Testowanie hipotez

Ekonometria WNE

Sebastian Zalas

University of Warsaw
s.zalas@uw.edu.pl

Klasyczny Model Regresji Liniowej

1. $\boldsymbol{y} = \boldsymbol{X}\beta + \varepsilon$ - model jest liniowy
2. Zmienne losowe $\{(y_1, X_1), ..., (y_i, y_i), ..., (y_n, y_n)\}$ są niezależne oraz wylosowane z tego samego rozkładu (*independently and identically distributed - iid.*)
3. $rz[\boldsymbol{X}_{n \times k}] = k$ - rząd kolumnowy $\boldsymbol{X}$ jest pełny
4. $\mathbb{E}[\varepsilon|\boldsymbol{X}] = 0$ - wartość oczekiwana składnika losowego jest równa 0
5. $\mathbb{E}[\varepsilon\varepsilon'|\boldsymbol{X}] = \boldsymbol{I}\sigma^2$ - sferyczność wariancji
6. $\varepsilon|\boldsymbol{X} \sim \mathcal{N}(0, \boldsymbol{I}\sigma^2)$ składnik losowy ma rozkład normalny

Rozkład estymatora MNK

- w KMRL mamy następujące założenie:

$$\varepsilon \mid \boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}\sigma^2) \tag{1}$$

na jego podstawie wyprowadzimy rozkład estymatora MNK.

- Przepiszemy (1) założenie korzystając z $\hat{\beta} - \beta = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon$

$$\begin{aligned} \hat{\beta} - \beta \mid \boldsymbol{X} &\sim (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathcal{N}(0, \boldsymbol{I}\sigma^2) \\ &\sim \mathcal{N}(0, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}) \\ &= \mathcal{N}(0, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}) \end{aligned}$$

- co oznacza, że:

$$\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}) \tag{2}$$

Statystyka t

- Możemy znormalizować (2):

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2[(\boldsymbol{X}'\boldsymbol{X})^{-1}]_{j,j}}} \sim \mathcal{N}(0, 1)$$

- Nie obserwujemy σ^2; należy skorzystać z oszacowania:

$$t = \frac{\hat{\beta}_j - \beta_j}{\sqrt{s^2[(\boldsymbol{X}'\boldsymbol{X})^{-1}]_{j,j}}} \sim t_{n-k-1}$$

- Wtedy statystyka t ma rozkład t-studenta z $n - k - 1$ stopniami swobody

Test t

- Mamy prosty model:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- korzystając ze statystyki T możemy zweryfikować hipotezę o β_1:

$$H_0 : \beta_1 = d$$

$$H_1 : \beta_1 \neq d$$

- Przy założeniu, że hipoteza zerowa jest prawdziwa, obliczamy wartość **statystyki testowej**:

$$t = \frac{\hat{\beta} - d}{se(\hat{\beta})} \sim t(n - k - 1)$$

TESTOWANIE HIPOTEZY

- następnie ustalamy **poziom istotności** α; standardowo przyjmuje się $\alpha = 0.05$
- znajdujemy **wartości krytyczne**: $t^*_{\frac{\alpha}{2},n-k-1}$ oraz $t^*_{1-\frac{\alpha}{2},n-k-1}$
- przyjmujemy hipotezę zerową jako prawdziwą, jeżeli statystyka testowa znajduje się pomiędzy wartościami krytycznymi. W przeciwnym przypadku, odrzucamy hipotezę zerową na rzecz alternatywnej.
- alternatywnie, jeżeli $p - value < \alpha$ to odrzucamy hipotezę zerową

Testowanie istotności

- Sprowadza się do przetestowania hipotezy w postaci:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

- obliczmy wartość statystyki t:

$$T = \frac{\hat{\beta}_k}{se(\hat{\beta})} \sim t(n-k-1)$$

- znajdujemy **wartości krytyczne**: $t^*_{\frac{\alpha}{2},n-k-1}$ oraz $t^*_{1-\frac{\alpha}{2},n-k-1}$, bądź p – *value* i przeprowadzamy wnioskowanie.

ROZKŁAD t - STUDENTA

Obszary akceptacji i odrzucenia H_0 w teście t z obustronną H_1

Przedział ufności

- Estymator przedziałowy $\hat{C}$ taki, że będzie on zawierał prawdziwy parametr β_1 z ustalonym prawdopodobieństwem równym $1 - \alpha$.

$$\mathbb{P}\left(\beta \in \hat{C}\right) = 1 - \alpha$$

- Jest to prawdopodobieństwo. że **losowy** przedział $\hat{C}$ zawiera prawdziwy, **ustalony** parametr β_1

- Możemy skonstruować przedział ufności w oparciu o statystykę T i odchylenie standardowe β:

$$\mathbb{P}\left(t^*_{\frac{\alpha}{2},n-k-1} \leq t \leq t^*_{1-\frac{\alpha}{2},n-k-1}\right) = 1 - \alpha$$

$$\mathbb{P}\left(t^*_{\frac{\alpha}{2},n-k-1} \leq \frac{\beta_1 - \hat{\beta}_1}{se(\hat{\beta}_1)} \leq t^*_{1-\frac{\alpha}{2},n-k-1}\right) = 1 - \alpha$$

Przedziały ufności

- Po przekształceniach otrzymujemy:

$$\mathbb{P}\left(\hat{\beta}_1 + se(\hat{\beta}_1)\, t^*_{\frac{\alpha}{2},n-k-1} \leq \beta_1 \leq \hat{\beta}_1 + se(\hat{\beta}_1)\, t^*_{1-\frac{\alpha}{2},n-k-1}\right) = 1 - \alpha$$

- Rozkład t – *studenta* jest symetryczny, więc $\left|t^*_{\frac{\alpha}{2},n-k-1}\right| = t^*_{1-\frac{\alpha}{2},n-k-1}$

- Wtedy z prawdopodobieństwem 0.95:

$$\hat{\beta}_1 - se(\hat{\beta}_1)\, t^*_{1-\frac{\alpha}{2},n-k-1} \leq \beta_1 \leq \hat{\beta}_1 + se(\hat{\beta}_1)\, t^*_{1-\frac{\alpha}{2},n-k-1}$$

Co jeżeli ε nie ma rozkładu normalnego?

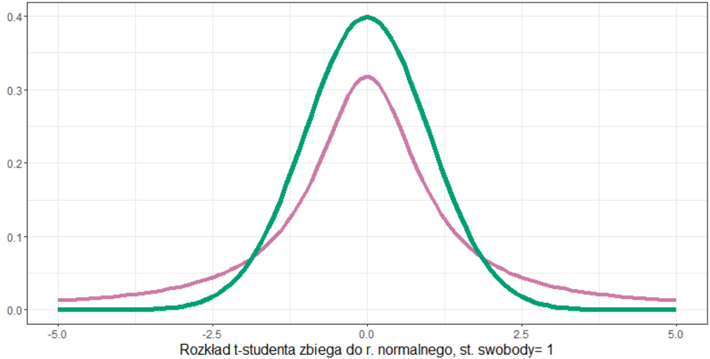
- jeśli $\varepsilon \not\sim \mathcal{N}(0, \sigma^2)$ to wtedy statystyka T nie ma rozkładu t – *studenta* i nie możemy przeprowadzić testu

- Możemy skorzystać z własności asymptotycznych. Na podstawie Centralnego Twierdzenia Granicznego:

$$\frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \xrightarrow{as.} \mathcal{N}(0, 1)$$

- nawet gdy składnik losowy nie ma rozkładu normalnego, nadal możemy przeprowadzić test jeśli mamy **wystarczająco dużą próbę**. Jednak wtedy odczytujemy wartości krytyczne ze standardowego rozkładu normalnego.

Rozkład t – *studenta* $\longrightarrow \mathcal{N}(0,1)$



Test F

- Test F pozwala na przetestowanie wielu restrykcji (warunków), które chcemy nałożyć na model.
- Mamy dany model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

- chcemy sprawdzić czy $\beta_1 = \beta_2 = 0$, zatem:

$$H_0 : \begin{cases} \beta_1 = 0 \\ \beta_2 = 0 \end{cases}$$

$$H_1 : \beta_1 \neq 0 \text{ lub } \beta_2 \neq 0$$

Test F

- szacujemy model bez restrykcji (oznaczmy go literą U):

$$y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + e$$

- następnie szacujemy z restrykcjami (oznaczmy go literą R):

$$y = \hat{\beta}_0 + \hat{\beta}_3 x_3 + e$$

- budujemy statystykę testową:

$$F = \frac{(SSR_R - SSR_U)\,\frac{1}{q}}{SSR_U \frac{1}{n-k-1}} = \frac{(R_U^2 - R_R^2)\,\frac{1}{q}}{(1 - R_U^2)\frac{1}{n-k-1}} \sim F(1-\alpha, q, n-k-1)$$

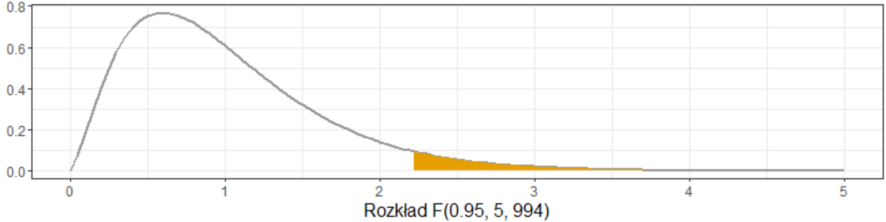
gdzie q to liczba restrykcji

Test F

- Korzystamy z rozkładu F z $(q, b-k-1)$ stopniami swobody by wyznaczyć wartość krytyczną F^*. Przyjmujemy H_0 jeśli $F < F^*$

- Im większa różnica $R^2_U - R^2_R \Rightarrow$ tym wyższa statystyka $F \Rightarrow$ lepsze dopasowanie do danych dzięki dodaniu zmiennych

Obszary akceptacji i odrzucenia H_0 w teście F



Test F łącznej istotności

- Model bez restrykcji:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + \varepsilon$$

- Model z ograniczeniami ma postać:

$$y = \beta_0 + \varepsilon$$

- Hipoteza zerowa i alternatywna:

$$H_0 : \begin{cases} \beta_1 = 0 \\ \beta_2 = 0 \\ \vdots \\ \beta_k = 0 \end{cases}$$

H_1 : chociaż jeden parametr <u>nie</u> jest równy zero

Test F łącznej istotności

- budujemy statystykę testową opartą na SSE lub R^2, pamiętając że w modelu tylko ze stałą $R^2 = 0$. Następnie szukamy wartości krytycznej korzystając z rozkładu $F(1 - \alpha, q, n - k - 1)$
- statystyka F oraz odpowiadające jej *p-value* są raportowane w pakietach statystycznych, np. w `R`.
- sprawdzamy czy nasz model jest lepszy niż model z tylko stałą

Restrykcje w zapisie macierzowym

- Hipotezy dotyczące kilku parametrów możemy zapisać w formie macierzowej:

$$\boldsymbol{H}\beta = \boldsymbol{h}$$

gdzie $\boldsymbol{H}$ to macierz $(q \times (k+1))$ opisująca restrykcje, q to liczba restrykcji, $\boldsymbol{h}$ to wektor stałych z każdej restrykcji.

- Przykład – test łącznej istotności:

$$\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{q \times k+1} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{k+1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{q}$$

Restrykcje w zapisie macierzowym

- Przykład – test następujących restrykcji:
  1. $\beta_1 = \beta_3$
  2. $\beta_2 = a$
  3. $\beta_1 + \beta_4 = b$

$$\begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ a \\ b \end{bmatrix}$$

Pytania? Wątpliwości?

Dziękuję!

e: s.zalas@uw.edu.pl