

Interakcje

Przybliżanie modeli nieliniowych

Stanisław Cichocki

Natalia Nehrebecka

Wykład 8

Plan wykładu

- ▶ 1. Interakcje
- ▶ 2. Przybliżanie modeli nieliniowych

Plan wykładu

- ▶ 1. Interakcje
- ▶ 2. Przybliżanie modeli nieliniowych

Modele z interakcjami

- ▶ W standardowym modelu liniowym zakładamy, że wpływ poszczególnych zmiennych niezależnych na oczekiwaną wartość zmiennej niezależnej 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 iloczyn zmiennych (interakcje).

Modele z interakcjami

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

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 niezależne: wiek, płeć, wykształcenie, interakcja: $\text{płeć} \times \text{wykształcenie}$
- ▶ 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
Model	1.0280e+09	10	102803871	F(10, 9676)	=	350.13
Residual	2.8410e+09	9676	293617.286	Prob > F	=	0.0000
Total	3.8691e+09	9686	399450.71	R-squared	=	0.2657
				Adj R-squared	=	0.2649
				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 CIĄGŁĄ 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
Model	212980427	5	42596085.4	F(5, 8770)	=	115.75
Residual	3.2273e+09	8770	367992.845	Prob > F	=	0.0000
Total	3.4403e+09	8775	392054.436	R-squared	=	0.0619
				Adj R-squared	=	0.0614
				Root MSE	=	606.62

placa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_Ikml1_1	-63.04935	66.61419	-0.95	0.344	-193.6288	67.53008
_Ikml1_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
_IkmlXwiek_1	-2.419013	1.62517	-1.49	0.137	-5.604726	.7667008
_IkmlXwiek_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

Modele wielomianowe

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

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

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

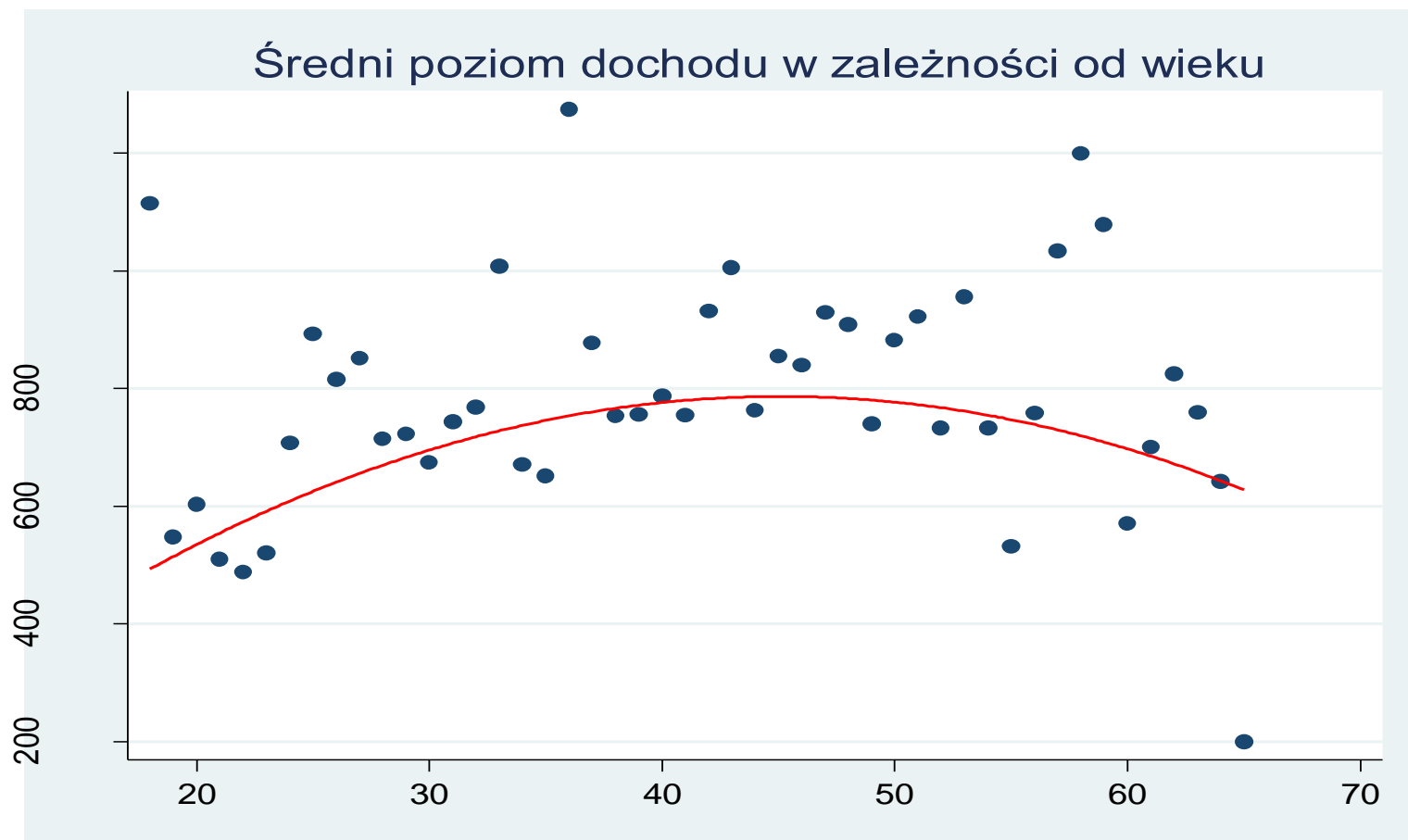
INNE FORMY FUNKCYJNE MODELU ZE WZGLĘDU NA WIEK - WIELOMIAN STOPNIA II

```
. regress dochod wiek wiek_2 plec srednie wyzsze
```

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



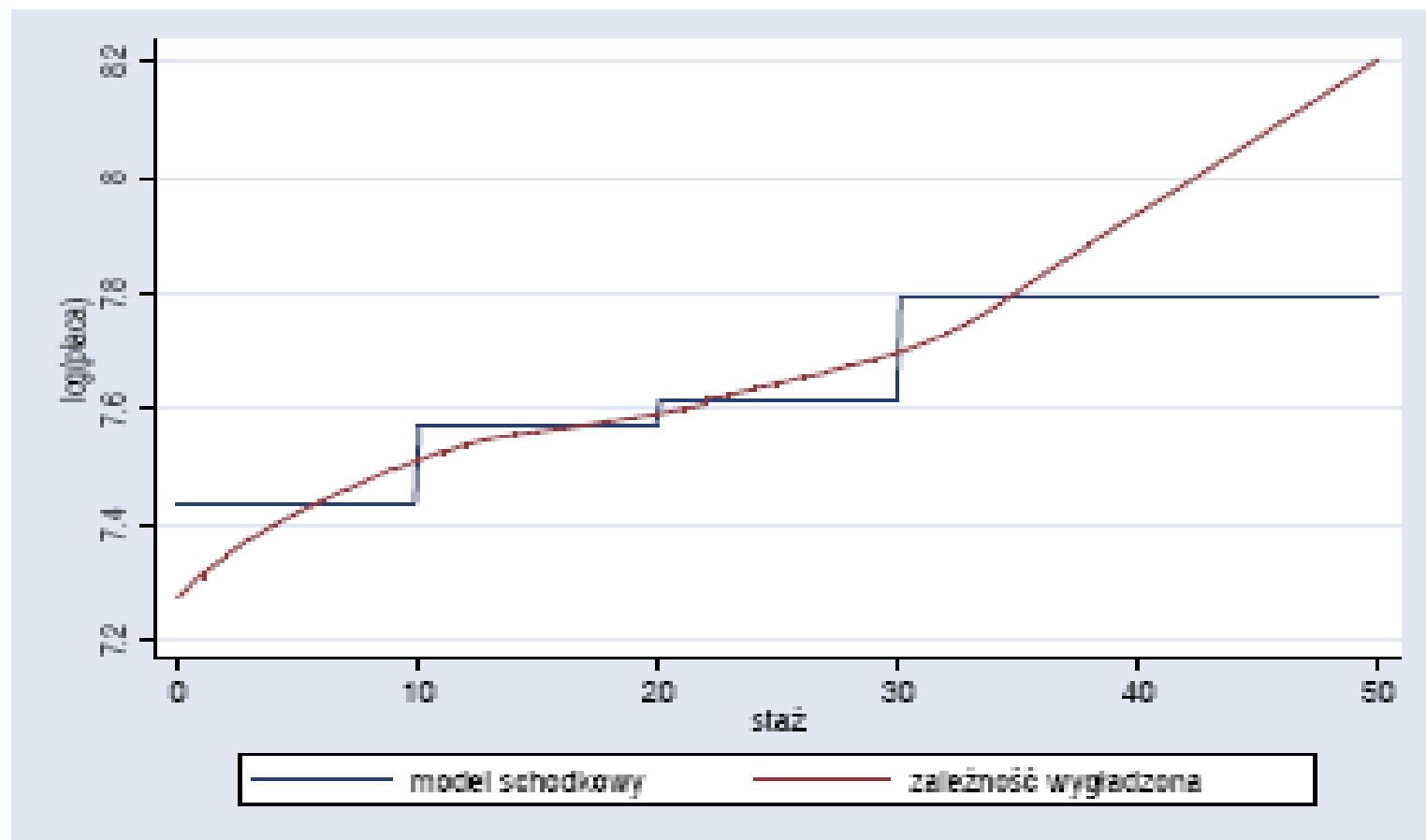
Model schodkowy

- ▶ Nieliniowa zależność między y a x można przybliżyć za pomocą modelu liniowego stosując model:
- ▶ **2. Model schodkowy**
- ▶ W tym przypadku definiujemy **zmienne zerojedynkowe**

$$D_{s,i} = \mathbb{I}(x_s^* < x_i \leq x_{s+1}^*)$$

- ▶ związane z przedziałami x_i i
- ▶ przeprowadzamy regresję na tych zmiennych zamiast na x_i

Model schodkowy



Model schodkowy

```
generate wiek_2 = (wiek > 25 & wiek <= 35)
generate wiek_3 = (wiek > 35 & wiek <= 45)
generate wiek_4 = (wiek > 45 & wiek <= 55)
generate wiek_5 = (wiek > 55)
```

```
regress dochod wiek_?
```

Source	SS	df	MS
Model	6403953.56	4	1600988.39
Residual	741077182	1078	687455.642
Total	747481135	1082	690832.842

```
Number of obs =      1083
F(   4,   1078) =      2.33
Prob > F       =    0.0544
R-squared      =    0.0086
Adj R-squared  =    0.0049
Root MSE      =    829.13
```

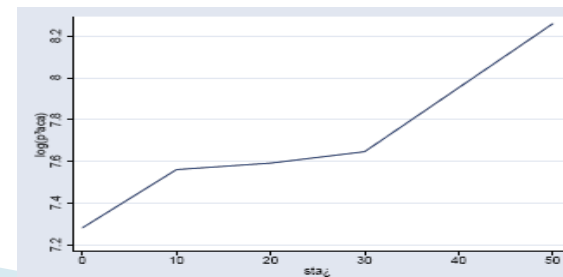
dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiek_2	126.6784	88.27104	1.44	0.152	-46.52407	299.881
wiek_3	239.7376	84.81751	2.83	0.005	73.31151	406.1637
wiek_4	206.697	91.38316	2.26	0.024	27.388	386.006
wiek_5	175.5193	141.5618	1.24	0.215	-102.2486	453.2873
_cons	639.0551	73.57334	8.69	0.000	494.6919	783.4183

Model krzywej łamanej

- ▶ Nieliniowa zależność między y a x można przybliżyć za pomocą modelu liniowego stosując model:
- ▶ **3. Model krzywej łamanej**

$$y = \begin{cases} \alpha + \beta_1 x + \varepsilon & \text{dla } x_i \leq x_1^* \\ \alpha + \beta_1 x + \beta_2 (x - x_1^*) + \varepsilon & \text{dla } x_1^* < x_i \leq x_2^* \\ \vdots \\ \alpha + \beta_1 x + \sum_{j=2}^{s-1} \beta_j (x_j^* - x_{j-1}^*) + \beta_s (x - x_s^*) + \varepsilon & \text{dla } x_i > x_s^* \end{cases}$$

- ▶ Zależność nieliniowa przybliżona jest w tym przypadku krzywą, którą można zilustrować rysunkiem:



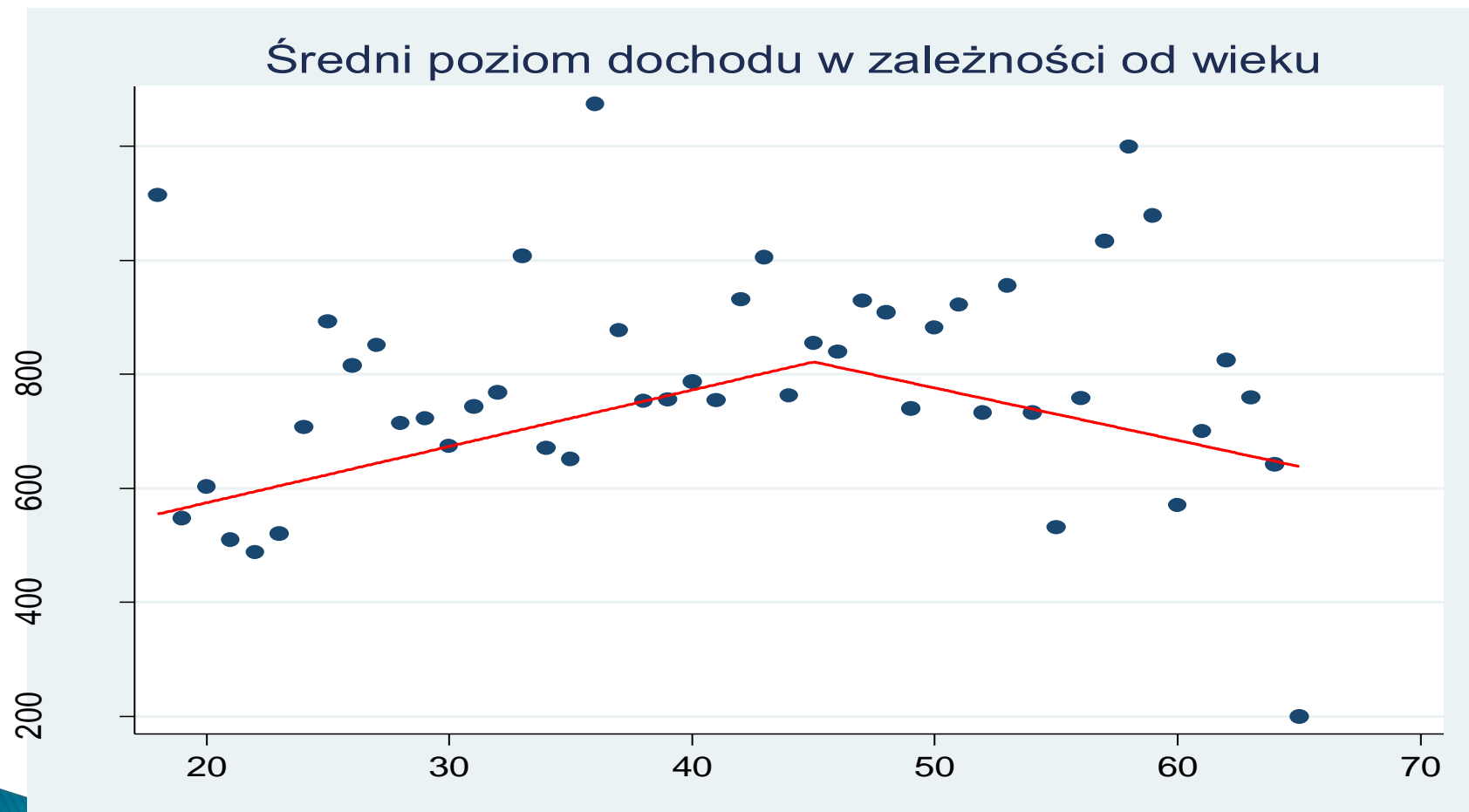
Model krzywej łamanej - DWIE LINIOWE FUNKCJE SKLEJONE W 45

```
regress dochod wiek wiek_45 plec srednie wyzsze
```

Source	SS	df	MS	Number of obs	=	1083
Model	71889880.6	5	14377976.1	F(5, 1077)	=	22.92
Residual	675591255	1077	627289.93	Prob > F	=	0.0000
				R-squared	=	0.0962
				Adj R-squared	=	0.0920
				Root MSE	=	792.02
Total	747481135	1082	690832.842			

dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiek	9.892845	3.449602	2.87	0.004	3.124143	16.66155
wiek_45	-19.06609	9.716528	-1.96	0.050	-38.13156	-.0006177
plec	-338.9919	48.27437	-7.02	0.000	-433.7144	-244.2694
srednie	211.058	77.6635	2.72	0.007	58.66912	363.447
wyzsze	712.6863	99.4661	7.17	0.000	517.517	907.8556
_cons	376.4752	145.4995	2.59	0.010	90.98058	661.9698

Model krzywej łamanej



Pytania teoretyczne

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

Dziękuję za uwagę