



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

# **Wprowadzenie do laboratorium 1**

**Estymacja jednorównaniowego modelu popytu  
na bilety lotnicze**



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## **Etapy budowy modelu ekonometrycznego**

- **Specyfikacja modelu**
- **Zebranie danych statystycznych**
- **Estymacja parametrów modelu**
- **Weryfikacja statystyczna modelu**
- **Praktyczne wykorzystanie modelu**



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## Specyfikacja modelu

- **Sformułowanie celu** i zakresu modelu oraz hipotez badawczych

**Cele:**

- **poznawcze**
  - **prognostyczne**
  - **normatywne**
- **Wybór** i zdefiniowanie **zmiennych** endogenicznych i egzogenicznych
  - **Wybór postaci analitycznej funkcji**



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## **Kolejne etapy budowy modelu**

- **Zebranie danych statystycznych**

**Struktura danych:**

- **dane przekrojowe**
- **szeregi czasowe**
- **dane panelowe**

- **Estymacja parametrów modelu**

**z wykorzystaniem oprogramowania GRET**



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## Weryfikacja modelu

- **testowanie istotności** wpływu poszczególnych **zmiennych niezależnych** na zmienną zależną ( **test t-Studenta** oraz **test F** )
- **ocena stopnia dopasowania modelu do danych empirycznych** (**błąd standardowy reszt  $S_e$** , **współczynnik zmienności resztowej  $V_e$** , **współczynnik determinacji  $R^2$** , **błędy standardowe parametrów**)
- **testowanie sferyczności / niesferyczności składnika losowego**:
  - **autokorelacji składnika losowego** (**test Durbina-Watsona**)
  - **heteroskedastyczności składnika losowego** (test White'a)
- **ocena liniowości postaci analitycznej modelu**



## Interpretacja parametrów

- w przypadku funkcji liniowej

Interpretuje się je jak **pochodne cząstkowe**:

Współczynnik  $\hat{a}_i$  oznacza **o ile średnio zmieni się zmienna objaśniana  $y$ , jeśli zmienna objaśniająca  $x_i$  wzrośnie *ceteris paribus* (przy niezmiennych pozostałych zmiennych objaśniających) o jednostkę.**

- w przypadku funkcji potęgowej

Interpretuje się je jak **współczynniki elastyczności**:

Współczynnik  $\hat{a}_i$  oznacza **o ile procent średnio zmieni się zmienna objaśniana  $y$ , jeśli zmienna objaśniająca  $x_i$  wzrośnie, *ceteris paribus*, o jeden procent.**



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## **Kierunki wykorzystania modelu**

- do celów **poznawczych**  
badanie zachowań podmiotów gospodarczych, analiza zależności ekonomicznych, badanie funkcjonowania systemów ekonomicznych, weryfikacja hipotez i teorii ekonomicznych
- do celów **prognostycznych**
- do celów **normatywnych**  
poszukiwanie efektywnych decyzji gospodarczych, analiza alternatywnych polityk ekonomicznych



## **Przesłanki uwzględnienia składnika losowego w modelu ekonometrycznym:**

- a. **niedeterministyczny charakter zjawisk społeczno-gospodarczych, konieczność uwzględnienia czynnika losowego**
- b. **błędy** wynikające z niedokładności **pomiaru** statystycznego, błędy obserwacji
- c. błędy wynikające z **nieuwzględnienia wśród zmiennych objaśniających niektórych czynników** mogących mieć wpływ na kształtowanie się zmiennej objaśnianej
- d. **błędy** wynikające z **przyjętej postaci analitycznej** (niedokładnie odzwierciedlającej rzeczywistą zależność funkcyjną)

Przesłanka (a) odzwierciedla immanentną, niezależną od badającego, własność zjawisk gospodarczych – niedeterministyczny, losowy charakter.

Przesłanki (b, c i d) odzwierciedlają błędy, które można ograniczyć w wyniku doskonalenia metod gromadzenia i analizy danych statystycznych oraz metod estymacji.



## **Założenia modelu KMNK**

1.  $y = Xa + \varepsilon$

(Każda obserwacja  $y_t$  jest liniową funkcją obserwacji  $x_{tk}$  oraz składnika losowego  $\varepsilon_t$ )

2.  $E \varepsilon = 0$

(Składnik losowy ma wartość oczekiwaną równą zeru.)

3.  $E \varepsilon \varepsilon' = \sigma^2 I$

(Założenie o sferyczności składnika losowego)

3a.  $E \varepsilon_t^2 = \sigma^2 I$

(Wariancja składnika losowego jest stała, tzn. występuje jednorodność wariancji składnika losowego)

3b.  $E \varepsilon_s \varepsilon_t = 0 \quad \forall s \neq t$

(Składnik losowy jest nieskorelowany, nie występuje autokorelacja składnika losowego)

4.  $X$  jest macierzą  $n \times (k+1)$  o elementach ustalonych w powtarzalnych próbach

5.  $r(x) = k + 1 \leq n$

Między zmiennymi objaśniającymi nie ma zależności liniowej.



Irena Woroniecka *EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## Klasyczna metoda najmniejszych kwadratów

**Z tw. Gaussa-Markowa:**

Przy powyższych założeniach klasyczna metoda najmniejszych kwadratów (**KMNK**) daje najlepsze (o najniższej wariancji) estymatory wśród liniowych i nieobciążonych.

**BLUE** – *Best Linear Unbiased Estimators* - najlepsze nieobciążone estymatory liniowe

$$\hat{a} = (X^T X)^{-1} X^T y$$



Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2

Wydział Informatycznych Technik Zarządzania

Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT

---

## Test Fishera-Snedecora

**Test F Fishera –Snedecora umożliwia całościową ocenę przydatności modelu**

Hipoteza  $H_0 : a_1 , a_2 , \dots , a_i = 0$  (wszystkie parametry przy zmiennych objaśniających są równe zero)

wobec hipotezy  $H_1$  , że przynajmniej jeden parametr jest różny od zera

Wartość statystyki F obliczona dla modelu: 
$$F = \frac{\frac{R^2}{k}}{\frac{1 - R^2}{n - k - 1}}$$

ma rozkład F o poziomie istotności  $\alpha$  oraz  $s_1 = k$  ,  $s_2 = n - k - 1$

$$P \{ F < F_{kr} (\alpha , s_1 , s_2) \} = 1 - \alpha$$

$$P \{ F \geq F_{kr} (\alpha , s_1 , s_2) \} = \alpha$$



## Test Fishera-Snedecora - cd

$F < F_{kr}(\alpha, s_1, s_2)$       ►      taki wynik testu wskazuje, że brak podstaw do odrzucenia hipotezy  $H_0$ , praktycznie oznacza to, że wszystkie współczynniki stojące przy zmiennych objaśniających są nieistotnie różne od zera, a więc wszystkie zmienne objaśniające mają nieistotny statystycznie wpływ na zmienną  $y$  (podsumowując – **wszystkie zmienne  $x_i$  są nieistotne**, żadna z nich nie ma istotnego wpływu na zmienną objaśnianą  $y$ , **model jest nieprzydatny** z tego punktu widzenia).

$F \geq F_{kr}(\alpha, s_1, s_2)$       ►      taki wynik testu wskazuje, że istnieją podstawy do odrzucenia hipotezy  $H_0$ , tym samym należy przyjąć hipotezę  $H_1$ . Oznacza to, że przynajmniej jeden współczynnik  $a_i$  jest istotnie różny od zera, a tym samym **przynajmniej jedna zmienna objaśniająca ma istotny statystycznie wpływ na zmienną  $y$**  (podsumowując, test oparty na statystyce  $F$  daje pozytywną, z punktu widzenia jakości dopasowania modelu, odpowiedź – oszacowany model zawiera istotne zmienne objaśniające).

gdzie  $F_{kr}(\alpha, s_1, s_2)$  – wartość krytyczna statystyki  $F$  o poziomie istotności  $\alpha$  oraz

$$s_1 = k, \quad s_2 = n - k - 1$$



## Test t-Studenta

**Test t-Studenta: umożliwia wyselekcjonować i odrzucić nieistotne zmienne objaśniające**

$$H_0 : a_i = 0 \quad \text{wobec} \quad H_1 : a_i \neq 0$$

zmienna losowa  $t_i = \frac{\hat{a}_i}{S_{ai}}$  ma rozkład t-Studenta o poziomie istotności  $\alpha$  oraz

liczbie stopni swobody  $r$  (jest to obliczona wartość statystyki t-Studenta dla danej zmiennej objaśniającej  $x_i$  )

Liczba stopni swobody :  $r = t - (k + 1)$  dla modelu z wyrazem wolnym lub  
 $r = t - k$  bez wyrazu wolnego

$$P \{ |t_i| \leq t_{kr}(\alpha, r) \} = 1 - \alpha$$

$$P \{ |t_i| > t_{kr}(\alpha, r) \} = \alpha$$



## Test t-Studenta – cd.

$|t_i| \leq t_{kr}(\alpha, r)$  ► taki wynik testu wskazuje, że brak podstaw do odrzucenia hipotezy  $H_0$ , praktycznie oznacza to, że współczynnik  $a_i$  jest nieistotnie różny od zera, a zmienna  $x_i$  ma nieistotny statystycznie wpływ na zmienną  $y$  (krótko - **zmienna  $x_i$  jest nieistotna**).

$|t_i| > t_{kr}(\alpha, r)$  ► taki wynik testu wskazuje, że istnieją podstawy do odrzucenia hipotezy  $H_0$ , tym samym należy przyjąć hipotezę  $H_1$ . Oznacza to, że współczynnik  $a_i$  jest istotnie różny od zera, a zmienna  $x_i$  ma istotny statystycznie wpływ na zmienną  $y$  (krótko - **zmienna  $x_i$  jest istotna**).

gdzie  $t_{kr}(\alpha, r)$  – **wartość krytyczna** statystyki t-Studenta o **poziomie istotności  $\alpha$**  oraz **liczbie stopni swobody  $r$**

Wykorzystując ten test należy zastosować sekwencyjną metodę odrzucania nieistotnych zmiennych objaśniających – zaczynając od zmiennych najmniej istotnych (o najniższej, co do modułu, wartości statystyki t-Studenta).



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## Test t-Studenta w GRETL

Dla każdego parametru podawane są:

- wartość **statystyki t-Studenta**
- **p-value** - empiryczny poziom istotności (dwustronne prawdopodobieństwo związane z rozkładem t-Studenta)
- symboliczne oznaczenie stopnia istotności (**gwiazdki**)

**Uwaga:**

Liczba gwiazdek charakteryzuje istotność zmiennych:

- \*\*\* - zmienna istotna statystycznie przy poziomie istotności 0,01;
- \*\* - zmienna istotna przy poziomie istotności 0,05;
- \* - zmienna istotna przy poziomie istotności 0,1.



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## Test F a test t-Studenta

### Uwaga:

**Test F nie rozstrzyga czy wszystkie zmienne objaśniające są istotne, odpowiedź na takie pytanie daje test oparty na statystyce t-Studenta.**



Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2

Wydział Informatycznych Technik Zarządzania

Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT

---

## Badanie założeń dotyczących składnika losowego

**Założenie o sferyczności składnika losowego:**

$E \varepsilon \varepsilon' = \sigma^2 I$  oznacza, że: macierz kowariancji jest macierzą diagonalną z jednakowymi wartościami na przekątnej równymi  $\sigma^2$  i zerami poza diagonalną ( $\xi$ - wektor)

można rozbić na dwa założenia:

$E \varepsilon_t^2 = \sigma^2 I$  występuje jednorodność wariancji składnika losowego

Niespełnienie tego założenia oznacza, że występuje heteroskedastyczność składnika losowego.

$E \varepsilon_s \varepsilon_t = 0 \quad \forall s \neq t$  składnik losowy jest niezależny

Niespełnienie tego założenia oznacza, że występuje autokorelacja składnika losowego.



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## Test Durbina-Watsona

**Test Durbina-Watsona weryfikuje brak / występowanie autokorelacji pierwszego rzędu**

Hipoteza  $H_0 : \rho_1 = 0$  oznacza brak autokorelacji pierwszego rzędu, wobec hipotezy  $H_1 : \rho_1 > 0$ .

gdzie  $\rho_1$  - współczynnik autokorelacji pierwszego rzędu

Statystyka Durbina-Watsona  $d$  :

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=2}^n e_t^2}$$

przybliżenie:  $d \approx 2 - 2\rho_1 = 2(1 - \rho_1)$



## Test Durbina-Watsona - cd

$d \leq d_L(\alpha, n, k)$       ► taki wynik testu wskazuje, że są podstawy do odrzucenia hipotezy  $H_0$  (wniosek: **występuje autokorelacja wariancji składnika losowego**)

$d_L(\alpha, n, k) < d < d_U(\alpha, n, k)$       ► taki wynik testu **nie rozstrzyga kwestii autokorelacji składnika losowego**

$d \geq d_U(\alpha, n, k)$       ► taki wynik testu wskazuje, że brak podstaw do odrzucenia hipotezy  $H_0$  (wniosek: **nie występuje autokorelacja wariancji składnika losowego**)

$d_L(\alpha, n, k)$  - dolna wartość krytyczna statystyki Durbina-Watsona dla poziomu istotności  $\alpha$ , liczebności próby  $n$  i liczby zmiennych objaśniających  $k$

$d_U(\alpha, n, k)$  - górna wartość krytyczna statystyki Durbina-Watsona dla poziomu istotności  $\alpha$ , liczebności próby  $n$  i liczby zmiennych objaśniających  $k$



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## **Test Durbina-Watsona dla ujemnej korelacji**

**Test dla ujemnej korelacji:**

**Hipoteza  $H_0 : \rho_1 = 0$  oznacza brak autokorelacji pierwszego rzędu,  
wobec hipotezy  $H_1 : \rho_1 < 0$  .**

**wykorzystuje statystykę  $d' = 4 - d$  .**

**Dla ujemnej korelacji statystyka  $d$  przyjmuje wartości z przedziału  
(2,4). Wtedy należy dokonać przekształcenia:**

$$**d' = 4 - d .**$$



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## Ocena dopasowania modelu do danych empirycznych

- **błąd standardowy reszt  $S_e$**   
**oraz współczynnik zmienności resztowej  $V_e$**
- **współczynnik determinacji  $R^2$  : nieskorygowany i skorygowany**  
**oraz współczynnik zbieżności  $\phi^2$**
- **błędy standardowe parametrów  $S_{ai}$**



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## **Odchylenie standardowe składnika losowego**

**Błąd standardowy reszt:**

$$S_e = \sqrt{\frac{\sum e^2}{n - (k + 1)}}$$

**Współczynnik zmienności resztowej:**

$$V_e = \frac{S_e}{\bar{y}}$$



## Współczynnik $R^2$

Współczynnik determinacji  $R$  kwadrat (*Unadjusted R-squared*)

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y - \bar{y})^2} = 1 - \frac{\sum e^2}{\sum (y - \bar{y})^2}$$

gdzie:

SST (*total sum of squares*) – całkowita (ogólna) wariancja zmiennej objaśnianej  $y$ , suma kwadratów odchyleń wartości empirycznych od średniej (zmiennność całkowita)

SSR (*regression sum of squares*) – objaśniona wariancja zmiennej objaśnianej  $y$ , suma kwadratów odchyleń wartości teoretycznych od średniej (zmiennność objaśniona)

SSE (*error sum of squares*) – nieobjaśniona wariancja zmiennej objaśnianej  $y$ , suma kwadratów reszt, czyli suma kwadratów odchyleń wartości teoretycznych od empirycznych (zmiennność nieobjaśniona)



Irena Woroniecka *EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## Współczynnik zbieżności

**Współczynnik zbieżności**  $\phi^2 : \phi^2 = 1 - R^2$        $\phi^2 = \frac{\sum e^2}{\sum (y - \bar{y})^2}$

**Skorygowany współczynnik determinacji (*Adjusted R-squared*)**  
umożliwia porównywalność różnych modeli ekonometrycznych

$$\tilde{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$$



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## **Jak zapisujemy ostateczny wynik estymacji**

$$\hat{y} = 54,353 - 2,871 x_1 + 1,784 x_2 + 0,873 x_3$$

**(29,410) (0,446) (0,539) (0,310)**



*Irena Woroniecka EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

# Prognoza

## na podstawie jednorównaniowego modelu ekonometrycznego

Jeżeli dla klasycznego modelu regresji liniowej o postaci:

$$y_i = a_0 + a_1 x_{i1} + \dots + a_k x_{ik} + \varepsilon_i, \quad i = 1, \dots, n$$

spełnione są wszystkie założenia schematu Gaussa-Markowa, wtedy MNK-estymator jest BLUE, a prognoza na okres  $n+s$  wynosi:

$$\hat{y}_{n+s} = a_0 + a_1 x_{n+s,1} + \dots + a_k x_{n+s,k}, \quad s = 1, \dots, T$$



Irena Woroniecka *EKONOMIA MENEDŻERSKA - dodatek do W2*

*Wydział Informatycznych Technik Zarządzania*

*Wyższa Szkoła Informatyki Stosowanej i Zarządzania WIT*

---

## Błędy prognozy

**Błąd prognozy *ex ante* w okresie  $n+s$  -  $V_{n+s}$  :**

$$V_{n+s}^2 = S_e^2 [ 1 + x_{n+s} (X^T X)^{-1} x_{n+s}^T ]$$

$$V_{n+s} = \sqrt{V_{n+s}^2}$$

**Błąd prognozy *ex post* w okresie  $n+s$  -  $\delta_{n+s}$  :**

$$\delta_{n+s} = y_{n+s} - \hat{y}_{n+s}$$

**Względne błędy prognozy :**

$$V_{n+s}^* = \frac{V_{n+s}}{\hat{y}_{n+s}} 100 \%$$

$$\delta_{n+s}^* = \frac{\delta_{n+s}}{\hat{y}_{n+s}} 100 \%$$



## Źródła błędów prognoz

- a. **błędy estymacji** (wartości estymatorów różnią się od rzeczywistych wartości parametrów)
- b. **błędy struktury stochastycznej** (jeśli założenia dotyczące składnika losowego nie są spełnione, estymatory tracą pożądane własności)
- c. **błędy losowe** (wynikające z niedeterministycznego charakteru zjawisk społeczno-gospodarczych)
- d. **błędy pomiaru** wynikające z niedokładności pomiaru statystycznego
- e. **błędy specyfikacji:**
  - błędy wynikające z **nieuwzględnienia** wśród zmiennych objaśniających **niektórych czynników** mogących mieć wpływ na kształtowanie się zmiennej objaśnianej
  - błędy wynikające z **przyjętej postaci analitycznej**, niedokładnie odzwierciedlającej rzeczywistą zależność funkcyjną
- f. **błędy warunków endogenicznych** (zmienia się siła oddziaływania między zmiennymi)
- g. **błędy warunków egzogenicznych** (błędnie przyjęte wartości zmiennych egzogenicznych w okresie prognozy)