Metoda najmniejszych kwadratów

Własności algebraiczne

Model liniowy

- Zapis modelu

$$zarobki = \beta_0 + \beta_1 * plec + \beta_2 * wiek + \varepsilon$$

- Oszacowania wartości współczynników

$$zarobki = b_0 + b_1 * plec + b_2 * wiek + e$$

Model liniowy

Tabela: Oszacowania współczynników

zarobki	współczynnik	błąd oszacowania
płeć	-96,78	2,81
wiek	3,17	0,15
stała	316,21	5,86
Liczba obserwacji	16162	

Źródło: Obliczenia własne na podstawie BAEL (2000)

Model logliniowy

- Zapis modelu

$$\ln(zarobki) = \beta_0 + \beta_1 * plec + \beta_2 * wiek + \varepsilon$$

- Oszacowania wartości współczynników

$$\ln(zarobki) = b_0 + b_1 * plec + b_2 * wiek + e$$

Model logliniowy

Tabela: Oszacowania współczynników

ln(zarobki)	współczynnik	błąd oszacowania
płeć	-0,23	0,006
wiek	0,01	0,001
stała	5,70	0,012
Liczba obserwacji	16162	

Źródło: Obliczenia własne na podstawie BAEL (2000)

- Własności każdego modelu szacowanego MNK
  1. wektor reszt jest ortogonalny do macierzy obserwacji
  2. wektor reszt jest ortogonalny do wektora wartości dopasowanych dla zmiennej objaśnianej
- Własności wyłącznie dla modelu ze stałą
  3. suma reszt jest równa zero
  4. średnia wartość teoretyczna (dopasowana) jest równa średniej wartości empirycznej

Macierz idempotentna

Macierz idempotentna M to taka macierz, że

$$M^2 = MM = M$$

- W ekonometrii literą M jest oznaczana szczególna macierz idempotentna

$$M = I - X(X'X)^{-1}X'$$

Macierz rzutu

Macierz Projekcji (rzutu)

$$P = I - M = X(X'X)^{-1}X'$$

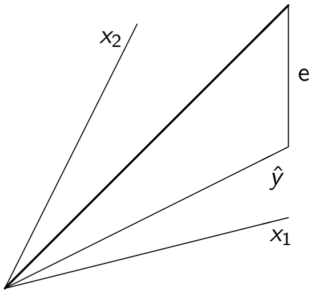
- Własności

$$MX = 0$$

$$PX = X$$

$$MP = PM = 0$$

Rysunek: Rzut wektora y na przestrzeń X



Analiza wariancji

- Równość analizy wariancji

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

- Zapis skrócony

$$TSS = ESS + RSS$$

Analiza wariancji

Tabela: Dekompozycja Wariancji

Wariancja	Suma kwadratów
Wyjaśniona	103115934
Resztowa	461687875
Całkowita	564803809

Źródło: Obliczenia własne na podstawie BAEL (2000)

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{\sum(y_i - Xb)^2}{\sum(y_i - \bar{y})^2}$$

$$\bar{R}^2 = 1 - \frac{n-1}{n-K}(1 - R^2)$$

- Model A: $\mathit{zarobki} = f(\text{płeć}, \text{wiek})$
- Model B: $\mathit{zarobki} = f(\text{płeć}, \text{wiek}, \text{wiek}^2)$

Tabela: Porównanie modeli

Model	TSS	RSS
A	2554.23525	2263.59451
B	2554.23525	2239.21981

Źródło: Obliczenia własne na podstawie BAEL (2000)

Tabela: Skorygowany współczynnik determinacji

Model	R^2	k	$\bar{R}^2$
A	0.0914	2	$\bar{R}^2 = 1 - \frac{16161}{16159}(1 - 0.0914) = 0.0913$
B	0.1138	3	$\bar{R}^2 = 1 - \frac{16161}{16158}(1 - 0.1138) = 0.1137$

Źródło: Obliczenia własne na podstawie BAEL (2000)

```
      Source |       SS       df       MS              Number of obs =   16162
-------------+------------------------------           F(  3, 16158) =  757.71
       Model |  315.015445     3  105.005148           Prob > F      =  0.0000
    Residual |  2239.21981 16158  .138582733           R-squared     =  0.1233
-------------+------------------------------           Adj R-squared =  0.1232
       Total |  2554.23525 16161  .158049332           Root MSE      =  .37227

------------------------------------------------------------------------------
    lzarobki |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Iplec_2 |  -.2335237   .0058857   -39.68   0.000    -.2450604    -.221987
        wiek |   .0356602   .0021362    16.69   0.000      .031473    .0398473
       wiek2 |  -.0003642   .0000275   -13.26   0.000     -.000418   -.0003103
       _cons |   5.197807   .0398962   130.28   0.000     5.119606    5.276008
------------------------------------------------------------------------------
```