciwnym przypadku), wielkość firmy (*large:* 1 – gdy duża firma, 0 – w przeciwnym przypadku).
Hipotezy testować na poziomie istotności 0,1. Odpowiedzi uzasadnić podając *p-value*.

```
 sum tothrs avgsal lavgsal sales lsales dy1 dy2 dy3 large

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      tothrs |       304    33.39474    52.42831          0        320
      avgsal |       304    18723.52    6967.911       4237      42583
     lavgsal |       304    9.773076    .3590829   8.351611   10.65921
       sales |       304     6413918     7899873     110000   4.90e+07
      lsales |       304    15.07133    1.126338   11.60824   17.70733
-------------+--------------------------------------------------------
         dy1 |       304    .3190789    .4668882          0          1
         dy2 |       304    .3322368     .471792          0          1
         dy3 |       304    .3486842    .4773396          0          1
       large |       304    .2138158    .4106743          0          1
```

1. Przeprowadzono regresję zmiennej *tothours na* logarytmie *avgsal* i logarytmie *sales* używając losowej próbki 304 obserwacji, pobranych dla lat 1987, 1988 i 1989:

```
 reg tothrs lavgsal lsales

  Source |     SS          df       MS                 Number of obs =    304
-------------+------------------------------           F( 2,  301)   =  10.14
   Model | 52584.9232    2     26292.4616              Prob > F      = 0.0001
Residual | 780279.708  301     2592.29139              R-squared     = 0.0631
-------------+------------------------------           Adj R-squared = 0.0569
   Total | 832864.632  303 2748.72816                  Root MSE      = 50.915

------------------------------------------------------------------------------
  tothrs |      Coef.  Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 lavgsal |   17.08952  8.314689     2.06   0.041     .7272393    33.4518
  lsales |   -11.5002   2.65077    -4.34   0.000   -16.71659  -6.283811
   _cons |   39.70086  83.09774     0.48   0.633   -123.8252    203.227
------------------------------------------------------------------------------
```

a) Zinterpretować wartość współczynnika dla *lavgsal*.

2. Następnie dodano do modelu zmienne *dy2* i *dy3*:

```
 reg tothrs lavgsal lsales dy2 dy3

  Source |       SS       df       MS              Number of obs =     304
-------------+------------------------------           F( 4,   299)   =    5.46
   Model | 56680.5896     4  14170.1474            Prob > F      = 0.0003
Residual | 776184.042   299  2595.93325            R-squared     = 0.0681
-------------+------------------------------           Adj R-squared = 0.0556
   Total | 832864.632   303 2748.72816             Root MSE      =  50.95

------------------------------------------------------------------------------
  tothrs |      Coef.  Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 lavgsal |   15.59959  8.405549    1.86   0.064    -.9419353   32.14112
  lsales |  -11.66302  2.655821   -4.39   0.000    -16.88949  -6.436551
     dy2 |   4.221359  7.271943    0.58   0.562    -10.08931   18.53203
     dy3 |   9.090815  7.254422    1.25   0.211    -5.185378   23.36701
   _cons |   52.14363  83.74939    0.62   0.534    -112.6693   216.9565
------------------------------------------------------------------------------
```

a) Zaproponować sposób przetestowania hipotezy, iż **rok** nie wpływa na całkowitą liczbę godzin szkoleń.

b) Zinterpretować wartość współczynnika dla *dy3* (proszę pominąć kwestię jego istotności).

3. Czy na podstawie poniższych tabel możliwa jest identyfikacja heteroskedastyczności? Odpowiedź uzasadnić. Jeśli możliwa jest ta identyfikacja to która zmienna wywołuje problem heteroskedastyczności?

Tabela 1

```
predict uhat, resid
gen uhat2 = uhat^2
reg uhat lavgsal lsales dy2 dy3

  Source |       SS       df       MS              Number of obs   =     304
-------------+------------------------------           F( 4,   299)     =    0.00
   Model | 1.1642e-10     4  2.9104e-11            Prob > F        =  1.0000
Residual | 776184.046   299  2595.93326            R-squared       =  0.0000
-------------+------------------------------           Adj R-squared   = -0.0134
   Total | 776184.046   303 2561.66352             Root MSE        =   50.95

------------------------------------------------------------------------------
    uhat |      Coef.  Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 lavgsal |   1.92e-07  8.405549    0.00   1.000    -16.54153   16.54153
  lsales |  -6.72e-08  2.655821   -0.00   1.000    -5.226469   5.226469
     dy2 |  -2.04e-07  7.271943   -0.00   1.000    -14.31067   14.31067
     dy3 |  -1.75e-07  7.254422   -0.00   1.000    -14.27619   14.27619
   _cons |  -7.52e-07  83.74939   -0.00   1.000    -164.8129   164.8129
```

Tabela 2

```
gen uhat2 = uhat^2
reg uhat2 lavgsal lsales dy2 dy3

  Source |       SS       df       MS              Number of obs =     304
-------------+------------------------------           F( 4,   299)   =    3.82
   Model | 1.0166e+09     4  254151538             R-squared     = 0.0486
Residual | 1.9889e+10   299 66516778.3             Adj R-squared = 0.0359
-------------+------------------------------           Root MSE      = 8155.8
   Total | 2.0905e+10   303 68993804.9

------------------------------------------------------------------------------
   uhat2 |      Coef.  Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------------------------------------------------------------------
 lavgsal |   232.6349  1345.504     0.17   0.863    -2415.222    2880.492
  lsales |  -1631.846   425.126    -3.84   0.000    -2468.464   -795.2283
     dy2 |  -143.6411  1164.044    -0.12   0.902    -2432.397    2147.115
     dy3 |     534.56  1161.239     0.46   0.646    -1750.677    2819.797
   _cons |    24735.1  13406.04     1.85   0.006    -1647.041    51117.25
```

4. Czy w oparciu o poniższe wyniki można stwierdzić, iż model szacowany w punkcie 1 jest różny dla dużych i małych firm? Odpowiedź uzasadnić.

```
gen large_lavgsal = large*lavgsal

gen large_lsales = large*lsales

reg tothrs lavgsal lsales large large_lavgsal large_lsales


  Source |       SS       df       MS              Number of obs =     304
-------------+------------------------------           F( 5,   298)   =    5.08
   Model | 65423.7882     5  13084.7576            Prob > F      = 0.0002
Residual | 767440.843   298  2575.30484            R-squared     = 0.0786
-------------+------------------------------           Adj R-squared = 0.0631
   Total | 832864.632   303 2748.72816             Root MSE      = 50.747

------------------------------------------------------------------------------
      tothrs |      Coef.  Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lavgsal |   22.45324  8.665497     2.59   0.010     5.399921    39.50656
      lsales |  -9.971127  3.131076    -3.18   0.002    -16.13295   -3.809305
       large |   582.7495  273.2314     2.13   0.034     45.04194    1120.457
large_lavgsal|  -59.82844  30.77553    -1.94   0.053    -120.3933    .7364569
large_lsales |  -.0325676  9.070745    -0.00   0.997     -17.8834    17.81826
       _cons |  -34.48482  90.38981    -0.38   0.703     -212.368    143.3984
------------------------------------------------------------------------------
```

5. W kolejnej regresji uwzględniono interakcję pomiędzy zmienną *large* a zmienną *lavgsal*. Nowo powstałą zmienną oznaczono: *large_lavgsal*

```
reg tothrs lavgsal lsales large large_lavgsal

  Source |       SS       df       MS              Number of obs =     304
-------------+------------------------------           F( 4,   299)   =    6.37
   Model |  65423.755     4   16355.9388           Prob > F       = 0.0001
Residual | 767440.877   299   2566.69189           R-squared      = 0.0786
-------------+------------------------------           Adj R-squared = 0.0662
   Total | 832864.632   303 2748.72816             Root MSE       = 50.663

------------------------------------------------------------------------------
      tothrs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     lavgsal |   22.45481   8.640065     2.60   0.010     5.451766    39.45785
      lsales |  -9.975008   2.933707    -3.40   0.001    -15.74834    -4.20168
       large |   582.6985   272.4064     2.14   0.033     46.62183    1118.775
large_lavgsal|   -59.8758   27.75878    -2.16   0.032    -114.5031   -5.248465
       _cons |  -34.44257    89.4705    -0.38   0.701    -210.5142    141.6291
------------------------------------------------------------------------------
```

a) Zinterpretować wielkość współczynnika przy interakcji pomiędzy zmienną *large* a *lavgsal*.

Rozwiązanie Zadanie 2

1. a) Wzrost rocznego wynagrodzenia o 1% powoduje wzrost całkowitej liczby godzin szkoleń średnio o 0,17godziny *ceteris paribus*.

2. a) Musimy przetestować dwa ograniczenia, które można zapisać w następujący sposób:

$$H_0: \beta_{dy2} = \beta_{dy3} = 0$$

Następnie szacujemy model bez ograniczeń. Korzystając z modelu bez ograniczeń i z ograniczeniami obliczamy statystykę:

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

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 $F^* = F(g, N-K)$, to odrzucamy hipotezę zerową o braku wpływu roku na całkowitą liczbę godzin szkoleń.

b) Całkowita liczba godzin szkoleń jest większa w 1989 średnio o 9,09 godziny od całkowitej liczby godzin szkoleń w 1987 *ceteris paribus*.

3. Tak, możliwa jest identyfikacja heteroskedastyczności. Przeprowadzając regresję pomocniczą kwadratów reszt na zbiorze zmiennych od których chcemy uzależnić wariację składnika losowego (Tabela 2) otrzymujemy, iż zmienna *lsales* wydaje się być istotna (*t=-3.84; p-value=0.000*) w wyjaśnianiu zróżnicowania kwadratów reszt.

   Wysoka wartość statystyki F modelu, również sugeruje obecność heteroscedastyczności w składniku losowym, ponieważ zmienne są łącznie istotne, czyli wyjaśniają kwadrat błędu.

   Postępując dalej w sposób analogiczny, możemy dokładnie znaleźć funkcję, która jest odpowiedzialna za heteroscedastyczność.

4. Tak. W oparciu o uzyskane wyniki w punkcie 4 można stwierdzić, iż oszacowania parametrów modelu w punkcie 1 są różne w podpróbach wyodrębnionych za pomocą zmiennej *large* (dla dużych i małych firm).

   Problem niestabilności parametrów można rozwiązać poprzez:
   ◦ wprowadzenie do modelu interakcji pomiędzy zmiennymi 0 – 1 związanymi z podziałem na grupy a odpowiednimi zmiennymi objaśniającymi (w przypadku gdy jedynie część parametrów jest różna dla analizowanych podprób)

◦ estymacje osobnych regresji na wyodrębnionych podpróbach.

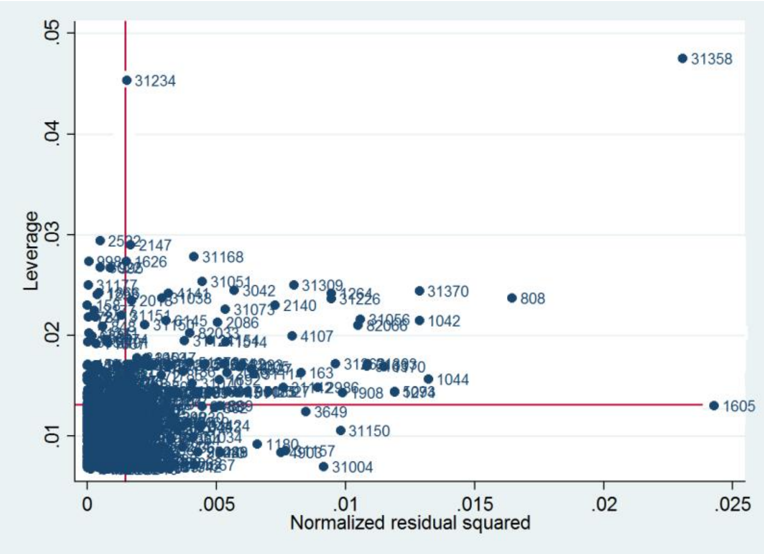
Parametry modelu szacowanego w pkt.4 nie są stabilne w podpróbkach ze względu na zmienną *lavgsal*. W związku z tym uzasadnione jest wprowadzenie interakcji*: large_lavgsal* (*p-value*=0,053).

5. a) Wzrost rocznego wynagrodzenia w dużych firmach o 1% prowadzi do spadku całkowitej liczby godzin szkoleń średnio o 0,3733 godzin (0,2254-0,5987=-0,3733) *ceteris paribus*.

Zadanie 3

Na podstawie danych BAEL z 2006 roku oszacowano wysokość wynagrodzenia dla studentów (*ln_placa* – logarytm wynagrodzenia). Zmiennymi objaśniającymi są wiek (*wiek*), miejsce zamieszkania (*duze_miasto*: 0 – miejscowość do 100 tys. mieszkańców 1 – miasto powyżej 100 tys. mieszkańców), rodzaj studiów (*rodzaj*: 0 – studia stacjonarne, 1 – studia niestacjonarne, 2 – studia zaoczne), płeć (*plec*: 0 – mężczyzna, 1 – kobieta), rodzaj umowy o pracę (*umowa*: 0 – umowa na czas określony; 1 – umowa na czas nieokreślony), rodzaj własności pracodawcy (*niepubliczny*: 0 – publiczny pracodawca, 1 – niepubliczny pracodawca), interakcja między płcią a rodzajem umowy. Oszacowania parametrów znajdują się poniżej. Hipotezy testować na poziomie istotności 0,1. Odpowiedzi uzasadnić podając *p-value*.

1. Po oszacowaniu regresji badacz uzyskał wykres przedstawiający zależność między dźwignią a wystandaryzowanymi resztami podniesionymi do kwadratu oraz obliczył standaryzowane reszty (*reszty_st),* dźwignię (*dzwignia*) i odległość Cook’a (*cook*). Obserwacje, dla których statystyki te były największe znajdują się w tabeli poniżej. Wyjaśnić na postawie wykresu i tabeli, które obserwacje budzą podejrzenia i dlaczego?



numer	placa	reszty_st	dzwignia	cook
31358	100	-3.782185	.0473836	.0634435
31234	1000	-1.119193	.0439123	.0396513
1605	2000	3.35655	.0159973	.0158197
31226	1200	2.55908	.0236324	.0176125
1264	1600	2.561957	.024149	.0180475

2. W modelu oszacowano efekty progowe dla wielkości miejscowości (*miasto*: 0 – przynajmniej wieś, 1 – przynajmniej miejscowość do 20 tys. mieszkańców, 2 – przynajmniej miejscowość do 100 tys. mieszkańców, 3 – przynajmniej miejscowość powyżej 100 tys. mieszkańców). Zinterpretować wyniki dla poszczególnych poziomów zmiennej *miasto*.

```
      Source |       SS       df       MS              Number of obs =     686
-------------+------------------------------           F( 10,   675) =   78.65
       Model |  204.248481    10  20.4248481           Prob > F      =  0.0000
    Residual |  175.299788   675   .25970339           R-squared     =  0.5381
-------------+------------------------------           Adj R-squared =  0.5313
       Total |   379.54827   685  .554085065           Root MSE      =  .50961

------------------------------------------------------------------------------
 ln_placa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     wiek |   .0311772    .0033795      9.23   0.000    .0245417    .0378128
 miasto_1 |  -.0633077   .0594773    -1.06   0.288   -.1800906    .0534751
 miasto_2 |   .1134291   .0536598     2.11   0.035     .008069    .2187893
 miasto_3 |   .3205136   .0515128     6.22   0.000     .219369    .4216583
 rodzaj_1 |   .9838117   .0623152    15.79   0.000    .8614567    1.106167
 rodzaj_2 |   .8212391   .1074394     7.64   0.000    .6102835    1.032195
     plec |  -.1136168   .0559014    -2.03   0.042   -.2233785   -.0038552
    umowa |   .3723659   .0587933     6.33   0.000    .2569261    .4878057
plecXumowa|  -.1580358    .079756    -1.98   0.048   -.3146356   -.0014361
niepublicz|   .0322562   .0476451     0.68   0.499   -.0612942    .1258066
    _cons |   4.762497   .1118487    42.58   0.000    4.542884    4.982111
```

3. W modelu uwzględniono zmienną dotyczącą czasu poszukiwania pracy (*czas*) oraz kwadrat tej zmiennej (*czas2*) i obliczono wielkość statystyk VIF.

```
      Source |       SS       df       MS              Number of obs =      66
-------------+------------------------------           F( 10,    55) =    7.89
       Model |  9.43134561    10  .943134561           Prob > F      =  0.0000
    Residual |  6.57137374    55  .119479523           R-squared     =  0.5894
-------------+------------------------------           Adj R-squared =  0.5147
       Total |  16.0027193    65  .246195682           Root MSE      =  .34566

------------------------------------------------------------------------------
 ln_placa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |   .0835472  .0148878   5.61   0.000    .0537114    .1133829
 duze_miasto |   .322613   .0980731   3.29   0.002    .1260701    .5191558
    rodzaj_1 |   .4086878  .1771117   2.31   0.025    .0537481    .7636275
    rodzaj_2 |  -.2577182  .2625386  -0.98   0.331   -.7838573    .2684208
        plec |  -.0585738  .1043441  -0.56   0.577   -.2676841    .1505366
       umowa |   .2165013  .1441093   1.50   0.139   -.0723001    .5053028
  plecXumowa |  -.3741311   1938654  -1.93   0.059    -.762646    .0143838
niepubliczny |   .2921906   .1283472  2.28   0.027     .034977    .5494042
        czas |   .0017727   .0159383  0.11   0.912   -.0301683    .0337137
       czas2 |  -.0007592   .0005032 -1.51   0.137   -.0017677    .0002493
       _cons |   3.896169   .4554267  8.55   0.000    2.983474    4.808865
------------------------------------------------------------------------------
```

Variable	VIF	1/VIF
czas	12.35	0.080972
czas2	11.30	0.088496
plecXumowa	2.45	0.408996
umowa	2.35	0.425209
rodzaj_1	1.85	0.541783
rodzaj_2	1.65	0.605326
wiek	1.63	0.615275
plec	1.50	0.665691
niepubliczny	1.35	0.738736
duze_miasto	1.28	0.778591
Mean VIF	3.77	

a) Sprawdzić czy w modelu występuje problem niedokładnej współliniowości. Jeśli występuje niedokładna współliniowość to w jaki sposób można rozwiązać ten problem?

4. Jeśli na wynagrodzenie wpływa ilość przepracowanych godzin, a estymowany jest model bez tej zmiennej to jakie będą własności estymatora MNK?

Rozwiązanie Zadanie 3

1. Na podstawie wykresu przedstawiającego zależność pomiędzy dźwignią a wystandaryzowanymi resztami podniesionymi do kwadratu podejrzenia budzą obserwacje nr 1605, 31234 i 31538.

   Na podstawie tego wykresu możemy stwierdzić, iż obserwacja nr 31358 charakteryzuje się zarówno wysoką wartością dźwigni jak i wysokimi wartościami wystandaryzowanych reszt podniesionymi do kwadratu. Znaczący wpływ na oszacowania ma dodanie obserwacji o nr 31358, która jest nietypowa ze względu na zmienną objaśniającą oraz nietypowa ze względu na zmienną objaśnianą (otrzymujemy duże reszty). Jest to obserwacja nietypowa, która nie pasuje do prostej regresji.

   Obserwacja nr 31234 cechuje się wysoką wartością dźwigni. Obserwacja ta jest co prawda nietypowa ze względu na zmienną objaśniającą, ale pasuje do linii regresji (czyli reszta dla niej jest mała).

   Obserwacja nr 1605 cechuje się wysoką wartością wystandaryzowanych reszt podniesionych do kwadratu. Dodanie obserwacji nr 1605 do próby nie powoduje znacznych zmian w oszacowanych parametrach (jest to obserwacja o stosunkowo dużej reszcie, ale nie jest nietypowa ze względu na zmienną objaśniającą).

2. miasto_1: Wynagrodzenie studenta z miejscowości do 20 tys. mieszkańców jest mniejsze o 6,3% od wynagrodzenia studenta ze wsi ceteris paribus.

   miasto_2: Wynagrodzenie studenta z miejscowości do 100 tys. mieszkańców jest większe o 12,0% $[100\% * (\exp(0{,}1134291) - 1)]$ od wynagrodzenia studenta z miejscowości do 20 tys. mieszkańców ceteris paribus.

   miasto_3: Wynagrodzenie studenta z miejscowości powyżej 100 tys. mieszkańców jest większe o 37,78% $[100\% * (\exp(0{,}3205136) - 1)]$ od wynagrodzenia studenta z miejscowości do 100 tys. mieszkańców ceteris paribus.

3. Wynik wskazuje na zbyt silną korelację zmiennych czas i czas2 (V IF>10). Nie jest to jednak wynik zaskakujący – zmienną czas2 wprowadzono do modelu w celu uwzględnienia nieliniowego wpływu czasu poszukiwania pracy na wynagrodzenia.

   Warto zauważyć, że zmienne czas i czas2 są w modelu nieistotne. Wobec tego warto przetestować hipotezę o łącznej nieistotności. W przypadku braku podstaw do odrzucenia hipotezy o łącznej nieistotności zmiennych czas i czas2 warto spróbować uwzględnić nieliniową zależność między wysokością wynagrodzenia a czasem poszukiwania pracy w inny sposób np. za pomocą dwóch funkcji liniowych „sklejonych” w punkcie wierzchołka paraboli.

4. Jest to problem zmiennej pominiętej. Estymator MNK będzie obciążony.