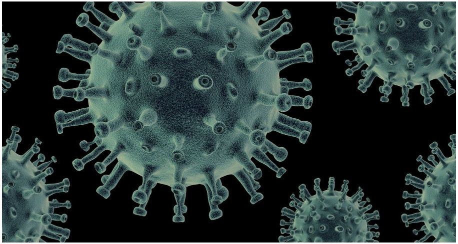
Modele ekonometryczne, zmienne objaśniające i objaśniane

8 marca 2021



Niech $(\Omega, \mathcal{F}, P)$ będzie przestrzenią probabilistyczną, gdzie

- Ω zbiór **zdarzeń elementarnych**
- $\mathcal{F}$ zbiór **zdarzeń**, gdzie typowy element oznacza (mierzalny) zbiór zdarzeń elementarnych;
- P jest **funkcją prawdopodobieństwa**, która każdemu zdarzeniu wyznacza **prawdopodobieństwo**, które jest liczbą między 0, a 1. Ponadto, $P(\emptyset) = 0$, $P(\Omega) = 1$, oraz

$$P(A_1 \cup A_2) = P(A_1) + P(A_2), \quad \text{dla wszystkich } A_1, A_2 \in \mathcal{F},$$

o ile $A_1 \cap A_2 = \emptyset$.

Przykład 1

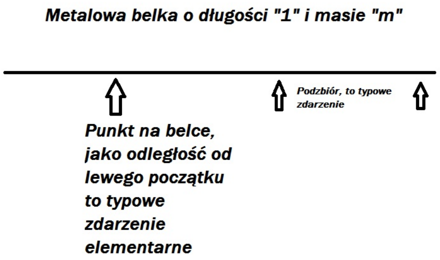
Rozważmy rzut tradycyjną kostką do gry:

$$
\begin{aligned}
\Omega &= \{1,2,3,4,5,6\} \\
\mathcal{F} &= \textit{rodzina wszystkich podzbiorów } \Omega \\
P(A) &= \frac{\#A}{\#\Omega}
\end{aligned}
$$

Przykładowo,

- *prawdopodobieństwo, że kostka pokaże liczbę nieparzystą* $P(\{1,3,5\}) = \frac{3}{6} = \frac{1}{2}$,
- *z kolei prawdopodobieństwo, że kostka pokaże liczbę podzielną przez* 3 *wynosi* $P(\{3,6\}) = \frac{2}{6} = \frac{1}{3}$.

Masa również ma własności prawdopodobieństwa.



Przykład 2

Własności prawdopodobieństwa ma również rozkład masy metalowego pręta o długości 1 *i masie* $m > 0$. *Zdarzeniem elementarnym jest punkt na belce jako odległość od lewego końca, ma on masę* 0, *a zdarzenie to dowolny wycinek pręta, lub zestaw wycinków. Przestrzeń zdarzeń to* $\Omega = [0,1]$. *Masa pręta wyraża się wzorem*

$$P(A) = \frac{1}{m}\int_A \rho(x)dx$$

gdzie ρ *jest funkcją gęstości masy, tzn.* $\rho(x) \geq 0$ *dla* $x \in [0,1]$ *i* $\int_0^1 \rho(x)dx = m$. *Wtedy* $\rho(x)/m$ *ma własności gęstości prawdopodobieństwa.*

Definicja 1

Zmienna losowa *to borelowska funkcja, która każdemu zdarzeniu wyznacza liczbę rzeczyswistą. Matematycznie,* X *jest zmienną losową jeśli*

- $X : \Omega \mapsto \mathbb{R}$ - X *przekształca przestrzeń zdarzeń w zbiór liczb rzeczywistych,*
- *dla dowolnego zbioru borelowskiego* $A \subset \mathbb{R}$, $X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$ - *przeciwobraz zbioru borelowskiego jest zdarzeniem.*

Definicja 2

Dla każdej zmiennej losowej X, odwzorowanie

$$A - \text{ Borel } \mapsto P(X^{-1}(A))$$

nazywamy ***rozkładem prawdopodobieństwa zmiennej losowej X***. *W skrócie piszemy $P(X \in A)$*

Zmienne losowe i ich parametry

W praktyce spotykamy 2 typy rozkładów prawdopodobieństwa:

- **dyskretny:** przestrzeń zdarzeń Ω jest skończona lub przeliczalna (tzn. $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ lub $\Omega = \{\omega_1, \omega_2, \ldots, \}$), a rozkład jest wyrażony wzorem

$$P(X \in A) = \sum_{a \in A} P(X = a).$$

- **ciągły:** przestrzeń zdarzeń to prosta rzeczywista $\mathbb{R}$, a rozkład wyraża się wzorem:

$$P(X \in A) = \int_A f(x)dx$$

gdzie f nazywamy **gęstością** prawdopodobieństwa; każda gęstość spełnia warunki:

1. $f(x) \geq 0$ dla $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x) = 1.$

Uwaga 1

Zmienna losowa wyznacza nową przestrzeń probabilistyczną na prostej $\mathbb{R}$*:*

- $X(\omega)$ *zdarzenie elementarne;*
- $\{X \in A\}$ *możliwe zdarzenie;*
- $P(X \in A)$ *funkcja prawdopodobieństwa (na* $\mathbb{R}$*, wcześniej było na* Ω*).*

Definicja 3

Wartością oczekiwaną *zmiennej losowej nazywamy wielkość*

$$EX = \int_{\Omega} X(\omega) P(d\omega).$$

Definicja 4

Wariancją X *zmiennej losowej nazywamy wielkość*

$$Var(X) = E(X - E(X))^2 = EX^2 - (EX)^2.$$

Liczbę $\sigma_X := \sqrt{Var(X)}$ *nazywamy* ***odchyleniem standardowym****.*

W praktyce:

- Jeśli X ma rozkład dyskretny taki, że $p_i = P(X = \omega_i)$, $i = 1, 2, \ldots, n$ $(i = 1, 2, \ldots)$

$$EX = \sum_i X(\omega_i) p_i$$

- Jeśli X ma rozkład ciągły o gęstości f to

$$EX = \int_{-\infty}^{\infty} x f(x) dx.$$

Definicja 5

Para zmiennych losowych (X, Y) *nazywamy* ***wektorem losowym****.*

Definicja 6

Odwzorowanie

$$A \subset \mathbb{R}^2 (\textit{zbiór borelowski}) \mapsto P((X, Y) \in A)$$

nazywamy ***rozkładem wektora losowego****.*

Definicja 7

Kowariancją *wectora* (X, Y) *nazywamy*

$$Cov(X, Y) := E(X - EX)(Y - EY),$$

a ***współczynnikiem korelacji*** *nazywamy*

$$Corr(X, Y) = \rho_{X,Y} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Uwaga 2

Współczynnik korelacji zawsze spełnia $\rho_{X,Y} \in [-1, 1]$, *natomiast* $|\rho_{X,Y}| = 1$ *wtedy i tylko wtedy gdy istnieją liczby* $\beta_0, \beta_1 \in \mathbb{R}$, *takie że* $Y = \beta_0 + \beta_1 X$ *or* $X = \beta_0 + \beta_1 Y$ *z prawdopodobieństwem 1.*

Estymatory.

Niech $(X_1, X_2, ..., X_n)$ będzie **próbą** (niezależne o jednakowym rozkładzie)

$$EX \approx \bar{X} := \frac{1}{n}\sum_{t=1}^{n} X_t \text{ (średnia z próby)}$$

$$Var(X) \approx S_X^2 := \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})^2 \text{ (wariancja z próby)}$$

Niech $((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ będzie **wektorem próbkowym**, wtedy

$$Cov(X, Y) \approx \hat{\gamma}_{X,Y} := \frac{1}{n}\sum_{t=1}^{n}(X_t - \bar{X})(Y_t - \bar{Y}) \text{ (kowariancja z próby)}$$

$$\rho_{X,Y} \approx \frac{\hat{\gamma}_{X,Y}}{S_X S_Y} \text{ (korelacja Pearsona).}$$

Czasami stosujemy powyższe estymatory również dla zależnych zmiennych losowych, ale nie jest to reguła.

Model liniowy

Rozważmy następujący model

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \ldots + \beta_k X_{t,k} + \epsilon_t,$$

dla $t = 1, 2, \ldots, n$, gdzie

- $X_{t,1}, X_{t,2}, \ldots, X_{t,k}$ - zaobserwowane zmienne objaśniające (tzn. znamy ich wartości)
- Y_t - zaobserwowane zmienne objaśniane
- ϵ_t - nieznana wartość składnika losowego (nazywanego czasami białym szumem, lub szokiem)
- $\beta_0, \beta_1, \ldots, \beta_k$ - nieznane parametry.

Przykład 3

Rozważmy model miesięcznej temperatury w danym kraju:

$$Y_t = \beta_0 + \beta_1 t + \beta_2 \cos\left(\frac{\pi}{6} t\right) + \beta_k \sin\left(\frac{\pi}{6} t\right) + \epsilon_t,$$

gdzie Y_t średnia miesięczna temperatura w miesiącu t. Jest to szczególny przypadek modelu liniowego, gdzie

$X_{t,1} = t$ *zmienna czasowa (indeks miesiąca)*

$$\left.\begin{array}{rcl} X_{t,2} & = & \cos\left(\frac{\pi}{6} t\right) \\ X_{t,3} & = & \sin\left(\frac{\pi}{6} t\right). \end{array}\right\} \text{ fale sezonowości.}$$

Innymi słowy, zmienne objaśniające to indeks miesiąca ($X_{t,1}$) oraz dwie fale sezonowości ($X_{t,2}$ oraz $X_{t,3}$).

Przykład 4

Rozważmy przykład z poprzedniego wykładu:

$$\ln(G_t) = \beta_0 + \beta_1 N_t + \beta_2 P_t + \beta_3 K_t + \beta_4 N_t^2 + \beta_5 P_t^2 + \beta_6 K_t^2 + \epsilon_t,$$

- $t = 1, 2, ..., 100$ *indeks sektora;*
- P_t *- dawka nawozu fosforowego dla sektora* t*;*
- N_t *- dawka nawozu azotowego dla sektora* t*;*
- K_t *- dawka nawozu fosforowego dla sektora* t*;*
- G_t *- plony z nawozu* t*;*

Jest to liniowy model ze ***zmiennymi objaśniającymi*** $X_{t,1} = N_t$, $X_{t,2} = P_t$, $X_{t,3} = K_t$, $X_{t,4} = N_t^2$, $X_{t,5} = P_t^2$, $X_{t,6} = K_t^2$, *oraz* ***zmienną objaśnianą*** $Y_t = \ln(G_t)$.

Następujący model

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \ldots + \beta_k X_{t,k} + \epsilon_t,$$

można przekształcić jako układ równań liniowych:

$$\begin{cases} Y_1 & = & \beta_0 + \beta_1 X_{1,1} + \beta_2 X_{1,2} + \ldots + \beta_k X_{1,k} + \epsilon_1 \\ Y_2 & = & \beta_0 + \beta_1 X_{2,1} + \beta_2 X_{2,2} + \ldots + \beta_k X_{2,k} + \epsilon_2 \\ \vdots & \vdots & \vdots \\ Y_n & = & \beta_0 + \beta_1 X_{n,1} + \beta_2 X_{n,2} + \ldots + \beta_k X_{n,k} + \epsilon_n \end{cases}$$

lub w formie macierzowej.

Model liniowy - forma macierzowa

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{\mathbf{Y}} = \underbrace{\begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\epsilon}.$$

Cel

- **Argumenty wejściowe:** dla każdego t znamy wartości zmiennych objaśniaących $X_{t,1}$, $X_{t,2}$, ...,$X_{t,k}$ oraz zmiennej objaśnianej Y_t;
- **Parametry zakłócające:** nie znamy wartości ϵ_t, one nas nawet nie interesują, ale nie możemy ich ignorować, bo jednak utrudniają nam wyciąganie wniosków i podejmowanie decyzji;
- **Argumenty wyjściowe:** staramy się znaleźć $\beta_0, \beta_1, \ldots, \beta_k$ za pomocą **metody najmniejszych kwadratów**: to znaczy znaleźć $\beta_0, \beta_1, \ldots, \beta_k \in \mathbb{R}$ dla których poniższa wartość

$$f(\beta) = \sum_{t=1}^{n} \left(Y_t - \beta_0 - \beta_1 X_{t,1} - \beta_2 X_{t,2} - \ldots - \beta_k X_{t,k} \right)^2$$

jest najmniejsza.

Najprostszy jednowymiarowy model

Rozważmy najprostszy model $k = 1$:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t.$$

Metoda najmniejszych kwadratów sprowadza się do kryterium:

$$f(\beta) = \sum_{t=1}^{n} (Y_t - \beta_0 - \beta_1 X_t)^2$$

Najprostszy jednowymiarowy model

- Obliczamy pochodne cząstkowe:

$$\begin{cases} \frac{\partial f}{\partial \beta_0} & = & -2\sum_{t=1}^{n}\left(Y_t - \beta_0 - \beta_1 X_t\right) \\ \frac{\partial f}{\partial \beta_1} & = & -2\sum_{t=1}^{n}\left(Y_t - \beta_0 - \beta_1 X_t\right)X_t \end{cases}$$

- Obliczając hesjan, macierz drugich pochodnych zauważamy, że jest **dodatnio określona**, stąd f jest **funkcją wypukłą** (omijamy szczegóły).

Najprostszy jednowymiarowy model

Stąd jeśli gradient funkcji f znika w $\beta = (\beta_0, \beta_1)$, wtedy f osiąga minimum w β.

- Rozwiązujemy:

$$\begin{cases} \frac{\partial f}{\partial \beta_0} = -2\sum_{t=1}^{n}(Y_t - \beta_0 - \beta_1 X_t) = 0 \\ \frac{\partial f}{\partial \beta_1} = -2\sum_{t=1}^{n}(Y_t - \beta_0 - \beta_1 X_t) X_t = 0 \end{cases}$$

- Zatem

$$\begin{cases} \beta_0 n + \beta_1 \sum_{t=1}^{n} X_t = \sum_{t=1}^{n} Y_t \\ \beta_0 \sum_{t=1}^{n} X_t + \beta_1 \sum_{t=1}^{n} X_t^2 = \sum_{t=1}^{n} X_t Y_t \end{cases}$$

Najprostszy jednowymiarowy model

Rozwiązując powyższy układ równań liniowych otrzymamy

$$\begin{cases} \hat{\beta}_0 & = & \dfrac{\sum\limits_{t=1}^{t} Y_t \sum\limits_{t=1}^{n} X_t^2 - \sum\limits_{t=1}^{n} X_t Y_t \sum\limits_{t=1}^{n} \sum\limits_{t=1}^{n} X_t}{n \sum\limits_{t=1}^{n} X_t^2 - \left(\sum\limits_{t=1}^{n} X_t \right)^2} \\ \hat{\beta}_1 & = & \dfrac{n \sum\limits_{t=1}^{n} X_t Y_t - \sum\limits_{t=1}^{n} X_t \sum\limits_{t=1}^{n} Y_t}{n \sum\limits_{t=1}^{n} X_t^2 - \left(\sum\limits_{t=1}^{n} X_t \right)^2} \end{cases}$$

Uwaga 3

*Zwykle oznaczamy $\hat{\theta}$ jako estymator parametru θ, aby powiązać i jednocześnie rozróżnić estymator od parametru. W tym przypadku β_0 i β_1 to nieznane **parametry**, a $\hat{\beta}_0$ oraz $\hat{\beta}_1$ to ich **estymatory**.*

Najprostszy jednowymiarowy model

Możemy uprościć powyższe wzory za pomocą estimatorów $\hat{\gamma}_{X,Y}$, S_X^2, S_Y^2, $\bar{X}$, oraz $\bar{Y}$ i orzymujemy równoważne wzory na estymatory β_0, β_1:

$$\begin{cases} \hat{\beta}_1 & = & \frac{\hat{\gamma}_{X,Y}}{S_X^2} \\ \hat{\beta}_0 & = & \bar{Y} - \hat{\beta}_1 \bar{X} \end{cases}$$

Przykład -populacja wilków i zajęcy

Przykład

Przykład 5 (Wilki i zające)

Pewien zoolog bada związek między populacją wilków i zajęcy poprzez obserwację populacji zwierząt w 4 lasach. Wyniki są zraportowane w tabeli:

	Las 1	*Las 2*	*Las 3*	*Las 4*
Populacja zajęcy	140	130	150	120
Populacja wilków	30	15	35	36

Znajdź najlepiej dopasowany model liniowy.

Mamy

- Y_t - populacja zajęcy, zmienna objaśniana
- X_t - populacja wilk'ow, zmienna objaśniająca.

Metoda najmniejszych kwadratów sprowadza się do minimalizacji wyrażenia

$$f(\beta) = (140 - \beta_0 - 30\beta_1)^2 + (130 - \beta_0 - 15\beta_1)^2$$

$$+(150 - \beta_0 - 35\beta_1)^2 + (120 - \beta_0 - 36\beta_1)^2.$$

Przyrównujemy pierwsze pochodne cząstkowe do 0:

$$\frac{\partial f}{\partial \beta_0} = -2(140 - \beta_0 - 30\beta_1) - 2(130 - \beta_0 - 15\beta_1)$$

$$-2(150 - \beta_0 - 35\beta_1) - 2(120 - \beta_0 - 36\beta_1) = 0$$

$$\frac{\partial f}{\partial \beta_1} = -60(140 - \beta_0 - 30\beta_1) - 30(130 - \beta_0 - 15\beta_1)$$

$$-70(150 - \beta_0 - 35\beta_1) - 72(120 - \beta_0 - 36\beta_1) = 0.$$

Przekształcamy wyrażenie

$$\frac{\partial f}{\partial \beta_0} = -2(140 - \beta_0 - 30\beta_1) - 2(130 - \beta_0 - 15\beta_1)$$

$$-2(150 - \beta_0 - 35\beta_1) - 2(120 - \beta_0 - 36\beta_1) = 0$$

$$\frac{\partial f}{\partial \beta_1} = -60(140 - \beta_0 - 30\beta_1) - 30(130 - \beta_0 - 15\beta_1)$$

$$-70(150 - \beta_0 - 35\beta_1) - 72(120 - \beta_0 - 36\beta_1) = 0.$$

Rozwiązanie przykładu - pierwszy sposób

$$\begin{cases} 4\beta_0 + 116\beta_1 = 540 \\ 116\beta_0 + 3646\beta_1 = 15720 \end{cases}$$

Za pomocą standardowych wzorów Cramér:

$$\hat{\beta}_0 = 128.83, \quad \hat{\beta}_1 = 0.21.$$

Zatem najlepiej dopasowany model liniowy to

$$Y_t = 0.21X_t + 128.83 + \epsilon_t$$

Rozwiązanie przykładu - pierwszy sposób

Z powodu licznych przekształceń algebraicznych, powyższy sposób jest żmudny. Dla wygody lepiej zastosować gotowy wzór:

$$\begin{cases} \hat{\beta}_1 & = & \frac{\hat{\gamma}_{X,Y}}{S_X^2} \\ \hat{\beta}_0 & = & \bar{Y} - \hat{\beta}_1 \bar{X} \end{cases}$$

Rozwiązanie przykładu - drugi sposób

Mamy $(X_1, X_2, X_3, X_4) = (30, 15, 35, 36)$ and $(Y_1, Y_2, Y_3, Y_4) = (140, 130, 150, 120)$. Z definicji

$$\bar{X} = \frac{1}{4}(30+15+35+36) = 29, \quad \bar{Y} = \frac{1}{4}(140+130+150+120) = 135.$$

Dalej,

$$S_X^2 = \frac{1}{4}\left((30-29)^2 + (15-29)^2 + (35-29)^2 + (36-29)^2\right) = 70.5,$$

$$S_Y^2 = \frac{1}{4}\left((140-135)^2 + (130-135)^2 + (150-135)^2 + (120-135)^2\right) = 125,$$

Rozwiązanie przykładu - drugi sposób

oraz

$$\hat{\gamma}_{X,Y} = \frac{1}{4}((30-29)(140-135)+(15-29)(130-135)$$

$$+(35-29)(150-135)+(36-29)(120-135)) = 15.$$

Na koniec podstawiamy do gotowego wzoru:

$$\hat{\beta}_1 = \frac{\hat{\gamma}_{X,Y}}{S_X^2} = \frac{15}{70.5} = 0.21277 \approx 0.21$$

oraz

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_0\bar{X} = 135 - 0.21 * 29 = 128.83.$$

Uwaga 4

Obie metody są algebraicznie równoważne. Stąd wyniki otrzymane za pomocą obu metod są powinny być podobne z dokładnością do ***błędów zaokrągleń***.