# Interakcje

# Klasyczny Model Regresji Liniowej cz.I

**Stanisław Cichocki**

**Natalia Nehrebecka**

Wykład 7

# Plan wykładu

- 1. Interakcje
- 2. Przybliżanie modeli nieliniowych
- 3. Założenia KMRL

# Plan wykładu

- 1. Interakcje
- 2. Przybliżanie modeli nieliniowych
- 3. Założenia KMRL

# Modele z interakcjami

- W standardowym modelu liniowym zakładamy, że wpływ poszczególnych zmiennych niezależnych na oczekiwaną wartość zmiennej zależnej jest **addytywny**.
- W ramach modelu liniowego można także uwzględnić efekt krzyżowego wzmacniania się efektów poszczególnych zmiennych.
- Efekt ten zachodzi, gdy siła oddziaływania jednej zmiennej niezależnej jest uwarunkowana wielkością innych zmiennych niezależnych.
- Ten efekt można uwzględnić, wstawiając do modelu iloczyny zmiennych (interakcje).

# Modele z interakcjami

$$y_i = \beta_1 + x_{2i}\beta_2 + \ldots + x_{Ki}\beta_K + \sum_{r=1}^{K}\sum_{s=1}^{r-1}\gamma_{rs}x_{ri}x_{si} + \varepsilon_i$$

Pochodną cząstkową warunkowej wartości oczekiwanej po $x_{Ki}$ :

$$\frac{\partial E(y_i)}{\partial x_k} = \beta_k + \sum_{r=1, r\neq k}^{K}\gamma_{rk}x_{ri}$$

# Interakcje między zmiennymi dyskretnymi

- Interakcje między zmiennymi zerojedynkowymi bierzemy pod uwagę, jeśli wpływ poszczególnych zmiennych nie jest addytywny.

- **Sytuacja taka może wystąpić, jeśli pewne kombinacje charakterystyk jakościowych wpływają na zmienną zależną bardziej lub mniej, niż wynikałoby z wpływu poszczególnych zmiennych.**

- Np.
- Zmienna zależna: płaca
- Zmienne niezależna wiek, płeć, wykształcenie, interakcja: płećXwykształcenie
- Do modelu wprowadzamy interakcje, ponieważ spodziewamy się, iż wpływ zmiennej oznaczającej wykształcenie zależy od płci.

# INTERAKCJE MIĘDZY ZMIENNYMI DYSKRETNYMI - WYKSZTAŁCENIE I PŁEĆ

**placa- zmienna zależna,**
**wiek, plec oraz interakcje między wykształceniem i płcią - zmienne niezależne**

```
xi: regress placa wiek i.plec*i.wykszt

      Source |       SS       df       MS              Number of obs =    9687
-------------+------------------------------           F( 10,  9676) =  350.13
       Model |  1.0280e+09    10   102803871           Prob > F      =  0.0000
    Residual |  2.8410e+09  9676  293617.286           R-squared     =  0.2657
-------------+------------------------------           Adj R-squared =  0.2649
       Total |  3.8691e+09  9686   399450.71           Root MSE      =  541.86

------------------------------------------------------------------------------
       placa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |   12.94226   .5317025    24.34   0.000     11.90001    13.98451
     _Iplec_1 |  -493.8996   25.95657   -19.03   0.000    -544.7799   -443.0193
   _Iwykszt_2 |  -727.0777   35.19716   -20.66   0.000    -796.0715   -658.0839
   _Iwykszt_3 |  -673.6041   25.35675   -26.57   0.000    -723.3086   -623.8995
   _Iwykszt_4 |  -871.1952   23.29587   -37.40   0.000      -916.86   -825.5305
   _Iwykszt_5 |  -1086.103   31.84348   -34.11   0.000    -1148.523   -1023.683
_IpleXwyk_~2 |   283.3427   43.05827     6.58   0.000     198.9395    367.7459
_IpleXwyk_~3 |   220.2259   33.77979     6.52   0.000     154.0104    286.4413
_IpleXwyk_~4 |    212.681   33.03096     6.44   0.000     147.9334    277.4286
_IpleXwyk_~5 |   297.2812   46.36358     6.41   0.000     206.3989    388.1635
       _cons |   1388.519   29.31611    47.36   0.000     1331.053    1445.984
------------------------------------------------------------------------------
```

# Interakcje między zmiennymi dyskretnymi i ciągłymi

- Wprowadzenie do modelu interakcji pomiędzy zmiennymi dyskretnymi i ciągłymi ma sens, jeśli **wpływ pewnej zmiennej niezależnej ciągłej na zmienną zależną zależy od poziomów zmiennej dyskretnej**.

# INTERAKCJE MIĘDZY ZMIENNĄ CIĄGŁA I DYSKRETNĄ - WIEK I MIEJSCE ZAMIESZKANIA

## interakcje między zmienną klm1 a wiekiem

```
xi: regress placa i.klm1*wiek

      Source |       SS       df       MS              Number of obs =    8776
-------------+------------------------------           F(  5,  8770) =  115.75
       Model |   212980427     5  42596085.4           Prob > F      =  0.0000
    Residual |  3.2273e+09  8770  367992.845           R-squared     =  0.0619
-------------+------------------------------           Adj R-squared =  0.0614
       Total |  3.4403e+09  8775  392054.436           Root MSE      =  606.62

------------------------------------------------------------------------------
       placa |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Iklm1_1 |  -63.04935   66.61419    -0.95   0.344    -193.6288    67.53008
    _Iklm1_2 |   -158.479   63.10371    -2.51   0.012     -282.177   -34.78088
        wiek |   12.61045   1.220073    10.34   0.000     10.21882    15.00208
_IklmXwiek_1 |  -2.419013    1.62517    -1.49   0.137    -5.604726    .7667008
_IklmXwiek_2 |  -2.062711   1.560181    -1.32   0.186    -5.121032    .9956097
       _cons |   743.0838   50.34411    14.76   0.000     644.3976    841.7701
------------------------------------------------------------------------------
```

# Plan wykładu

- 1.Interakcje
- 2. Przybliżanie modeli nieliniowych
- 3. Założenia KMRL

# Modele wielomianowe

- Nieliniowa zależność między ***y*** a ***x*** można przybliżyć za pomocą modelu liniowego stosując model:
- **Model wielomianowy**

$$y_i = \beta_0 + x_i\beta_1 + x_i^2\beta_2 + \ldots + x_i^k\beta_K + \varepsilon_i$$

- Przy większej liczbie zmiennych objaśniających wstawia się do modelu ich kwadraty i iloczyny

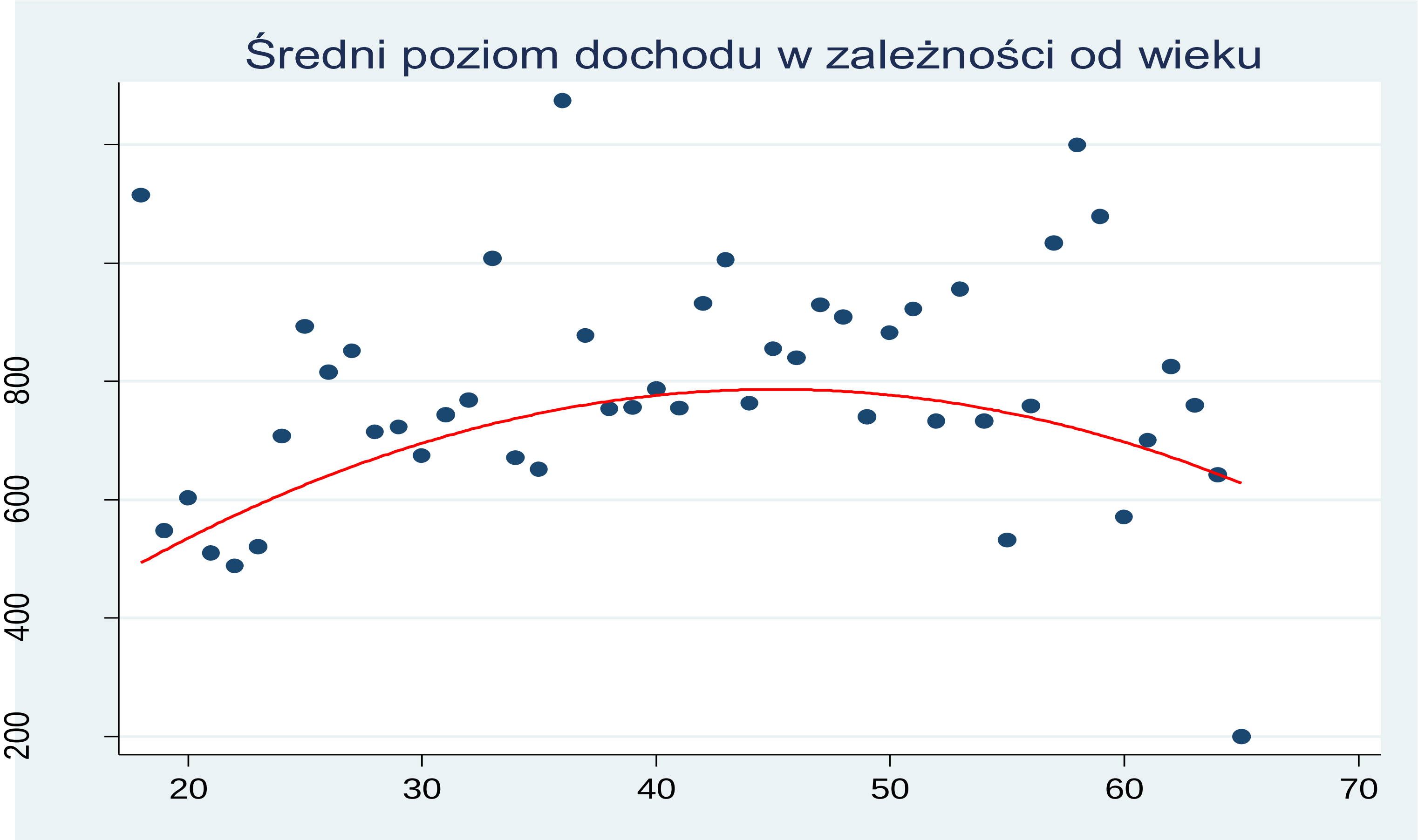
# INNE FORMY FUNKCYJNE MODELU ZE WZGLĘDU NA WIEK - WIELOMIAN STOPNIA II

```
. regress dochod wiek wiek_2 plec srednie wyzsze

      Source |       SS       df       MS              Number of obs =    1083
-------------+------------------------------           F(  5,  1077) =   22.98
       Model |  72048793.8     5  14409758.8           Prob > F      =  0.0000
    Residual |   675432341  1077  627142.378           R-squared     =  0.0964
-------------+------------------------------           Adj R-squared =  0.0922
       Total |   747481135  1082  690832.842           Root MSE      =  791.92

------------------------------------------------------------------------------
      dochod |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        wiek |   36.06131   15.48328     2.33   0.020     5.680494    66.44212
      wiek_2 |  -.3998842   .1973767    -2.03   0.043    -.7871707   -.0125977
        plec |  -338.0671   48.25867    -7.01   0.000    -432.7588   -243.3755
     srednie |   208.5538   77.72619     2.68   0.007     56.04182    361.0657
      wyzsze |   708.2862   99.55596     7.11   0.000     512.9406    903.6318
       _cons |  -26.64989   298.3288    -0.09   0.929    -612.0215    558.7217
------------------------------------------------------------------------------
```

# INNE FORMY FUNKCYJNE MODELU ZE WZGLĘDU NA WIEK - WIELOMIAN STOPNIA II



# Uwaga!

Praca na ćwiczeniach:

- Model schodkowy
- Model krzywej łamanej

# Plan wykładu

- 1.Interakcje
- 2. Przybliżanie modeli nieliniowych
- 3. Założenia KMRL

# Klasyczny model regresji liniowej

- Na poprzednich wykładach pokazaliśmy, iż estymator MNK daje oszacowania parametrów, które są najlepiej dopasowane do danych
- Obecnie zajmiemy się własnościami statystycznymi tego estymatora i w tym celu przyjmujemy pewne dodatkowe założenia
- Najprostszym i najpopularniejszym układem założeń jest KMRL

# Założenia klasycznego modelu regresji liniowej

- 1. Związek pomiędzy zmienną zależną a zmiennymi niezależnymi opisany jest równaniem:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_K x_{Ki} + \varepsilon_i \qquad i = 1,2,3 \cdot\cdot n$$

- 2. Zmienne objaśniające $x_{2i}, x_{3i}, \ldots, x_{Ki}$ są nielosowe dla $i = 1,2,3 \cdot\cdot n$
- 3. Wartość oczekiwana błędu losowego jest równa zeru:

$$E(\varepsilon) = \mathbf{0}$$

- Pozostałe założenia zostaną omówione na kolejnym wykładzie.

# Liniowość w modelu

- Liniowość w modelu względem:
- Po pierwsze **- zmiennych objaśniających** $X_{Ki}$, które są w pierwszej potędze,

$$y_i = \beta_1 + \beta_2 X_{2i}^{1} + \ldots + \beta_K X_{Ki}^{1} + \varepsilon_i$$

- a po drugie - **względem parametrów** $\beta_k$, które są również w pierwszej potędze.

$$y_i = \beta_1^{1} + \beta_2^{1} X_{2i} + \ldots + \beta_K^{1} X_{Ki} + \varepsilon_i$$

# Pytania teoretyczne

1. Wyjaśnić, co to znaczy, że między zmiennymi w modelu występują interakcje.
2. Opisać sposoby przybliżania zależności nieliniowej za pomocą modelu liniowego.

# Dziękuję za uwagę