

# Ekonometria

## Dekompozycja wariancji zmiennej zależnej, $R^2$ i jego własności – cz. I

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

**Wykład 4**

# Plan wykładu

- ▶ Dobroć dopasowania równania regresji. Współczynnik determinacji  $R^2$ 
  - Dekompozycja wariancji zmiennej zależnej
  - Współczynnik determinacji  $R^2$

# Plan wykładu

- ▶ Dobroć dopasowania równania regresji. Współczynnik determinacji  $R^2$ 
  - Dekompozycja wariancji zmiennej zależnej
  - Współczynnik determinacji  $R^2$

# Dobroć dopasowania równania regresji

## Współczynnik determinacji $R^2$

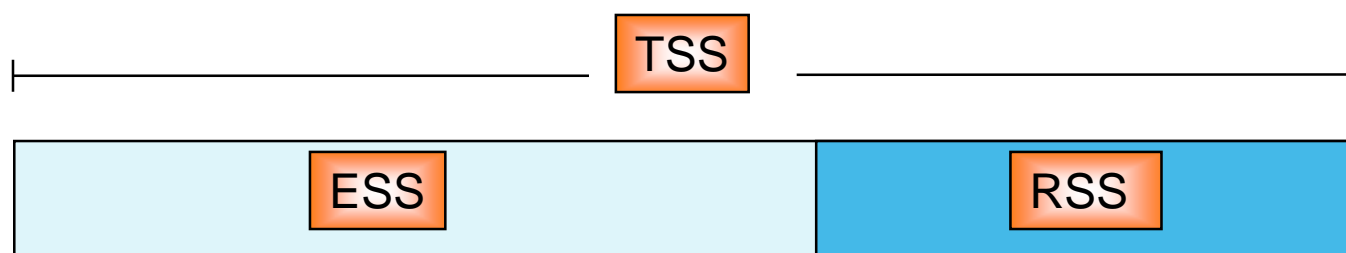
- ▶ Dobroć dopasowania równania regresji (do danych empirycznych) wyrażona jest przez tak zwany **współczynnik determinacji liniowej** oznaczany przez  $R^2$ .



- ▶ Współczynnik ten określa jaka część zmienności zmiennej objaśnianej  $y$  jest wyjaśniona łącznie przez zmienność wszystkich zmiennych objaśniających.
- ▶ Jedną z miar zmienności zmiennej jest **WARIANCJA**.

# Dekompozycja wariancji zmiennej zależnej

Wariancje zmiennej zależnej  $y$  można przedstawić jako dekompozycje (podział) na część wyjaśnioną przez model i na część niewyjaśnioną przez model.



Dekompozycja wariancji jest możliwa **JEDYNYE** dla  
**modelu ze stałą**

# Oznaczenia (1/3)

## ▶ Całkowita suma kwadratów:

Zmienność całkowitą zmiennej objaśnianej  $y$ , oznaczaną w literaturze angielskim skrótem TSS (***Total Sum of Squares***), mierzymy za pomocą sumy kwadratów odchyleń obserwacji zmiennej objaśnianej od średniej:

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2$$

# Oznaczenia (2/3)

## ▶ Wyjaśniona suma kwadratów:

**Jeśli model zawiera stałą**, to całkowitą sumę kwadratów możemy zdekomponować na dwa składniki, na wyjaśnioną (równaniem regresji) sumę kwadratów, oznaczaną przez ESS (***Explained Sum of Squares***)

$$ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

# Oznaczenia (3/3)

- ▶ **Resztowa suma kwadratów:**
- ▶ i resztową (*niewyjaśnioną*) sumę kwadratów, oznaczaną przez RSS (*Residual Sum of Squares*).

$$RSS = \sum_{i=1}^n e_i^2$$



# STATA

## Tablica analizy wariancji

Źródło zmienności	Suma kwadratów	Liczba stopni swobody	Średnia suma kwadratów
Source	SS	df	MS
Model (ESS)	34.9709245	3 (K-1)	11.6569748
Residual (RSS)	334.808247	1079 (N-K)	.310294946
Total (TSS)	369.779172	1082 (N-1)	41755242

ESS/df

RSS/df

# Plan wykładu

- ▶ Dobroć dopasowania równania regresji. Współczynnik determinacji  $R^2$ 
  - Dekompozycja wariancji zmiennej zależnej
  - Współczynnik determinacji  $R^2$

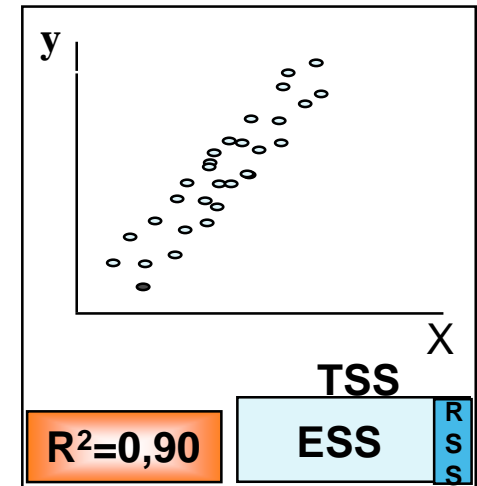
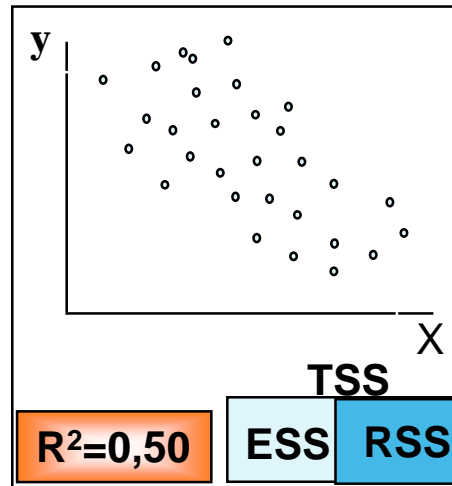
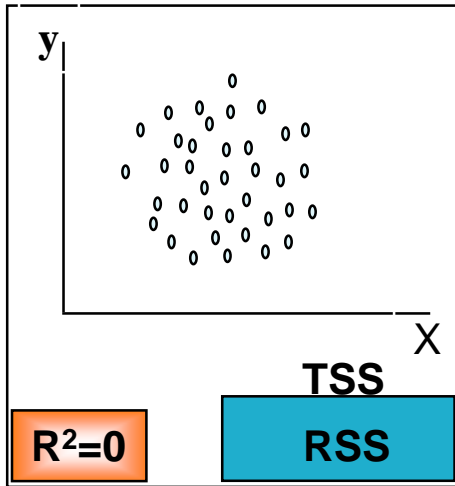
# Współczynnik determinacji $R^2$ (1/3)

$$R^2 = \frac{\text{wyjasniona suma kwadratów}}{\text{całkowita suma kwadratów}} = \frac{ESS}{TSS} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS}$$

Dla  
modelu ze stałą

$$0 \leq R^2 \leq 1$$

# Współczynnik determinacji $R^2$ (2/3)



# Współczynnik determinacji $R^2$ (3/3)

- ▶  $R^2$  przyjmuje wartości z przedziału domkniętego między 0 i 1.
- ▶ Jeśli  $R^2 = 1$ , to model regresji w 100% wyjaśnia zmienność  $y$ ,

$$y = \hat{y}, \quad e = 0 \quad \text{oraz} \quad RSS = 0$$

- ▶ a jeśli  $R^2 = 0$ , to model regresji w ogóle nie wyjaśnia zmienności  $y$ .

$$\hat{y} = \bar{y}, \quad ESS = 0$$

- ▶ Jeśli na przykład wynosi  $R^2 = 0,7$  to możemy powiedzieć, że
- ▶ 70% zmienności zmiennej objaśnianej  $y$  jest wyjaśnione przez łączną zmienność wszystkich zmiennych objaśniających, a 30% zmienności jest niewyjaśnione (jest zmiennością resztową).

# Współczynnik determinacji $R^2$

- ▶ Współczynnik determinacji  $R^2$  nazywany jest niekiedy **współczynnikiem korelacji wielorakiej**.
- ▶  $R^2$  jest równy kwadratowi współczynnika korelacji między  $y_i$  i  $\hat{y}_i$ .
- ▶ **W modelu ze stałą i jedną** zmienną współczynnik determinacji  $R^2$  jest równy kwadratowi współczynnika korelacji  $\hat{\rho}_{xy}$

# Pytania teoretyczne

1. Udowodnić, że w modelu ze stałą  $TSS = ESS + RSS$ .
2. Podać interpretacje  $R^2$ .

**Dziękuję za uwagę**