Egzamin z ekonometrii 14.06.2011

Pytania teoretyczne

1. Pokazać, że w modelu ze stałą suma reszt jest równa zeru.
2. Wyjaśnić, dlaczego R^2 nie można używać do porównywania modeli.
3. Wymienić założenia Klasycznego Modelu Regresji Liniowej (KMRL).
4. Za pomocą jakich testów testuje się autokorelację? Jakiemu założeniu KMRL odpowiada H_0 w tych testach? Jakie są hipotezy alternatywne w tym testach?

ZADANIE 1 Pokazać, że w modelu $y_i = \beta + \varepsilon_i$ suma kwadratów reszt $\boldsymbol{e}'\,\boldsymbol{e} = \sum_{i=1}^{N} (y_i - \overline{y})^2$

Rozwiązanie: W modelu, w którym występuje tylko stała $\boldsymbol{X} = \boldsymbol{1}$ a $\boldsymbol{b} = (\boldsymbol{1}'\boldsymbol{1})^{-1}\,\boldsymbol{1}'\boldsymbol{y} = \frac{1}{N}\sum_{i=1}^{N} y_i = \overline{y}$. Z kolei

$$e_i = y_i - x_i\,\boldsymbol{b} = y_i - \overline{y}$$

$$\boldsymbol{e}'\boldsymbol{e} = \sum_{i=1}^{N} (y_i - \overline{y})^2$$

ZADANIE 2 Na podstawie danych z Diagnozy Społecznej z roku 2009 badacz oszacował model, który wyjaśnia poziom logarytmu dochodu respondenta ($ldochód$) za pomocą logarytmu czasu pracy ($lczas\_pracy$), płci (zmienna $plec$; kodowanie: 1 - mężczyzna 1, 2 - kobieta), wieku ($wiek$), wieku do kwadratu ($wiek2$), lat nauki ($lata\_nauki$), interakcją między płcią i latami nauki ($plec\_2Xlata$ - liczba lat nauki dla kobiet, dla mężczyzn 0) oraz znajomością języka angielskiego (ang, kodowanie: 1 - czynna znajomość języka, 2 - bierna, 3 - brak znajomości). Uzyskane przez badacza wyniki regresji wraz z wynikami testów diagnostycznych znajdują się poniżej.

```
      Source |       SS       df       MS              Number of obs =   10258
-------------+------------------------------           F(  8, 10249) =  479.44
       Model |  1071.42376     8  133.92797            Prob > F      =  0.0000
    Residual |  2863.00835 10249  .279345141           R-squared     =  0.2723
-------------+------------------------------           Adj R-squared =  0.2718
       Total |  3934.43212 10257  .383585075           Root MSE      =  .52853

------------------------------------------------------------------------------
     ldochód |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 lczas_pracy |   .0134909   .0005247    25.71   0.000     .0124624    .0145195
        wiek |     .03176    .003029    10.49   0.000     .0258225    .0376975
       wiek2 |  -.0002728   .0000351    -7.78   0.000    -.0003415    -.000204
     _Iplec_2 |  -.3891494   .0467148    -8.33   0.000    -.4807195   -.2975793
  lata_nauki |   .0713448   .0027206    26.22   0.000     .0660119    .0766776
_Iplec_2Xlata|   .0116455   .0035241     3.30   0.001     .0047376    .0185533
     _Iang_2 |  -.0329431   .0185773    -1.77   0.076    -.0693582     .003472
     _Iang_3 |  -.1750962       .018    -9.73   0.000    -.2103797   -.1398128
       _cons |     5.3019   .0668522    79.31   0.000     5.170857    5.432943
------------------------------------------------------------------------------

    Variable |       VIF       1/VIF
-------------+----------------------
        wiek |     47.60    0.021011
       wiek2 |     46.55    0.021482
_Iplec_2Xlata|     22.02    0.045410
    _Iplec_2 |     19.96    0.050096
     _Iang_3 |      2.63    0.380093
  lata_nauki |      2.43    0.412268
     _Iang_2 |      1.74    0.576161
 lczas_pracy |      1.12    0.893374
-------------+----------------------
    Mean VIF |     18.01
```

Breusch-Pagan test chi2(1) = 21.81 [0.000]
Jarque - Bera test chi2(2) = 2254.93 [0.000]
RESET F(3, 115) = 5.27 [0.0012]

UWAGA: testy przeprowadzamy na poziomie istotności $\alpha = 0.05$. W nawiasach kwadratowych znajdują się wartości p. Odpowiedzi należy uzasadnić wielkościami odpowiednich statystyk testowych.

1. Zintepretuj wielkość R^2. Sprawdź, czy zmienne w modelu są łącznie istotne.
2. Wskaż zmienne nieistotne w modelu. Podaj interpretację ekonomiczną nieistotności wskazanych zmiennych.
3. Zintepretuj wielkość współczynnika przy zmiennej $lczas\_pracy$. Czy oszacowana wielkość współczynnika wydaje ci się sensowna? Odpowiedź uzasadnij.
4. Policz i zinterpretować semielastyczność cząstkową dla wieku. Średni wiek respondentów w próbie wynosi 42 lata.
5. Zintpretuj wielkość współczynnika przy zmiennej $plec\_2Xlata$.
6. O czym świadczą policzone wielkości statystyk VIF?
7. Zinterpretuj wielkości statystyk diagnostycznych.
8. Jakie są konsekwencje problemów wykrytych przez statystyki diagnostyczne? Wymień metody radzenia sobie z tym problemami.

Rozwiązanie:

1. Wielkość statystyki R^2świadczy o tym, że 27.23% zmienności zmiennej objaśnianej zostało wyjaśnione za pomocą zmienności zmiennych objaśniających. Na zadanym poziomie istotności odrzucamy hipotezę zerową o łącznej nieistotności wszystkich zmiennych (F=479, wartość p =0.000<0.05)
2. Zmienną nieistotną jest zmienna zerojedynkowa związana z bierną znajomością języka angielskiego (t=-1.77, wartość p = 0.076>0.05). Ponieważ poziomem bazowym jest czynna znajomość języka angielskiego oznacza to, że czynna i bierna znajomość języka angielskiego ma identyczny wpływ na poziom dochodu respondenta.
3. Współczynnik przy zmiennej $lczas\_p$ jest elastycznością cząstkową. Oszacowana wielkość oznacza, że wydłużenie czasu pracy o 1% powoduje wzrost zarobków o 0.01%. Oszacowanie to wydaje się zdecydowanie zbyt niskie, ponieważ sugerowuje nierealistycznie niskie wynagrodzenie za dodatkowe godziny pracy (nadgodziny).
4. Semielastyczność cząstkową dla wieku liczymy jako
$$\frac{\partial\,\mathrm{E}\,(ldoch\acute{o}d)}{\partial wiek} = 2 \times wiek \times \beta_{wiek^2} + \beta_{wiek},$$
przy czym za wiek przyjmujemy średnią w próbie dla zmiennej wiek
$$2 \times 42 \times (-.0002728) + .03176 = 0.009$$
co oznacza, że wzrost wieku i 1 rok powoduje wzrost dochodu o 0.9% w przypadku osób w wieku równym średniej w próbie.
5. Wielkość współczynnika przy zmiennej $plec\_2Xlata$ oznacza, że dochód kobiet rośnie z każdym rokiem nauki o 1.16% więcej niż rośnie dochód mężczyzn.
6. Wielkość statystyk VIF świadczy o tym, że między zmiennymi $wiek$, $wiek2$, $plec\_2Xlata$ oraz $plec$ występuje silna wspóliniowość.
7. Wielkość statystyki Breuscha-Pagana oznacza, że należy odrzucić hipotezę zerową o braku homoskedastyczności czynnika losowego i przyjąć hipotezę alternatywną o występowaniu heteroskedastyczności (21.81 wartość-p 0.000<0.05), wartość statystyki testu JB zmusza nas do odrzucenia hipotezy o normalności czynnika losowego (2254.93 wartość p 0.000<0.05). Wynik testu RESET skłania nas do odrzucenia hipotezy o liniowej formie zależności między zmienną objaśnianą a zmiennymi objaśniającymi (5.27 wartość p 0.0012<0.05)

8. Występowanie heteroskedastyczności spowoduje, że obciążony i niezgodny będzie estymator macierzy wariancji kowariancji, brak normalności czynnika losowego powoduje, że błędne będą rozkłady statystyk w małych próbach, brak liniowości może zarówno obciążyć estymmatory jak i negatywnie wpływnąć na poprawność wnioskowania statystycznego. Aby rozwiązać problem heteroskedastyczności składnika losowego można zastosować odporną macierz wariancji kowariancji lub UMNK, problem nieliniowości formy funkcyjnej i braku normalności składnika losowgo da się często rozwiązać modyfikując formę funkcyjną modelu.

ZADANIE 3 Oszacowano MNK następujący model:

$$c_t = \beta_0 + \beta_1 y_t + \beta_2 p_t + \varepsilon_t,$$

w którym: c_i jest logarytmem nominalnych wydatków konsumpcyjnych gospodarstwa domowego, y_i jest logarytmem jego nominalnego dochodu, a p_i logarytmem indeksu cen żywności. Otrzymano następujące wyniki dla 66 obserwacji:

zmienna	współczynnik	błąd standardowy
stała	6.0	4.0
y	5.0	2.0
p	2.0	1.0

$RSS = 256$ i macierz wariancji-kowariancji estymatorów:

$$\begin{bmatrix} 16 & 3 & 5 \\ 3 & 4 & 5 \\ 5 & 5 & 1 \end{bmatrix}.$$

1. Oblicz prognozę dla c_{t+1} jeżeli $y_{t+1} = 2$, a $p_{t+1} = 3$.
2. Oblicz wariancję prognozy.
3. Oblicz wariancję błędu prognozy.

Rozwiązanie:

1. Prognoza wynosi:
$$\widehat{c}_{t+1} = b_0 + b_1 y_{t+1} + b_2 p_{t+1} = 22.$$
2. Nieobciążony estymator wariancji błędu losowego jest równy:
$$s^2 = \frac{RSS}{T-k} = 4$$
Wariancja prognozy
$$\operatorname{Var}(\widehat{c}_{t+1}) = \boldsymbol{x}'_{t+1} \operatorname{Var}(\boldsymbol{b}) \boldsymbol{x}_{t+1} = 143$$
Wariancja błędu prognozy
$$\operatorname{Var}(c_{t+1} - \widehat{c}_{t+1}) = \boldsymbol{x}'_{t+1} \operatorname{Var}(\boldsymbol{b}) \boldsymbol{x}_{t+1} + s^2 = 147$$