

# Wybór formy funkcyjnej (CZ. IV)

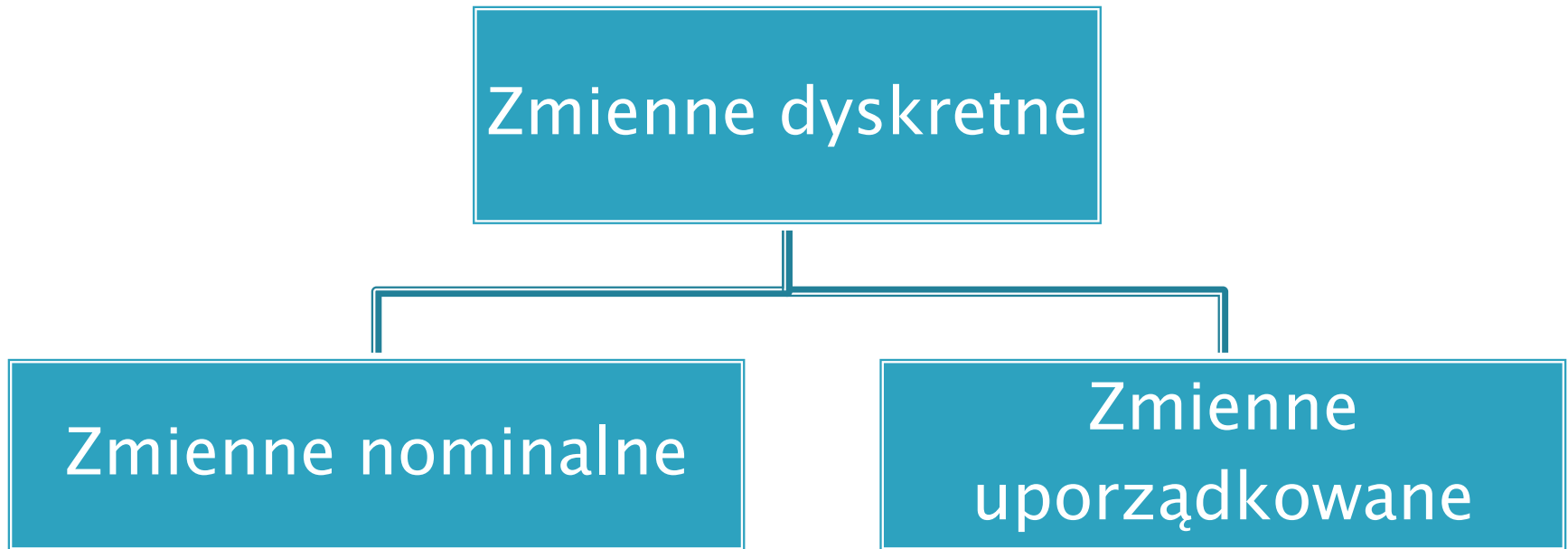
**Natalia Nehrebecka**  
**Stanisław Cichocki**

Wykład 7

# Plan zajęć

1. Zmienne dyskretne
2. Modele z interakcjami
3. Przybliżanie modeli nieliniowych

# Zmienne dyskretne



# Zmienne dyskretne

- ▶ Nieco bardziej skomplikowana jest sytuacja, gdy mamy do czynienia ze zmienną dyskretną która przyjmuje więcej niż 2 wartości.
- ▶ np. **wykształcenie** (1 – podstawowe, 2 – średnie, 3 - wyższe)
- ▶ W tym przypadku do każdego poziomu  $s$  zmiennej dyskretnej  $X_i$  musimy przypisać jedną zmienną zero-jedynkową  $D_{s,i}$

$$D_{s,i} = 1 \text{ gdy } X_i = s$$

$$D_{s,i} = 0 \text{ gdy } X_i \neq s \text{ dla } s = 1, 2, \dots, S$$

# Przykład

$$\text{wykształcenie}_i = \begin{cases} 1 & \text{podstawowe} \\ 2 & \text{średnie} \\ 3 & \text{wyższe} \end{cases}$$

$$\text{podstawowe}_i = \begin{cases} 1 & \text{podstawowe} \\ 0 & \text{w p.p.} \end{cases}$$

$$\text{średnie}_i = \begin{cases} 1 & \text{średnie} \\ 0 & \text{w p.p.} \end{cases}$$

$$\text{wyższe}_i = \begin{cases} 1 & \text{wyższe} \\ 0 & \text{w p.p.} \end{cases}$$

# Zmienne dyskretne

- ▶ Za **poziom bazowy** uznajemy jeden z poziomów (np. poziom 1), i zmienną zero-jedynkową związaną z tym poziomem usuwamy z modelu ze stałą.
- ▶ Np. dla zmiennej wykształcenie
- ▶ Poziom bazowy: wykształcenie podstawowe

$$placa_i = \beta_1 + \beta_2 \acute{s}rednie_i + \beta_3 wyzsze_i + \varepsilon_i$$

- ▶ **Dlaczego?**
- ▶ Nie jest możliwe, by w modelu była jednocześnie stała i wszystkie zmienne zero-jedynkowe (dla każdego poziomu zmiennej dyskretnej), ponieważ macierz  $\mathbf{X}^T\mathbf{X}$  byłaby osobliwa!

# Interpretacja parametrów przy zmiennych dyskretnych

- ▶ Interpretacja współczynników przy zmiennych 0-1 jest analogiczna jak w przypadku modelu z jedną tylko taką zmienną:

dany współczynnik opisuje różnicę między oczekiwaną wartością  $y_i$  dla respondenta o **charakterystyce bazowej** i dla respondenta charakterystyce **s**.

# Przykład – regresja ze zmiennymi dyskretnymi

Zmienna zależna: **dochod**

Zmienne niezależne: **wiek, wiek\_2, miasto**

miasto	Freq.	Percent	Cum.
<b>1 - wieś</b>	<b>323</b>	<b>29.82</b>	<b>29.82</b>
2 - miasto do 25tyś	194	17.91	47.74
3 - miasto od 25tyś do 250tyś	356	32.87	80.61
4 - miasto powyżej 250tyś	210	19.39	100.00
Total	1,083	100.00	



# Przykład

```
xi:regres dochod wiek wiek_2 i.miasto
```

```
i.miasto      _Imiasto 1-4      (naturally coded; _Imiasto_1_1 omitted)
```

Source	SS	df	MS	Number of obs =	1083
Model	23872603.5	5	4774520.71	F( 5, 1077) =	7.11
Residual	723608532	1077	671874.217	Prob > F =	0.0000
-----				R-squared =	0.0319
Total	747481135	1082	690832.842	Adj R-squared =	0.0274
-----				Root MSE =	819.68

dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiek	37.8833	16.01033	2.37	0.018	6.468336	69.29827
wiek_2	-.4486477	.2039518	-2.20	0.028	-.8488356	-.0484597
<b>_Imiasto_2</b>	<b>158.2807</b>	<b>74.50027</b>	<b>2.12</b>	<b>0.034</b>	<b>12.0986</b>	<b>304.4629</b>
<b>_Imiasto_3</b>	<b>265.9892</b>	<b>63.10008</b>	<b>4.22</b>	<b>0.000</b>	<b>142.1762</b>	<b>389.8022</b>
<b>_Imiasto_4</b>	<b>345.5604</b>	<b>73.00193</b>	<b>4.73</b>	<b>0.000</b>	<b>202.3182</b>	<b>488.8025</b>
_cons	-119.8138	303.7319	-0.39	0.693	-715.7871	476.1596

# Przykład

```
xi:regres ln_dochod wiek wiek_2 i.miasto
```

```
i.miasto      _Imiasto_1-4      (naturally coded; _Imiasto_1 omitted)
```

Source	SS	df	MS	Number of obs =	1083
Model	32.2360851	5	6.44721702	F( 5, 1077) =	20.57
Residual	337.543087	1077	.31341048	Prob > F =	0.0000
Total	369.779172	1082	.341755242	R-squared =	0.0872
				Adj R-squared =	0.0829
				Root MSE =	.55983

ln_dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiek	.0491115	.0109349	4.49	0.000	.0276554	.0705675
wiek_2	-.0006039	.0001393	-4.34	0.000	-.0008772	-.0003306
<b>_Imiasto_2</b>	<b>.2345958</b>	<b>.0508828</b>	<b>4.61</b>	<b>0.000</b>	<b>.1347553</b>	<b>.3344364</b>
<b>_Imiasto_3</b>	<b>.2870446</b>	<b>.0430966</b>	<b>6.66</b>	<b>0.000</b>	<b>.2024818</b>	<b>.3716073</b>
<b>_Imiasto_4</b>	<b>.4162679</b>	<b>.0498594</b>	<b>8.35</b>	<b>0.000</b>	<b>.3184353</b>	<b>.5141005</b>
_cons	5.351579	.2074451	25.80	0.000	4.944537	5.758622

# UWAGA!

- ▶ Samodzielnie:
  - Efekty progowe
  - Kontrasty w odchyleniach

# 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 iloczyn zmiennych (interakcje).

# Interakcje między zmiennymi zerojedynkowymi

- ▶ 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: dochód
  - Zmienne niezależne: płeć (0 – mężczyzna; 1 - kobieta), wykształcenie (0 – podstawowe; 1- średnie; 2 - wyższe), interakcja: płećXwykształcenie
- ▶ Do modelu wprowadzamy interakcje, ponieważ spodziewamy się, iż wpływ zmiennej oznaczającej wykształcenie zależy od płci.

# INTERAKCJE MIĘDZY ZMIENNYMI DYSKRETNymi - WYKSZTAŁCENIE I PŁEĆ

dochod - zmienna zależna,

wiek, wiek\_2 oraz interakcje między wykształceniem i płcią - zmienne niezależne

```
xi: regress dochod wiek wiek_2 i.plec*i.wyksztalzenie
```

Source	SS	df	MS	Number of obs = 1083	
Model	81648217.6	7	11664031.1	F( 7, 1075)	= 18.83
Residual	665832918	1075	619379.458	Prob > F	= 0.0000
-----				R-squared	= 0.1092
-----				Adj R-squared	= 0.1034
Total	747481135	1082	690832.842	Root MSE	= 787.01

dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wiek	36.38318	15.39846	2.36	0.018	6.168745	66.59762
wiek_2	-.4049352	.1962222	-2.06	0.039	-.7899572	-.0199131
<b>_Iplec_1</b>	<b>-144.4044</b>	143.4615	-1.01	0.314	-425.9008	137.0919
<b>_Iwyksztal~2</b>	<b>274.2703</b>	105.1538	2.61	0.009	67.94046	480.6002
<b>_Iwyksztal~3</b>	<b>1040.998</b>	137.1701	7.59	0.000	771.8461	1310.149
<b>_IpleXwyk_~2</b>	<b>-143.4455</b>	<b>153.4394</b>	<b>-0.93</b>	<b>0.350</b>	<b>-444.5201</b>	<b>157.6292</b>
<b>_IpleXwyk_~3</b>	<b>-682.341</b>	<b>197.7395</b>	<b>-3.45</b>	<b>0.001</b>	<b>-1070.34</b>	<b>-294.3418</b>
_cons	-121.1625	300.6773	-0.40	0.687	-711.1435	468.8184

# 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 miasto a wiekiem

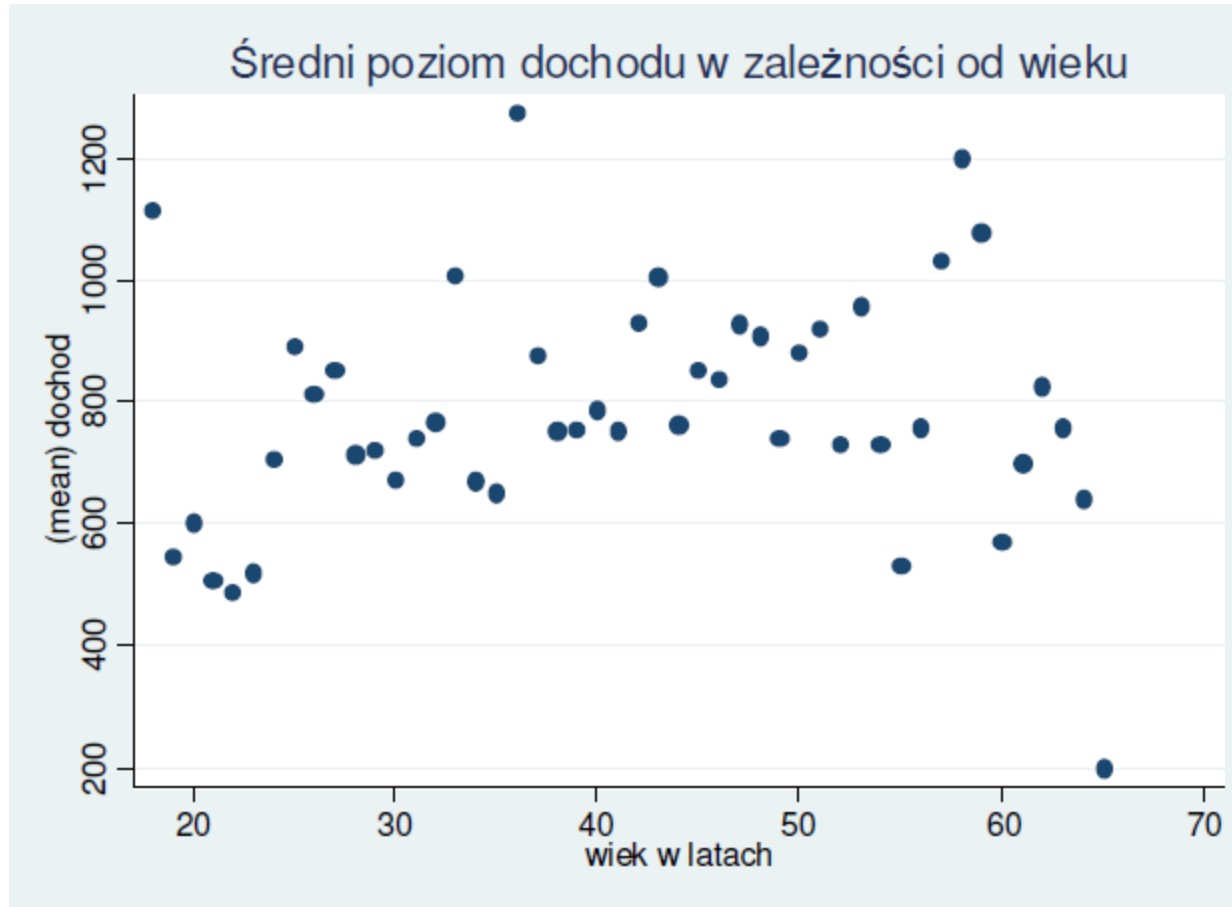
```
xi: regress dochod i.miasto_1*wiek
```

Source	SS	df	MS	Number of obs =	1083
Model	21268278.5	7	3038325.5	F( 7, 1075) =	4.50
Residual	726212857	1075	675546.843	Prob > F =	0.0001
				R-squared =	0.0285
				Adj R-squared =	0.0221
Total	747481135	1082	690832.842	Root MSE =	821.92

dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_Imiasto_1_2	28.34615	296.4254	0.10	0.924	-553.2919	609.9842
_Imiasto_1_3	53.41383	249.026	0.21	0.830	-435.2183	542.046
_Imiasto_1_4	135.6545	283.6069	0.48	0.633	-420.8315	692.1404
wiek	-.4870689	4.569159	-0.11	0.915	-9.452549	8.478412
<b>_ImiaXwiek_2</b>	<b>3.588019</b>	<b>7.603214</b>	<b>0.47</b>	<b>0.637</b>	<b>-11.3308</b>	<b>18.50684</b>
<b>_ImiaXwiek_3</b>	<b>5.698882</b>	<b>6.355967</b>	<b>0.90</b>	<b>0.370</b>	<b>-6.772626</b>	<b>18.17039</b>
<b>_ImiaXwiek_4</b>	<b>5.396286</b>	<b>7.063888</b>	<b>0.76</b>	<b>0.445</b>	<b>-8.464285</b>	<b>19.25686</b>
_cons	641.7219	175.9821	3.65	0.000	296.4145	987.0292



# Zależność między $y$ a $x$



# 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

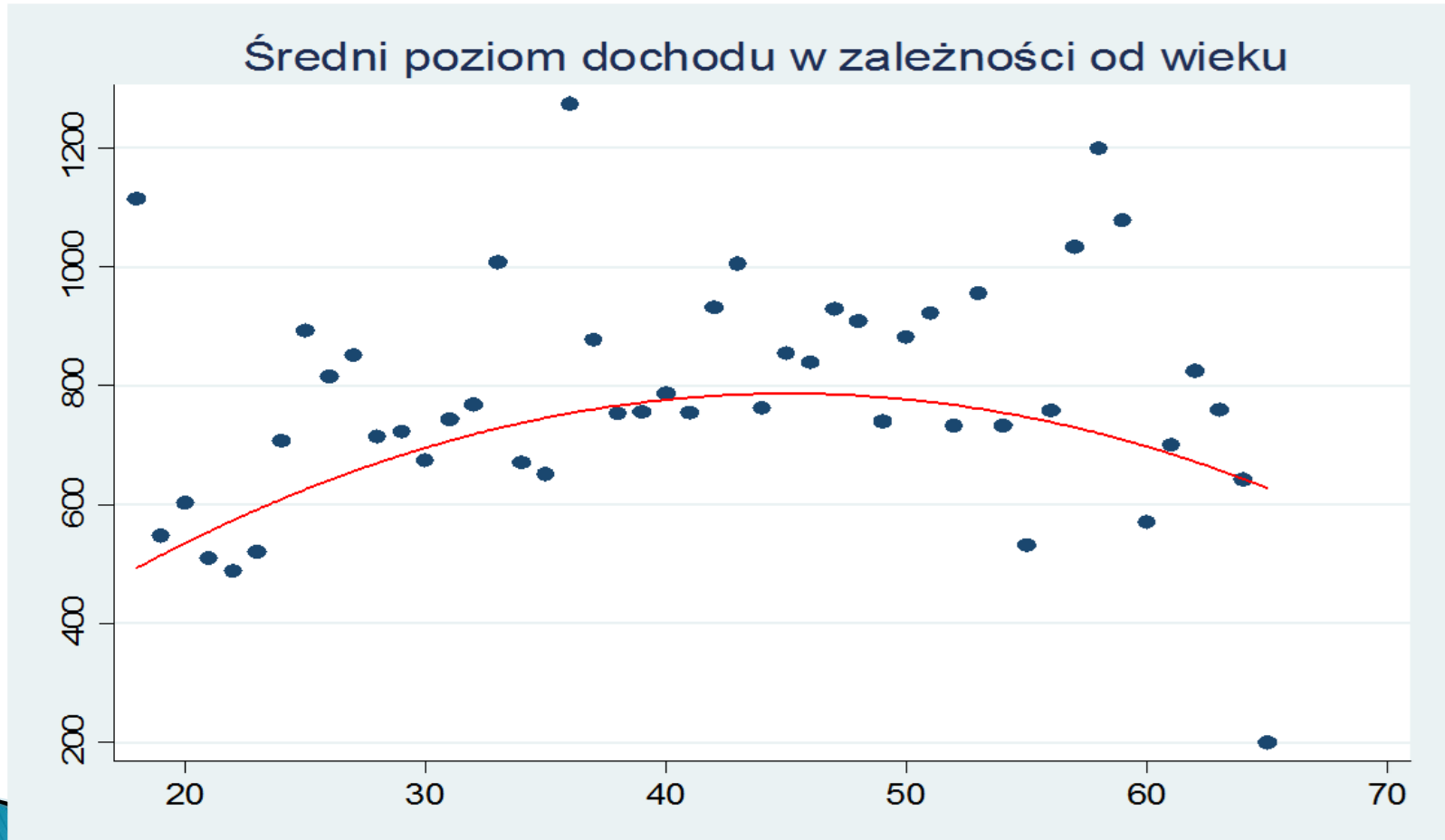
# INNE FORMY FUNKCYJNE MODELU ZE WZGLĘDU NA WIEK - WIELOMIAN STOPNIA II

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

Source	SS	df	MS			
Model	72048793.8	5	14409758.8	Number of obs = 1083		
Residual	675432341	1077	627142.378	F( 5, 1077) = 22.98		
Total	747481135	1082	690832.842	Prob > F = 0.0000		
				R-squared = 0.0964		
				Adj R-squared = 0.0922		
				Root MSE = 791.92		
dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
<b>wiek</b>	<b>36.06131</b>	<b>15.48328</b>	<b>2.33</b>	<b>0.020</b>	<b>5.680494</b>	<b>66.44212</b>
<b>wiek_2</b>	<b>-.3998842</b>	<b>.1973767</b>	<b>-2.03</b>	<b>0.043</b>	<b>-.7871707</b>	<b>-.0125977</b>
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

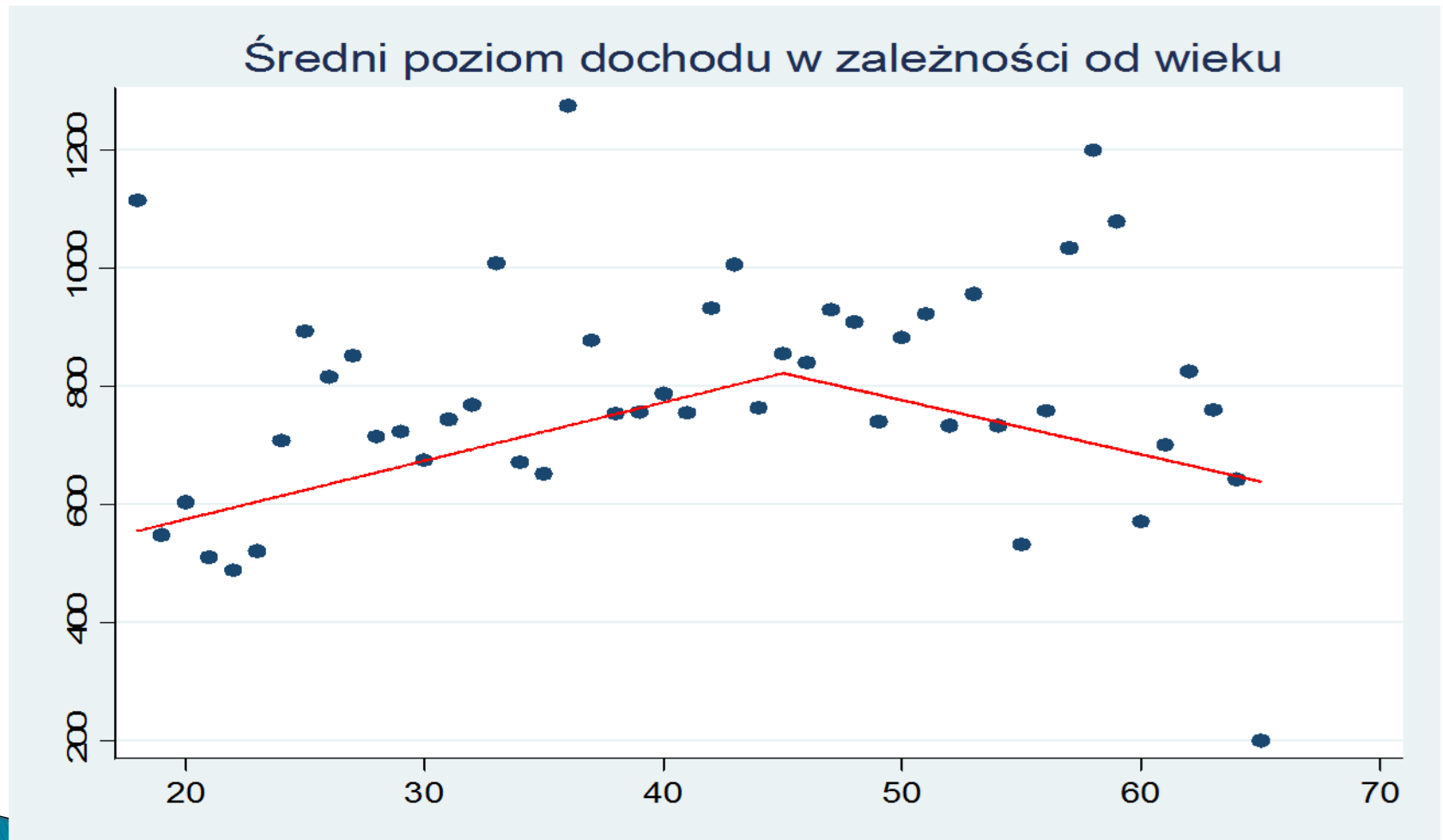
$$y = -26,64989 + 36,06131\text{wiek} - 0,3998842 \text{wiek}^2$$



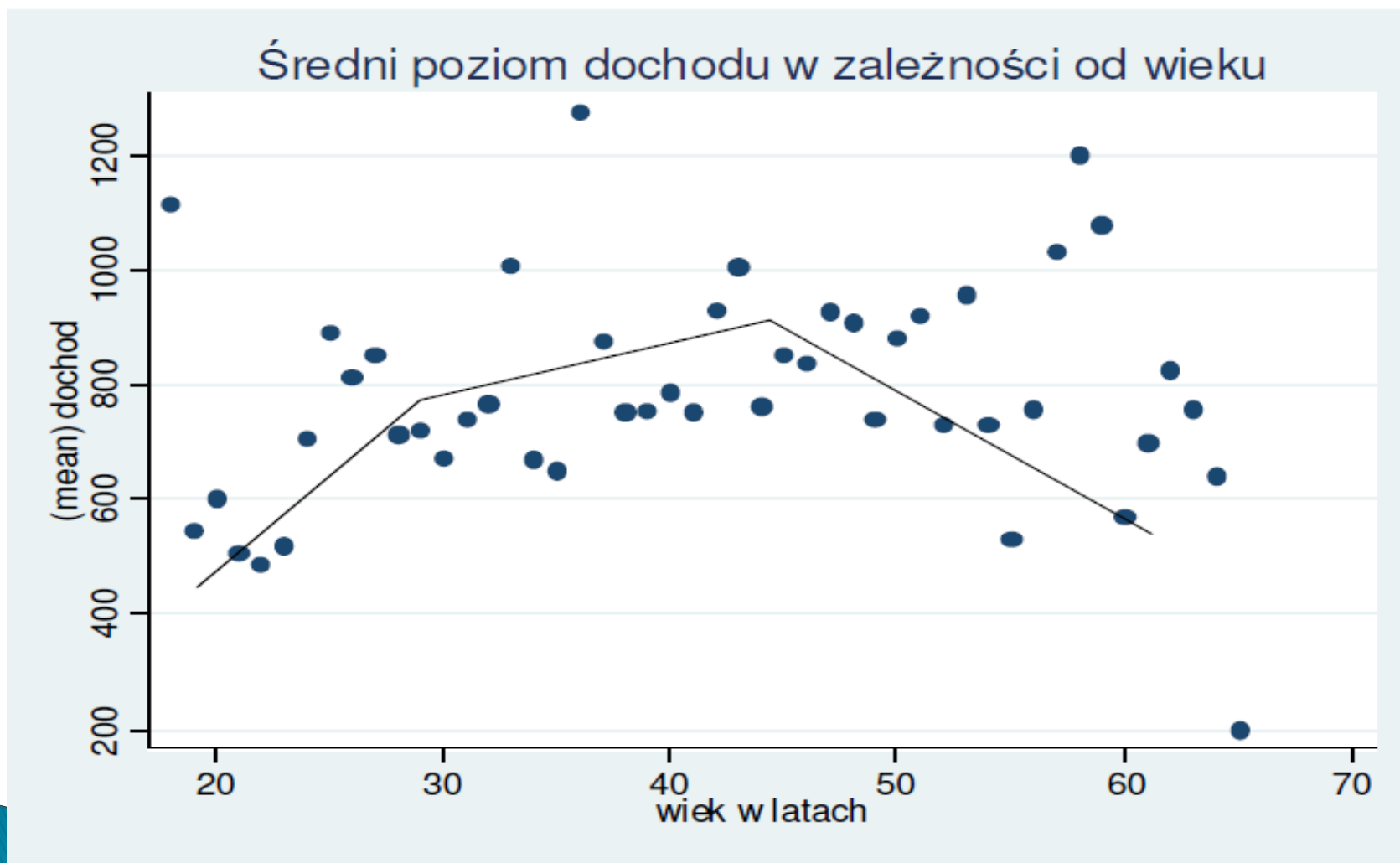
# UWAGA!

- ▶ Samodzielnie:
  2. Model krzywej łamanej
  3. Model schodkowy

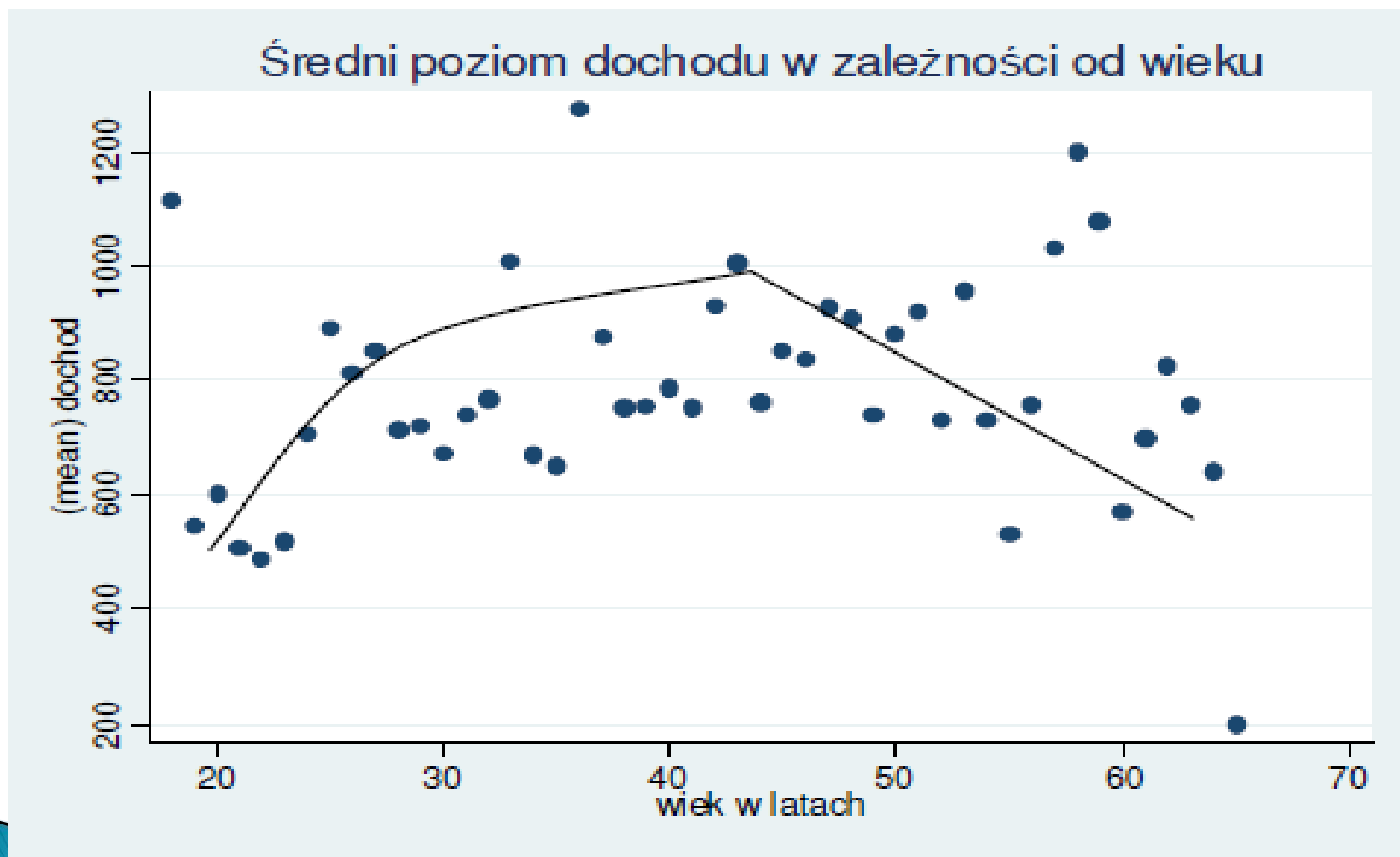
# Model krzywej łamanej



# Model krzywej łamanej



# Model krzywej łamanej





# Pytania teoretyczne

1. Dlaczego zmienna dyskretna rozkodowujemy na zmienne zerojedynkowe?
2. Dlaczego w modelu nie powinno się umieszczać stałej i wszystkich zmiennych zerojedynkowych, związanych z poziomami zmiennej dyskretnej?
3. Porównaj zastosowania znanych ci kontrastów ze standardowym sposobem rozkodowania zmiennej dyskretnej.
4. Wyjaśnić, co to znaczy, że między zmiennymi w modelu występują interakcje.
5. Opisać sposoby przybliżania zależności nieliniowej za pomocą modelu liniowego.

**Dziękuję za uwagę**