

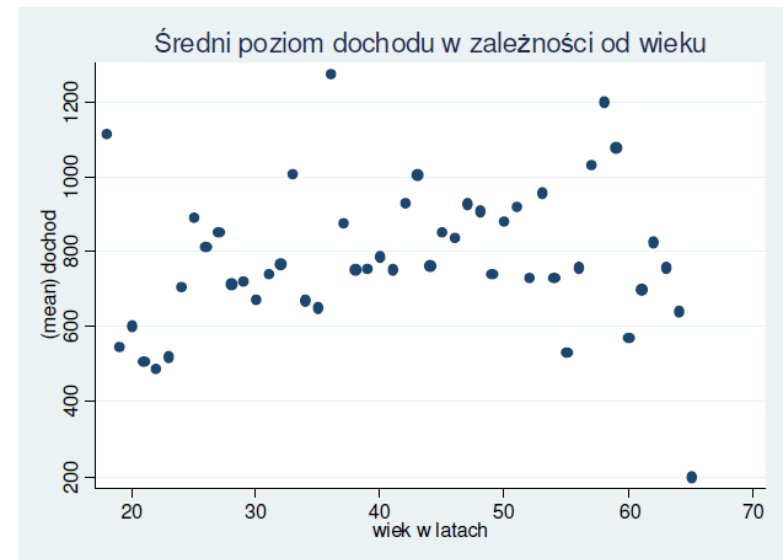
Wybór formy funkcyjnej (CZ. II)

Natalia Nehrebecka
Stanisław Cichocki

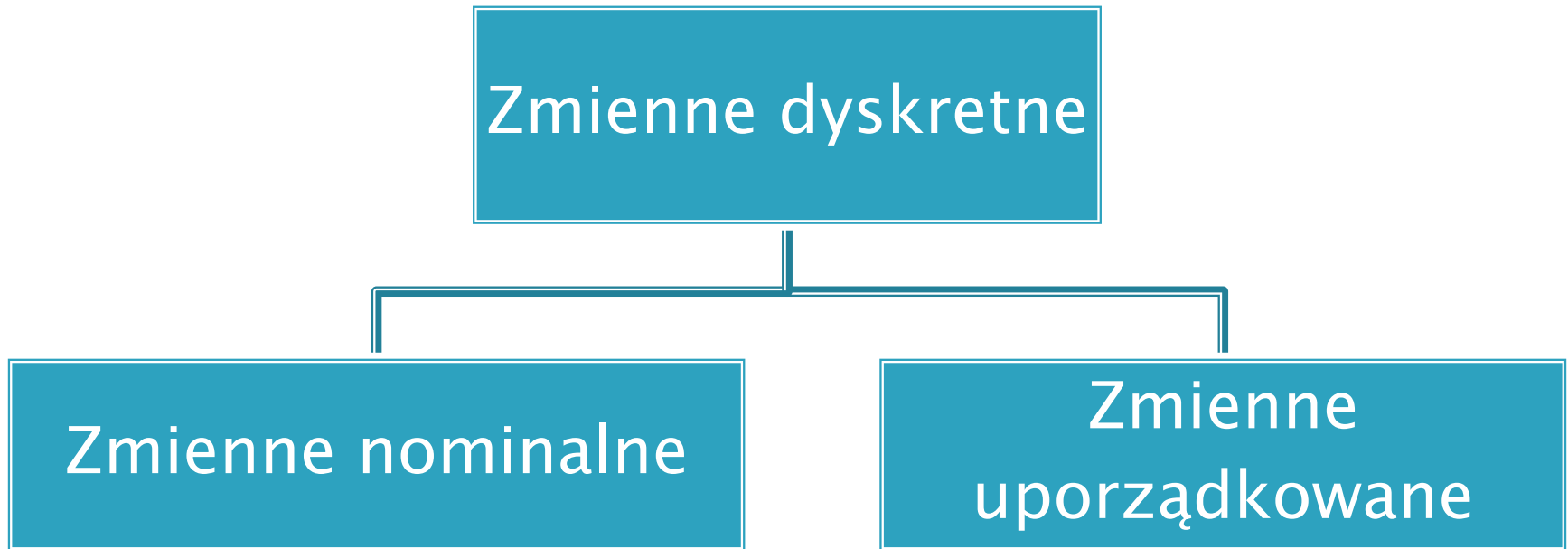
Wykład 7

Plan zajęć

1. Zmienne dyskretne
 - Zmienne zero-jedynkowe
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 $X^T X$ 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.
1 - wieś	323	29.82	29.82
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
_Imiasto_2	158.2807	74.50027	2.12	0.034	12.0986	304.4629
_Imiasto_3	265.9892	63.10008	4.22	0.000	142.1762	389.8022
_Imiasto_4	345.5604	73.00193	4.73	0.000	202.3182	488.8025
_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
_Imiasto_2	.2345958	.0508828	4.61	0.000	.1347553	.3344364
_Imiasto_3	.2870446	.0430966	6.66	0.000	.2024818	.3716073
_Imiasto_4	.4162679	.0498594	8.35	0.000	.3184353	.5141005
_cons	5.351579	.2074451	25.80	0.000	4.944537	5.758622

UWAGA!

- ▶ Praca na ćwiczeniach:
 - 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 iloczyny 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.wyksztalcenie
```

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
Total	747481135	1082	690832.842	Adj R-squared	= 0.1034
-----				Root MSE	= 787.01

dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
wiek	36.38318				
wiek_2	-.4049352				
_Iplec_1	-144.4044				
_Iwyksztal~2	274.2703				
_Iwyksztal~3	1040.998				
IpleXwyk~2	-143.4455				
IpleXwyk~3	-682.341				
_cons	-121.1625				

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

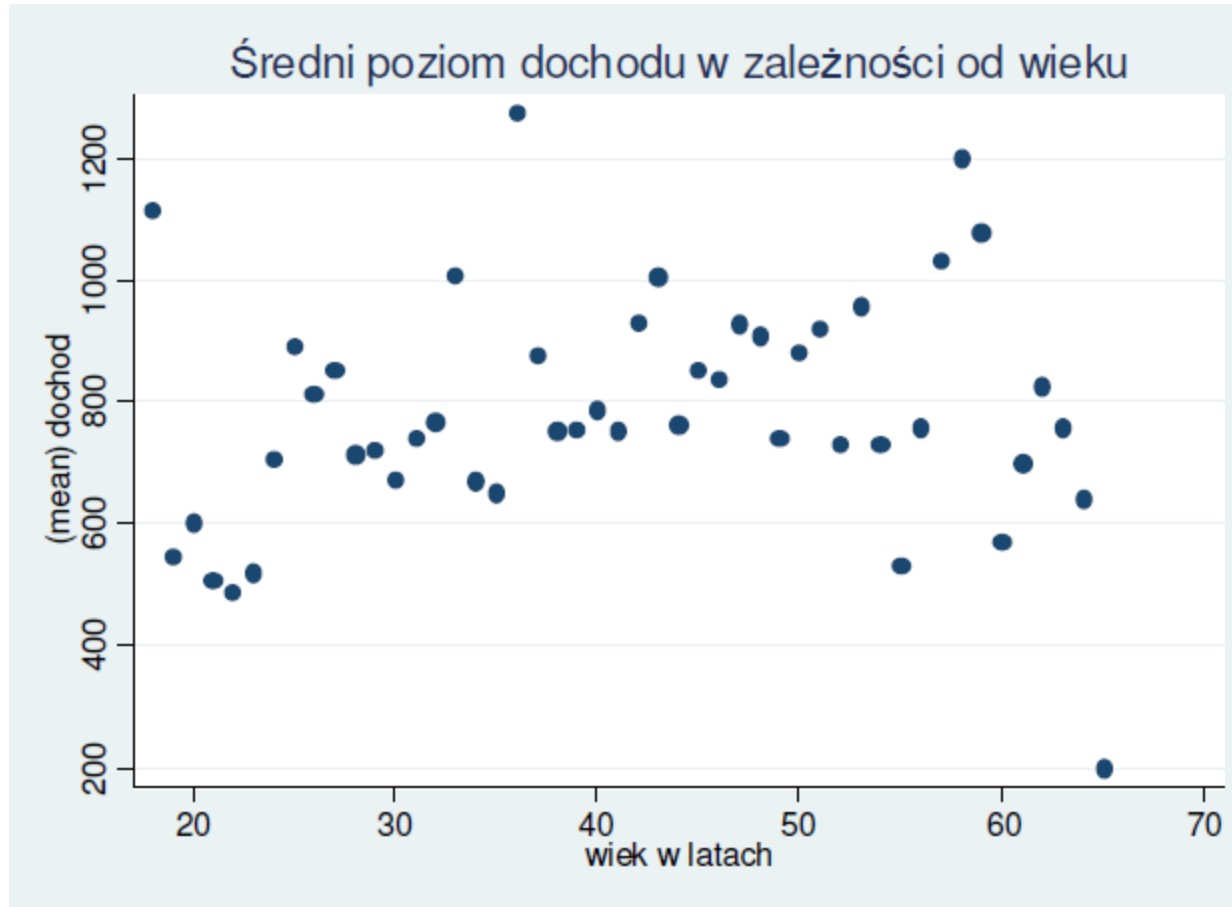
interakcje między zmienną miasto a wiekiem

```
xi: regress dochod i.miasto*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
Total	747481135	1082	690832.842	R-squared =	0.0285
				Adj R-squared =	0.0221
				Root MSE =	821.92

dochod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_Imiasto_2	28.34615				
_Imiasto_3	53.41383				
_Imiasto_4	135.6545				
wiek	-.4870689				
_ImiaXwiek_2	3.588019				
_ImiaXwiek_3	5.698882				
_ImiaXwiek_4	5.396286				
_cons	641.7219				

Zależność między y a x



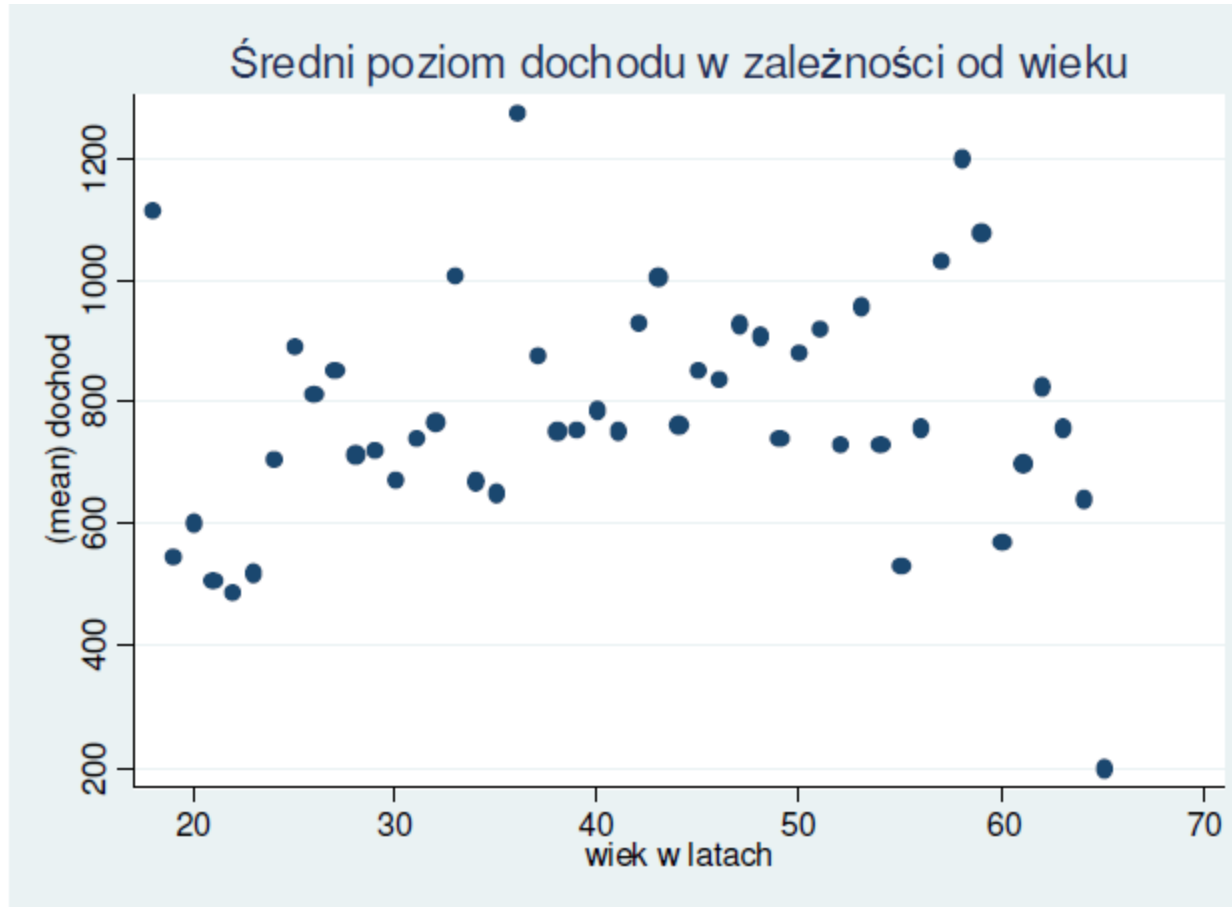
Modele wielomianowe

- ▶ 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

Zależność między y a x



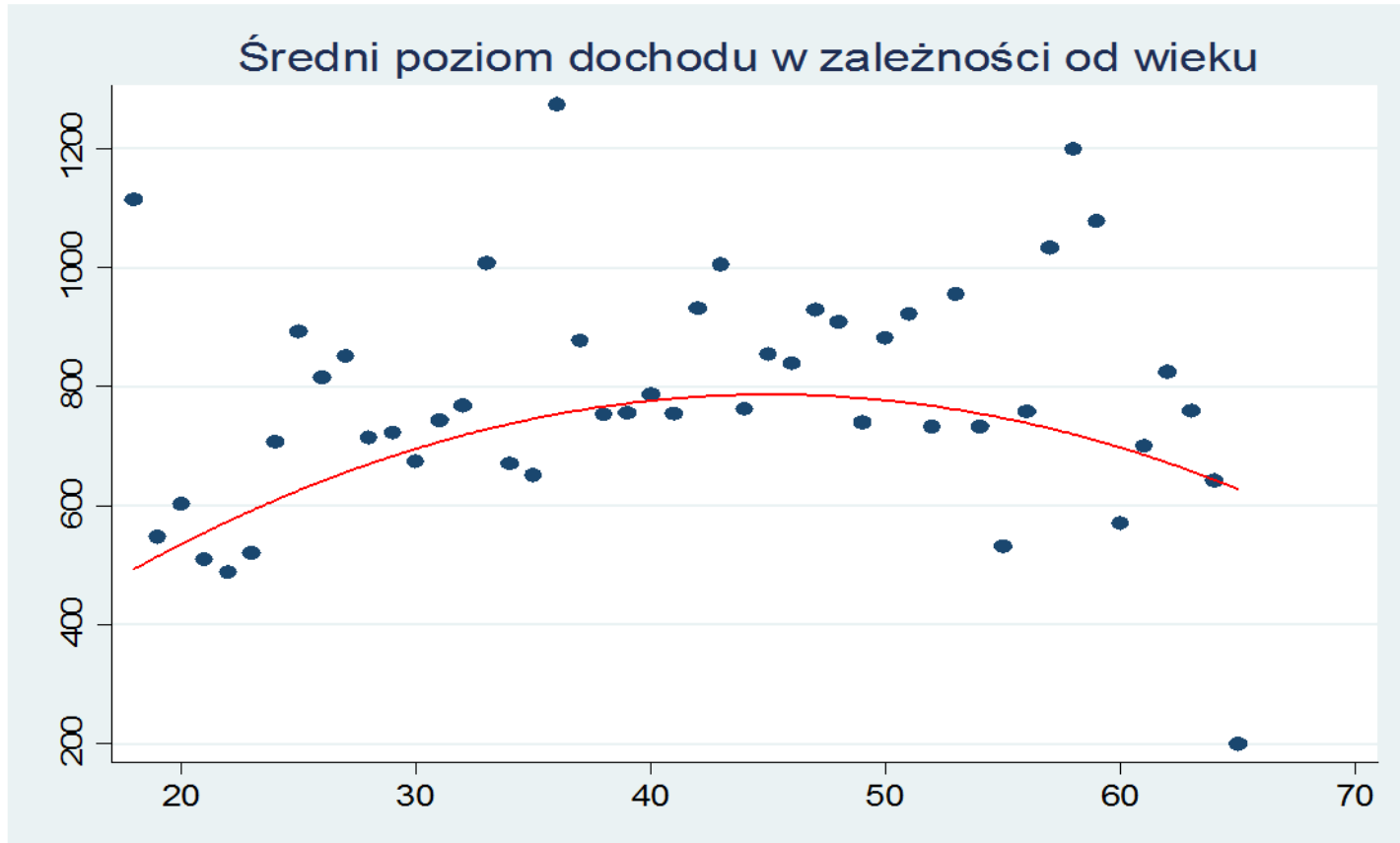
INNE FORMY FUNKCYJNE MODELU ZE WZGLĘDU NA WIEK - WIELOMIAN STOPNIA II

```
. 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	
				Prob > F =	0.0000	
				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
wysze	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

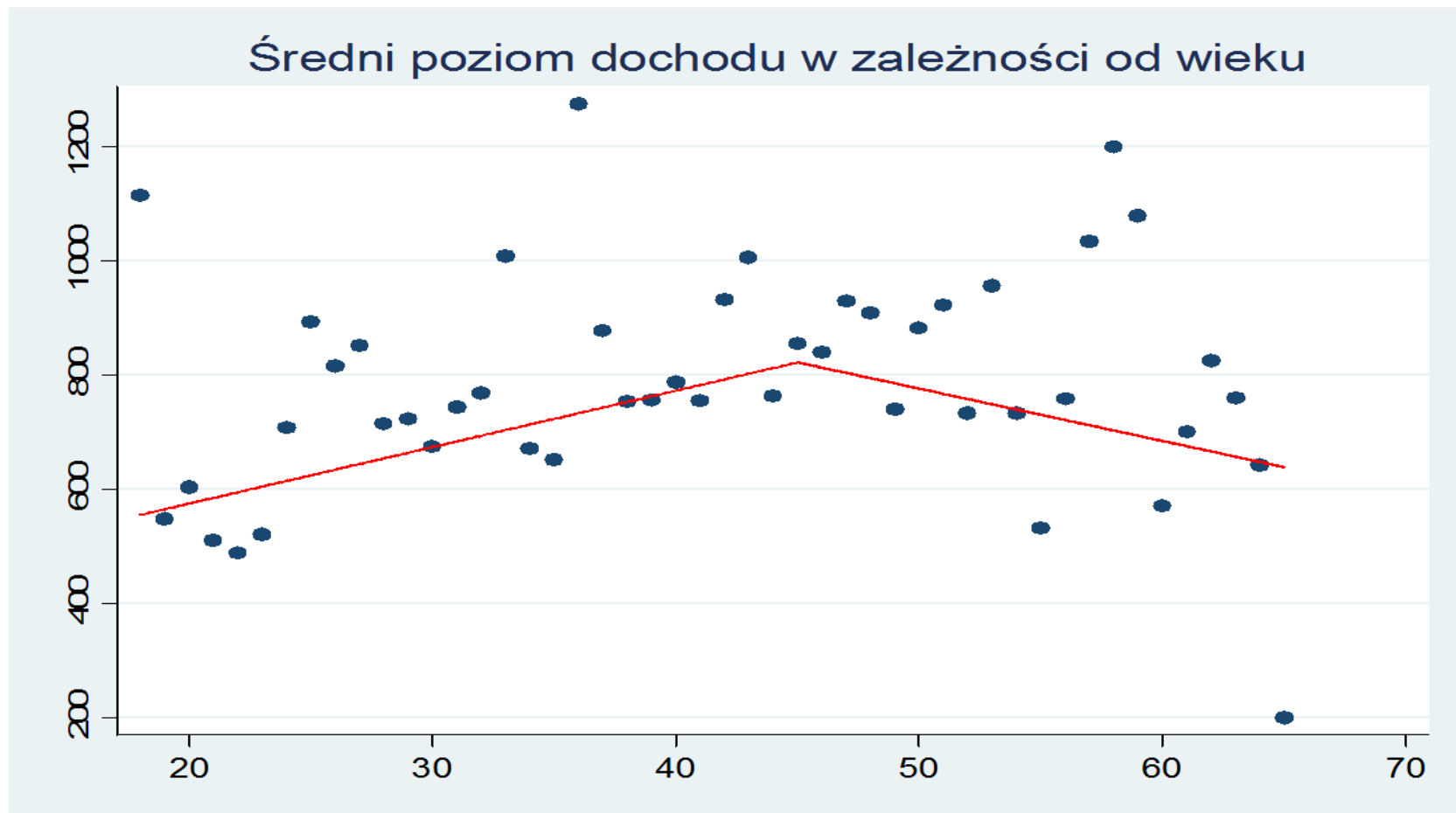
INNE FORMY FUNKCYJNE MODELU ZE WZGLĘDU NA WIEK - WIELOMIAN STOPNIA II



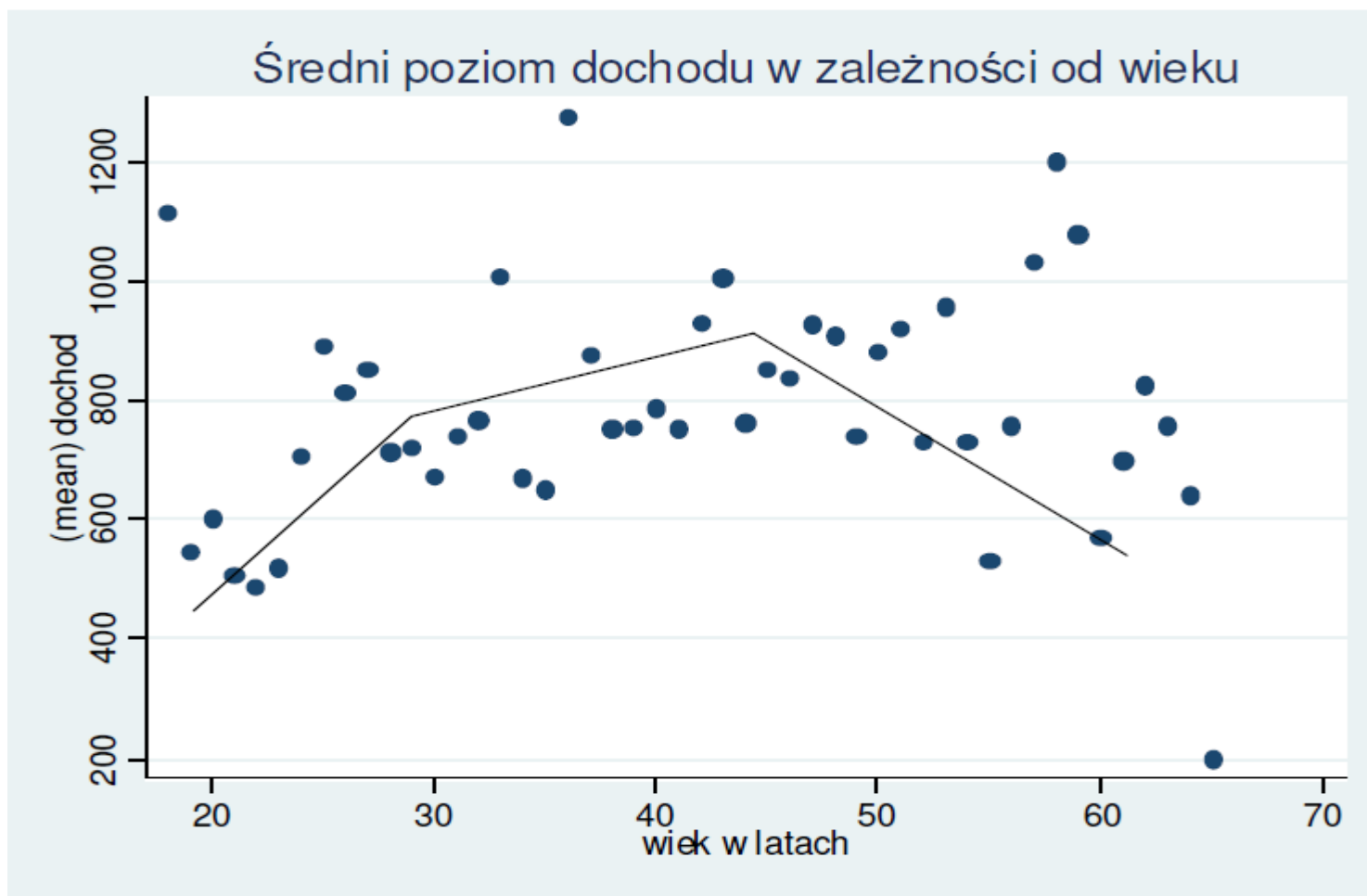
UWAGA!

- ▶ Praca na ćwiczeniach:
 - Model krzywej łamanej
 - Model schodkowy

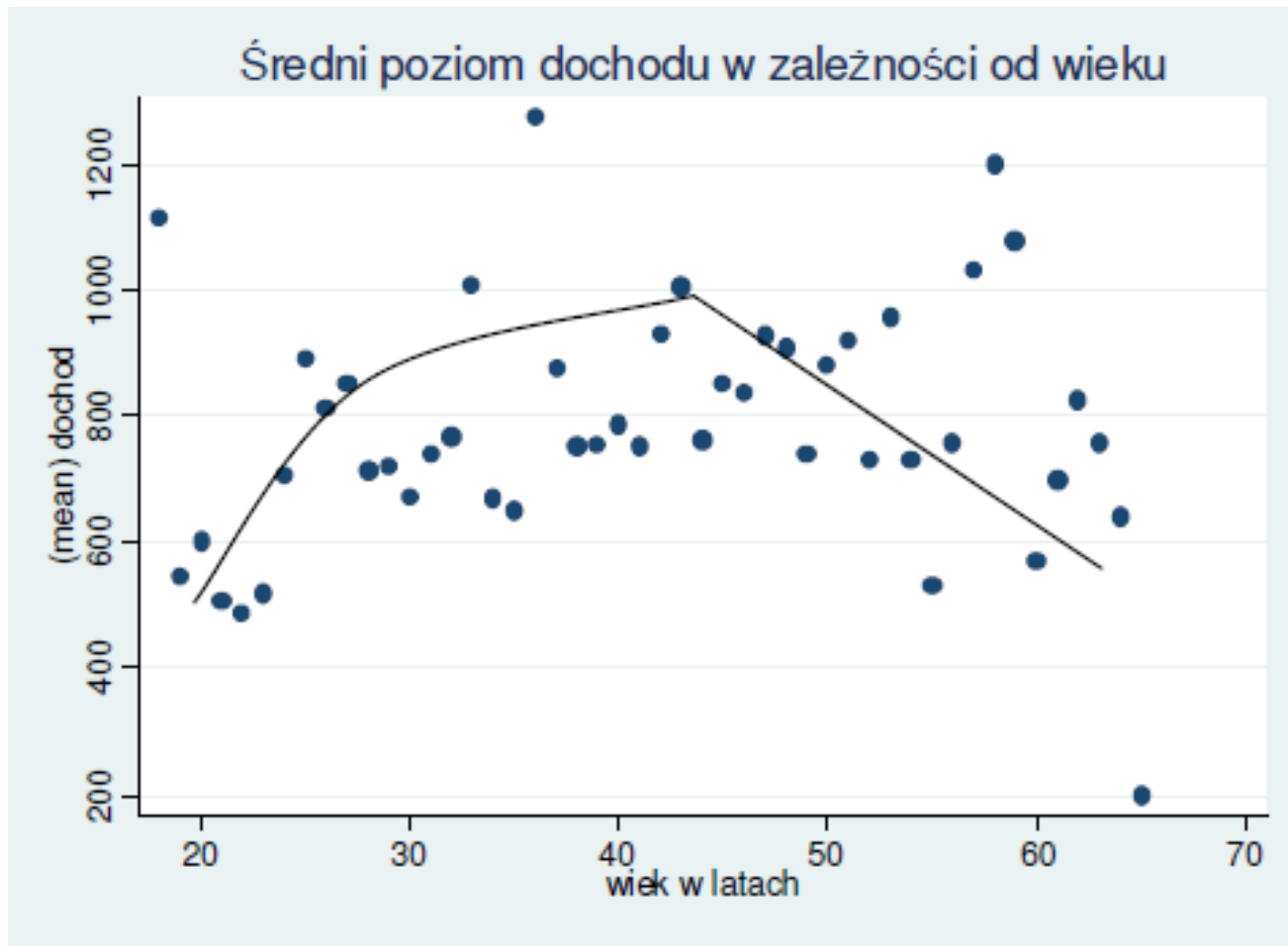
Model krzywej łamanej



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