Modele liniowe

Michał Kos

Uniwersytet Wrocławski

Plan wykładu

Table of Contents

Regresja liniowa wieloraka

Pomysł zawarty w regresji liniowej prostej można łatwo uogólnić na sytuację w której zmiennych objaśniających jest więcej niż jedna. Wówczas z każdą wartością zmienej odpowiedzi Y_i stowarzyszony jest wektor wartości kilku regresorów $X_i = (X_{i1}, ..., X_{ip-1})'$:

Model teoretyczny regresji liniowej wielorakiej w postaci skalarnej

Teoretyczny model dla regresji liniowej wielorakiej:

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + ... + X_{ip-1}\beta_{p-1} + \epsilon_i \quad i = 1, ..., n$$

gdzie: $\epsilon_1, ..., \epsilon_n$ – błędy losowe, ciąg niezależnych zmiennych losowych z rozkładu normalnego o średniej 0 i wariancji σ^2
X_{ij} – wartość j–tej zmiennej niezależnej stowarzyszona z i–tą obserwacją zmiennej zależnej Y_i,
β_0 – Intercept,
$\beta_k \quad k = 1, ..., p-1$ – parametry regresji związane z odpowiadającymi im zmiennymi objaśniającymi ("Slope"-y).

Łatwo zauważyć, że takie uogólnienie jest mocno pożądane. Zwykle badacz podejrzewa, że na badane zjawisko ma wpływ więcej niż jedna cecha.

Regresja liniowa wieloraka – ważne przykłady

Łatwo zauważyć, że takie uogólnienie jest mocno pożądane. Zwykle badacz podejrzewa, że na badane zjawisko ma wpływ więcej niż jedna cecha.

Przykład 1
Badania medyczne wskazują na to, że ciśnienie pacjenta Y_i zależy od wielu czynników np. wiek, BMI, przewlekłe stany chorobowe, pula genetyczna $X_i = (X_{i1}, ..., X_{ip-1})'$.

Regresja liniowa wieloraka – ważne przykłady

Łatwo zauważyć, że takie uogólnienie jest mocno pożądane. Zwykle badacz podejrzewa, że na badane zjawisko ma wpływ więcej niż jedna cecha.

Przykład 1
Badania medyczne wskazują na to, że ciśnienie pacjenta Y_i zależy od wielu czynników np. wiek, BMI, przewlekłe stany chorobowe, pula genetyczna $X_i = (X_{i1}, ..., X_{ip-1})'$.

Stosowanie regresji liniowej wielorakiej umożliwia równoczesną analizę wpływu kilku cech. Ponadto, (przy odpowiedniej wstępnej obróbce danych) daje możliwość porównywania siły poszczególnych cech.

Regresja liniowa wieloraka – ważne przykłady (relacje nieliniowe)

Regresja liniowa wieloraka pozwala na modelowanie relacji nieliniowych.

Regresja liniowa wieloraka – ważne przykłady (relacje nieliniowe)

Regresja liniowa wieloraka pozwala na modelowanie relacji nieliniowych.

Przykład 2
Z fizyki wiemy, że przemieszczenie w swobodnym spadku w polu grawitacyjnym ziemi wyraża się następującym wzorem:

$$Y(t) = \beta_0 + t\beta_1 + t^2\beta_2$$

gdzie: β_0 – początkowe położenie ciała, β_1 – początkowa szybkość ciała (w kierunku y), $\beta_2 = -\frac{g}{2}$ g – przyspieszenie ziemskie, t – czas, Y(t) – położenie w chwili t.
Powyższa relacja w oczywisty sposób jest funkcją kwadratową. W związku z tym dla par obserwacji $\{(Y_i, t_i)\}_{i=1}^{n}$ model uzyskany przy użyciu regresji liniowej prostej (dopasowujący prostą do danych; $\hat{Y}_i = \hat{\beta}_0 + t_i\hat{\beta}_1$) będzie nieadekwatny.

Regresja liniowa wieloraka – ważne przykłady (relacje nieliniowe)

Przykład 2 c.d.

$$Y(t) = \beta_0 + t\beta_1 + t^2\beta_2$$

Jednakże, alternatywnie, do zbioru danych $\{(Y_i, t_i)\}_{i=1}^n$ możemy użyć modelu regresji wielorakiej następującej postaci:

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i$$

gdzie $X_{i1} = t_i$, $X_{i2} = t_i^2$, oraz ϵ_i – błąd związany z pomiarem położenia ciała (Y_i) w chwili t_i.

Regresja liniowa wieloraka – ważne przykłady (relacje nieliniowe)

Przykład 2 c.d.

$$Y(t) = \beta_0 + t\beta_1 + t^2\beta_2$$

Jednakże, alternatywnie, do zbioru danych $\{(Y_i, t_i)\}_{i=1}^n$ możemy użyć modelu regresji wielorakiej następującej postaci:

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i$$

gdzie $X_{i1} = t_i$, $X_{i2} = t_i^2$, oraz ϵ_i – błąd związany z pomiarem położenia ciała (Y_i) w chwili t_i.

Powyższy model umożliwia dopasowanie funkcji kwadratowej do danych. Stosując analogiczne rozumowanie możemy dopasować do danych np. wielomian dowolnego stopnia lub inne nieliniowe relacje.

Regresja liniowa wieloraka – ważne przykłady (interakcja)

Czasami naturalnym wydaje się, że równoczense występowanie kilku czynników ma dodatkowy pozytywny wpływ na zmienną odpowiedzi Y.

Regresja liniowa wieloraka – ważne przykłady (interakcja)

Czasami naturalnym wydaje się, że równoczense występowanie kilku czynników ma dodatkowy pozytywny wpływ na zmienną odpowiedzi Y.

Przykład 3
Na zbiory rolnicze (Y) ma wpływ np. skład podłoża (np. nawóz) oraz odpowiednie nawodnienie. Możemy zatem zastosować następujący model:

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i \quad i = 1, ..., n$$

gdzie X_{i1} – zmienna niezależna będąca indykatorem opisującym to czy na polu Y_i zastosowano nawóz ($X_{i1} = 1$) czy nie zastosowano nawożenia ($X_{i1} = 0$),
X_{i2} – zmienna niezależna opisując poziom nawodnienia obszaru Y_i np. ilość wody na m^2 w jednostce czasu.

Regresja liniowa wieloraka – ważne przykłady (interakcja)

Przykład 3 c.d.

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i \quad i = 1, ..., n$$

Zgodnie z powyższym modelem zysk dla każdego czynnika (nawożenie, nawadnianie) jest niezależny. Wydaje się jednak, że równoczesne zastosowanie optymalnych nawozu i nawodnienia może skutkować pewnym "bonusowym" plonem ("1+1 = 5").

Regresja liniowa wieloraka – ważne przykłady (interakcja)

Przykład 3 c.d.

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \epsilon_i \quad i = 1, ..., n$$

Zgodnie z powyższym modelem zysk dla każdego czynnika (nawożenie, nawadnianie) jest niezależny. Wydaje się jednak, że równoczesne zastosowanie optymalnych nawozu i nawodnienia może skutkować pewnym "bonusowym" plonem ("1+1 = 5").

Powyższą obserwację możemy uwzględnić wykorzystując tzw. model liniowy z interakcjami:

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + X_{i1}X_{i2}\beta_3 + \epsilon_i \quad i = 1, ..., n$$

W powyższym modelu składnik $X_{i1}X_{i2}\beta_3$ opisuje "bonusowy" wpływ interakcji pomiędzy zmiennymi X_{i1} i X_{i2}. Wprowadzając oznaczenie $X_{i3} = X_{i1}X_{i2}$ otrzymujemy model regresji liniowej wielorakiej postaci:

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + X_{i3}\beta_3 + \epsilon_i \quad i = 1, ..., n$$

Table of Contents

Model teoretyczny w postaci skalarnej

Model teoretyczny w postaci skalarnej

Poznaliśmy już teoretyczny model dla regresji liniowej prostej w zapisie skalarnym:

$$Y_i = \beta_0 + X_{i1}\beta_1 + X_{i2}\beta_2 + \ldots + X_{ip-1}\beta_{p-1} + \epsilon_i \quad i = 1, \ldots, n$$

gdzie $\epsilon_1, \ldots, \epsilon_n$ – ciąg niezależnych zmiennych losowych z rozkładu normalnego o średniej 0 i wariancji σ^2.

Model ten bardzo często zapisywany jest alternatywnie w postaci macierzowej:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Model teoretyczny w postaci macierzowej – oznaczenia

Wprowadźmy następujące oznaczenia:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}; \quad \mathbb{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np-1} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}; \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

gdzie:
$Y \in \mathbb{R}^n$ nazywamy wektorem odpowiedzi,
$\mathbb{X} \in M_{n \times p}$ nazywamy macierzą planu lub eksperymentu (deterministyczna macierz),
$\beta \in \mathbb{R}^p$ nazywamy wektorem parametrów (deterministyczny wektor),
$\epsilon \in \mathbb{R}^n$ nazywamy wektorem błędów losowych.

Model teoretyczny w postaci macierzowej

Model teoretyczny w postaci macierzowej

Model teoretyczny dla regresji liniowej wielorakiej w zapisie macierzowym ma postać:

$$Y = \mathbb{X}\beta + \epsilon$$

gdzie $\epsilon \in \mathbb{R}^n$ jest wektorem losowym z rozkładu normalnego $N(0_n, \sigma^2 \mathbb{I}_n)$.

Model teoretyczny w postaci macierzowej

Model teoretyczny w postaci macierzowej

Model teoretyczny dla regresji liniowej wielorakiej w zapisie macierzowym ma postać:

$$Y = \mathbb{X}\beta + \epsilon$$

gdzie $\epsilon \in \mathbb{R}^n$ jest wektorem losowym z rozkładu normalnego $N(0_n, \sigma^2 \mathbb{I}_n)$.

Ma on zatem dokładnie tą samą postać co teoretyczny model regresji liniowej prostej. Jedyną róznicą jest liczba kolumn w macierzy planu $\mathbb{X}$ i liczba parametrów w wektorze β:

$$\begin{array}{c|ccccc} \text{równanie} & Y & = & \mathbb{X} & \beta & + & \epsilon \\ \text{wymiary (macierzowo)} & n \times 1 & & n \times p & p \times 1 & & n \times 1 \end{array}$$

Model teoretyczny w postaci macierzowej – oznaczenia

Uwaga o notacji w wektorze β i macierzy $\mathbb{X}$

W wektorze $\beta = (\beta_0, \beta_1, ..., \beta_{p-1})'$ jest p elementów, przy czym numerujemy je od 0 do $p-1$, zachowując "wyjątkowość" parametru β_0 jako interceptu. Pozostałe $p-1$ elemetów opisuje wpływ poszczególnych regresorów.
Podobnie w macierzy planu $\mathbb{X}$ jest p kolumn. Pierwsza kolumna to ciąg jedynek $(\forall i)\ \ \mathbb{X}_{i1} = 1$ stowarzyszony z Interceptem. Kolejne kolumny opisują zachowanie poszczególnych zmiennych niezależnych dla kolejnych obserwacji $(\forall i, j \geqslant 2)\ \ \mathbb{X}_{ij} = X_{ij-1}$

$$\mathbb{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np-1} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix};$$

Własności wektora odpowiedzi Y

Tw. 1

Jeżeli $B = c + DA \in \mathbb{R}^m$, jest afinicznym przekształceniem wektora losowego $A \sim N(\mu, \Sigma) \in \mathbb{R}^n$, gdzie $c \in \mathbb{R}^m$ jest deterministycznym wektorem oraz $D \in \mathbb{M}_{m \times n}$ jest deterministyczną macierzą, wówczas:

$$B \sim N(c + D\mu, D\Sigma D')$$

Na podstawie Tw. 1 otrzymujemy w natychmiastowy sposób że:

$$Y \sim N(\mathbb{X}\beta, \sigma^2 \mathbb{I})$$

Czyli:

1. Wektor Y pochodzi z n–wymiarowego rozkładu normalnego,
2. Wartość oczekiwana jest liniowym przekształceniem: $E(Y) = \mathbb{X}\beta$,
3. Macierz kowariancji jest postaci: $Cov(Y) = \sigma^2 \mathbb{I}$:
    - $Y_1, ..., Y_n$ jest ciągiem niezależnych zmiennych losowych o stałej wariancji σ^2.

Table of Contents

Metoda najmniejszych kwadratów

Kryterium estymacji parametrów metodą najmniejszych kwadratów w regresji liniowej wielorakiej:

$$\hat{\beta} = \arg\min_b \sum_{i=1}^{n} (Y_i - (b_0 + X_{i1}b_1 + ... + X_{ip-1}b_{p-1}))^2 = \arg\min_b \sum_{i=1}^{n} e_i^2(b)$$

gdzie $b = (b_0, b_1, ..., b_{p-1})' \in \mathbb{R}^p$, $e_i(b)$ – ciąg residuów stowarzyszony z wektorem b.

W notacji macierzowej możemy zapisać powyższe równanie w następującej postaci:

$$\hat{\beta} = \arg\min_b \| Y - \mathbb{X}b\|_2^2 = \arg\min_b \|e(b)\|_2^2$$

Residua dla estymatora OLS

W przypadku estymatora najmniejszych kwadratów we wzorze na residuum (wektor residuów) będziemy pomijać argument funkcji:

$$e_i := e_i(\hat{\beta}); \quad e := e(\hat{\beta})$$

Estymator metody najmniejszych kwadratów (1)

Zapis macierzowy umożliwia elegancki sposób wyrażenia estymatora metody najmniejszych kwadratów. Wyznaczmy pochodną z funkcji $\|Y - \mathbb{X}b\|_2^2$ po współczynnikach:

$$\partial_{b_0}\|Y - \mathbb{X}b\|_2^2 = \partial_{b_0}\sum_{i=1}^{n}(Y_i - (\mathbb{X}_{i1}b_0 + \mathbb{X}_{i2}b_1 + \ldots + \mathbb{X}_{ip}b_{p-1}))^2 =$$

$$= \sum_{i=1}^{n} -2\mathbb{X}_{i1}(Y_i - (\mathbb{X}_{i1}b_0 + \mathbb{X}_{i2}b_1 + \ldots + \mathbb{X}_{ip}b_{p-1})) =$$

$$= -2\mathbb{X}'_{\bullet 1}(Y - \mathbb{X}b) = -2\mathbb{X}'_{\bullet 1}e(b)$$

gdzie $\mathbb{X}_{\bullet j}$ oznacza j-tą kolumnę w macierzy $\mathbb{X}$.

Estymator metody najmniejszych kwadratów (2)

W analogiczny sposób można uzyskać pochodną po dowolnym b_i. Podsumowując, pochodna po b_i z funkcji $\|Y - \mathbb{X}b\|_2^2$ jest równa przeskalowanemu (-2) iloczynowi skalarnemu pomiedzy kolumną związaną z i-tym współczynnikiem w wektorze b a wektorem residów $e(b)$:

$$\partial_{b_i}\|Y - \mathbb{X}b\|_2^2 = -2\mathbb{X}'_{\bullet i}(Y - \mathbb{X}b)$$

Zapis wykorzystujący gradient funkcji $(\nabla_s f(s) = (\partial_{s_1} f(s), ..., \partial_{s_d} f(s))')$ umożliwia uzyskanie jeszcze bardziej zwartej postaci:

Gradient funkcji $\|Y - \mathbb{X}b\|_2^2$

$$\nabla_b\|Y - \mathbb{X}b\|_2^2 = -2\mathbb{X}'(Y - \mathbb{X}b)$$

Estymator metody najmniejszych kwadratów (3)

Wykonując dalsze przekształcenia łatwo jest wyznaczyć hesjan funkcji $\|Y - \mathbb{X}b\|_2^2$:

$$Hes_{jk} = \partial_{b_{k-1}}(\partial_{b_{j-1}}\|Y - \mathbb{X}b\|_2^2) =$$

$$= \partial_{b_{k-1}} \sum_{i=1}^{n} -2\mathbb{X}_{ij}(Y_i - (\mathbb{X}_{i1}b_0 + \mathbb{X}_{i2}b_1 + \ldots + \mathbb{X}_{ip}b_{p-1})) =$$

$$= \sum_{i=1}^{n} 2\mathbb{X}_{ij}\mathbb{X}_{ik} = 2(\mathbb{X}'\mathbb{X})_{jk}$$

Zatem hesjan funkcji $\|Y - \mathbb{X}b\|_2^2$ w macierzowej postaci ma formę:

Hesjan funkcji $\|Y - \mathbb{X}b\|_2^2$

$$Hes = 2\mathbb{X}'\mathbb{X}$$

Estymator metody najmniejszych kwadratów (4)

Łatwo pokazać, że hesjan $Hes = 2\mathbb{X}'\mathbb{X}$ jest dodatnio określony:
Niech $v \in \mathbb{R}^n$ będzie dowolnym niezerowym wektorem. Wówczas

$$v'(Hes)v = 2v'\mathbb{X}'\mathbb{X}v = 2(\mathbb{X}v)'(\mathbb{X}v) = 2\|\mathbb{X}v\|_2^2 \geqslant 0$$

W konsekwencji, ekstremum wyznaczone za pomocą przyrównania gradientu do zera musi być minimum dla funkcji $\|Y - \mathbb{X}b\|_2^2$ i stanowi estymator OLS:

$$-2\mathbb{X}'(Y - \mathbb{X}\hat{\beta}) = 0 \quad \Leftrightarrow \quad \mathbb{X}'\mathbb{X}\hat{\beta} = \mathbb{X}'Y$$

Ostatecznie uzyskujemy:

Estymator metody najmniejeszych kwadratów

Załóżmy, że macierz $\mathbb{X}'\mathbb{X}$ jest odwracalna ($det(\mathbb{X}'\mathbb{X}) \neq 0$). Wówczas estymator metody najmniejszych kwadratów ma postać:

$$\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y$$

Uwaga!
Można pokazać, że założenie o odwracalności macierzy $\mathbb{X}'\mathbb{X}$ jest równoważne liniowej niezależności kolumn macierzy $\mathbb{X}$.
W wielu sytuacjach jest to naturalny warunek w kontekście modeli liniowych i od tej pory na wykładzie będzie on zakładany.

Estymator metody największej wiarogodności (MLE)

Aby wyznaczyć estymator największej wiarogodności musimy znaleźć wektor maksymalizujący funkcję wiarogodności dla zbioru zmiennych odpowiedzi $Y_1, ..., Y_n$:

$$L(b, z^2|Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi z^2}} \exp\left(-\frac{1}{2z^2}(Y_i - (\mathbb{X}_{i1}b_0 + \mathbb{X}_{i2}b_1 + ... + \mathbb{X}_{ip}b_{p-1}))^2\right)$$

Estymator metody największej wiarogodności (MLE)

Aby wyznaczyć estymator największej wiarogodności musimy znaleźć wektor maksymalizujący funkcję wiarogodności dla zbioru zmiennych odpowiedzi $Y_1, ..., Y_n$:

$$L(b, z^2|Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi z^2}} \exp\left(-\frac{1}{2z^2}(Y_i - (\mathbb{X}_{i1}b_0 + \mathbb{X}_{i2}b_1 + ... + \mathbb{X}_{ip}b_{p-1}))^2\right)$$

Logarytm funkcji wiarogodności:

$$\log(L(b, z^2|Y)) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log z^2 - \frac{1}{2z^2}\sum_{i=1}^{n}(Y_i - (\mathbb{X}_{i1}b_0 + \mathbb{X}_{i2}b_1 + ... + \mathbb{X}_{ip}b_{p-1}))^2$$

Estymator metody największej wiarogodności (MLE)

Aby wyznaczyć estymator największej wiarogodności musimy znaleźć wektor maksymalizujący funkcję wiarogodności dla zbioru zmiennych odpowiedzi $Y_1, ..., Y_n$:

$$L(b, z^2|Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi z^2}} \exp\left(-\frac{1}{2z^2}(Y_i - (\mathbb{X}_{i1}b_0 + \mathbb{X}_{i2}b_1 + ... + \mathbb{X}_{ip}b_{p-1}))^2\right)$$

Logarytm funkcji wiarogodności:

$$\log(L(b, z^2|Y)) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log z^2 - \frac{1}{2z^2}\sum_{i=1}^{n}(Y_i - (\mathbb{X}_{i1}b_0 + \mathbb{X}_{i2}b_1 + ... + \mathbb{X}_{ip}b_{p-1}))^2$$

Z powyższej postaci widać, że dla dowolne wartości z^2, maksymalizacja po wektorze b jest równoważna minimalizacji wyrażenia:

$$\sum_{i=1}^{n}(Y_i - (\mathbb{X}_{i1}b_0 + \mathbb{X}_{i2}b_1 + ... + \mathbb{X}_{ip}b_{p-1}))^2$$

czyli minimalizacji sumy kwadratów residów. Wynika stąd, że estymatory $\hat{\beta}$ uzyskane metodami ML i OLS są takie same.

Metoda największej wiarogodności (Maximum Likelihood)

Aby wyznaczyć estymator ML dla σ^2 musimy policzyć pochodną z logarytmu funkcji wiarogodności po z^2 i przyrównać do 0 (ekstremum) oraz sprawdzić, że rozwiązanie faktycznie maksymalizuje badaną funkcję. Estmator ML jest postaci:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - (\mathbb{X}_{i1}\hat{\beta}_0 + \mathbb{X}_{i2}\hat{\beta}_1 + \ldots + \mathbb{X}_{ip}\hat{\beta}_{p-1}))^2$$

Można pokazać, że jest on obciążony. Nieobciążony estymator (zwykle stosowany) ma postać:

$$s^2 = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - (\mathbb{X}_{i1}\hat{\beta}_0 + \mathbb{X}_{i2}\hat{\beta}_1 + \ldots + \mathbb{X}_{ip}\hat{\beta}_{p-1}))^2 = \frac{1}{n-p}\|Y - \mathbb{X}\hat{\beta}\|_2^2$$

$$E(s^2) = \sigma^2$$

Predykcja Y

Znając estymator $\hat{\beta}$ w łatwy sposób możemy wyznaczyć przewidywane wartości zmiennej objaśnianej dla wartości ze zbioru danych $Y = (Y_1, ..., Y_n)$:

$$\hat{Y} = \mathbb{X}\hat{\beta}$$

Predykcja Y

Znając estymator $\hat{\beta}$ w łatwy sposób możemy wyznaczyć przewidywane wartości zmiennej objaśnianej dla wartości ze zbioru danych $Y = (Y_1, ..., Y_n)$:

$$\hat{Y} = \mathbb{X}\hat{\beta}$$

W analogiczny sposób możemy dokonać predykcji dla dowolnej wartości $Y \in \mathbb{R}$.
Załóżmy że chcemy znaleźć przewidywaną wartość zmiennej objaśnianej $\tilde{Y}$ dla wektora wartości zmiennych objaśniających $\tilde{X} = (1, \tilde{X}_1, ..., \tilde{X}_{p-1})$ (np. nowy wiersz w macierzy planu) na podstawie skonstruowanego modelu empirycznego. Wówczas:

$$\hat{\tilde{Y}} = \tilde{X}\hat{\beta}$$

Table of Contents

Rozkład $\hat{\beta}$

Rozkład estymatora $\hat{\beta}$

Estymator otrzymany metodą najmniejszych kwadratów ma następujacy rozkład:

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

- normalność wynika z faktu, że $\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y$ jest liniowym przekształceniem wektora $Y \sim N(\mathbb{X}\beta, \sigma^2\mathbb{I})$,
- wartość oczekiwana: $E(\hat{\beta}) = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'E(Y) = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{X}\beta = \beta$ (estymator jest nieobciążony),
- macierz kowariancji: $Cov(\hat{\beta}) = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\sigma^2\mathbb{I}\,((\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')' = \sigma^2(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{X}((\mathbb{X}'\mathbb{X})^{-1})' = \sigma^2((\mathbb{X}'\mathbb{X})')^{-1} = \sigma^2(\mathbb{X}'\mathbb{X})^{-1}$

W dowodzie postaci macierzy kowariancji korzystaliśmy z własności macierzy: $(AB)' = B'A'$ oraz $(A^{-1})' = (A')^{-1}$

Macierz H

Wprowadźmy następujące oznaczenie na macierz

$$H = \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'$$

Macierz H jest ważnym obiektem w analizie modeli liniowych. Przy jej pomocy można w łatwy sposób wyrazić predykcję wektora odpowiedzi $\hat{Y}$ i wektor residuów e:

Predykcja $\hat{Y}$ i residua wyrażone przy pomocy macierzy H

$$\hat{Y} = \mathbb{X}\hat{\beta} = \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y = HY$$

$$e = Y - \hat{Y} = (\mathbb{I} - H)Y$$

Kolejną ważną własnością macierzy H jest:

H – macierz rzutu ortogonalnego na Lin(X)

Macierz H jest macierzą rzutu ortogonalnego na przestrzeń Lin(X) rozpiętą na kolumnach macierzy eksperymentu.

Macierz H

Aby to wykazać skorzystamy z charakteryzacji rzutów ortogonalnych z algebry liniowej:

dowolna macierz P jest macierzą rzutu ortogonalnego wtedy i tylko wtedy gdy macierz P jest symetryczna $(P = P')$ i idempotentna $(P^2 = P)$.

Macierz H

Aby to wykazać skorzystamy z charakteryzacji rzutów ortogonalnych z algebry liniowej:

dowolna macierz P jest macierzą rzutu ortogonalnego wtedy i tylko wtedy gdy macierz P jest symetryczna ($P = P'$) i idempotentna ($P^2 = P$).

Łatwo pokazać, że macierz H ma obie powyższe cechy:

- symetryczność: $H' = (\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')' = \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' = H$
- idempotentność:
  $H^2 = (\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}')(\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}') = \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}' = H$

Macierz H

Pozostaje wykazać że przestrzeń na którą rzutuje macierz H to $Lin(\mathbb{X})$.

Rozważmy dowolny wektor $v \in \mathbb{R}^n$ i rozłóżmy go na dwie składowe $v = v_1 + v_2$, w taki sposób że $v_1 \in Lin(\mathbb{X})$, a $v_2 \perp Lin(\mathbb{X})$. Ponieważ v_1 należy do przestrzeni $Lin(\mathbb{X})$ zatem możemy go wyrazić jako kombinację liniową kolumn z macierzy $\mathbb{X}$:

$$v_1 = \mathbb{X}w$$

gdzie $w \in \mathbb{R}^p$. Z drugiej strony ze względu na to, że wektor V_2 należy do przestrzeni ortogonalnej do $Lin(\mathbb{X})$, wynika to że dla dowolnej kolumny w macierzy $\mathbb{X}$, iloczyn skalarny pomiędzy kolumną a wektorem v_2 wynosi zero. Stąd:

$$\mathbb{X}'v_2 = 0$$

Wykorzystując otrzymane relacje otrzymujemy:

$$Hv = Hv_1 + Hv_2 = \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{X}w + \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'v_2 = \mathbb{X}w = v_1$$

Wobec założonej dowolnosci wektora v, uzyskujemy ostatecznie, że macierz H rzutuje na przestrzeń rozpiętą przez kolumny macierzy $\mathbb{X}$.

Rozkłady $\hat{Y}$ oraz e

Łatwo pokazać (przez analogiczne przekształcenia), że macierz $\mathbb{I} - H$ jest macierzą rzutu ortogonalnego na przestrzeń ortogonalną do $Lin(\mathbb{X})$. W konsekwencji otrzymujemy, że predykcja $\hat{Y}$ oraz wektor residuów e dane wzorami:

$$\hat{Y} = HY; \qquad e = (\mathbb{I} - H)Y$$

są rozkładem wektora Y na dwie ortogonalne składowe: leżącą w przestrzeni $Lin(\mathbb{X})$ (wekt. $\hat{Y}$) i w przestrzeni ortogonalnej do niej (wekt. e).

Rozkłady $\hat{Y}$ oraz e

Łatwo pokazać (przez analogiczne przekształcenia), że macierz $\mathbb{I} - H$ jest macierzą rzutu ortogonalnego na przestrzeń ortogonalną do $Lin(\mathbb{X})$. W konsekwencji otrzymujemy, że predykcja $\hat{Y}$ oraz wektor residuów e dane wzorami:

$$\hat{Y} = HY; \qquad e = (\mathbb{I} - H)Y$$

są rozkładem wektora Y na dwie ortogonalne składowe: leżącą w przestrzeni $Lin(\mathbb{X})$ (wekt. $\hat{Y}$) i w przestrzeni ortogonalnej do niej (wekt. e).

Korzystając z powyższych wzorów oraz z faktu że $Y \sim N(\mathbb{X}\beta, \sigma^2\mathbb{I})$, w natychmiastowy sposób otrzymujemy rozkłady obu wektorów:

Rozkłady $\hat{Y}$ oraz e

Predykcja $\hat{Y}$ oraz wektor residuów e mają następujace rozkłady:

$$\hat{Y} \sim N(\mathbb{X}\beta, \sigma^2 H); \qquad e \sim N(0, \sigma^2(\mathbb{I} - H))$$

Na mocy powyższej zależności łatwo zauważyć, że elementy wektora e są skorelowane! (inaczej niż elementy wetora błędów losowych ϵ).

Rozkład estymatora s^2

Przypomnijmy, że estymator wariancji błędów losowych s^2 jest funkcją wektora residuów e i dany jest wzorem:

$$s^2 = \frac{1}{n-p}\|e\|_2^2$$

Można pokazać natępujacą własność estymatora s^2:

Rozkład estymatora s^2

Jeżeli wyrazimy estymator s^2 w następujacy sposób:

$$s^2 = \sigma^2 \frac{K}{n-p}$$

Wówczas K jest zmienną losową z rozkładu χ^2 z $n-p$ stopniami swobody. Wykorzystując zależność pomiędzy s^2 i e otrzymujemy:

$$K = \|\frac{e}{\sigma}\|_2^2 \sim \chi^2_{n-p}$$

dowód

Pokażemy, że $\|e/\sigma\|_2^2 \sim \chi^2_{n-p}$. W pierwszym kroku zauważmy, iż

$$e = (\mathbb{I} - H)Y = (\mathbb{I} - H)(\mathbb{X}\beta + \epsilon) = (\mathbb{I} - H)\epsilon$$

Niech $(\mathbb{I} - H) = Q\Lambda Q'$ będzie diagonalizacją macierzy.

Q – macierz wekt. własnych, $QQ' = Q'Q = \mathbb{I}$ (izometria),

Λ – macierz diagonalna o wartościach własnych λ_i macierzy $(\mathbb{I} - H)$ na diagonali.

Wykorzystując powyższy rozkład macierzy $(\mathbb{I} - H)$, otrzymujemy

$$\|e/\sigma\|_2^2 = \|(\mathbb{I}-H)\epsilon/\sigma\|_2^2 = \|Q(\Lambda Q'\epsilon/\sigma)\|_2^2 = (\Lambda Q'\epsilon/\sigma)'Q'Q(\Lambda Q'\epsilon/\sigma) = \|\Lambda Q'\epsilon/\sigma\|_2^2$$

W ostatnim kroku skorzystaliśmy z faktu że macierz Q jest izometrią (nie zmiena długości wektorów).

Wprowadźmy następujące oznaczenie na wektor $Z = Q'\epsilon/\sigma$. Zauważmy, że wektor Z pochodzi z wielowymiarowego standardowego rozkładu normalnego:

$$Z = (Q'/\sigma)\epsilon \sim N(0, (Q'/\sigma)\sigma^2\mathbb{I}(Q'/\sigma)') = N(0, Q'Q) = N(0, \mathbb{I})$$

Otrzymujemy zatem, że badane wyrażenie jest ważoną sumą kwadratów niezależnych zmiennych losowych ze standardowego rozkładu normalnego:

$$\|e/\sigma\|_2^2 = \|\Lambda Q'\epsilon/\sigma\|_2^2 = \|\Lambda Z\|_2^2 = \sum_{i=1}^{n} \lambda_i Z_i^2$$

Rozkład estymatora s^2

Pozostaje zbadać własności wartości własnych λ_i ($\|e/\sigma\|_2^2 = \sum_{i=1}^n \lambda_i Z_i^2$). Fakt, iż macierz $(\mathbb{I} - H)$ jest macierzą rzutu ortogonalnego implikuje że jej wartości własne λ_i muszą być równe 0 lub 1.

Niech P–rzut ort. Ponieważ $P^2 = P$ zatem dla dowolnego wektora własnego v_i i stowarzyszonej z nim wartości własnej λ_i mamy $\lambda_i v_i = Pv_i = (PP)v_i = P(Pv_i) = P(\lambda_i v_i) = \lambda_i(Pv_i) = \lambda_i^2 v_i$, stąd $\lambda_i^2 v_i - \lambda_i v_i = 0 \Rightarrow \lambda_i(\lambda_i - 1) = 0 \Rightarrow \lambda_i \in \{0, 1\}$

Korzystając z własności, że ślad z macierzy jest równy sumie jej wartości własnych, możemy wyznaczyć liczbę wartości własnych równych 1 w $(\mathbb{I} - H)$:

$$\sum_{i=1}^n \lambda_i = Tr(\mathbb{I} - H) = Tr(\mathbb{I}_{n\times n}) - Tr(H) = n - Tr(\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}') =$$

$$= n - Tr((\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'\mathbb{X}) = n - Tr(\mathbb{I}_{p\times p}) = n - p$$

gdzie między pierwszą a drugą linią skorzystaliśmy z cykliczności śladu macierzy $Tr(ABC) = Tr(BCA)$

Ostatecznie uzyskujemy że $n - p$ wartosci własnych jest równa 1 oraz p jest rowna 0. W konsekwencji:

$$\|e/\sigma\|_2^2 = \sum_{i=1}^n \lambda_i Z_i^2 = \sum_{\{i:\lambda_i=1\}} 1 * Z_i^2 + \sum_{\{i:\lambda_i=0\}} 0 * Z_i^2 = \sum_{\{i:\lambda_i=1\}} Z_i^2 \sim \chi^2_{n-p}$$

Table of Contents

Testowanie dotyczące wektora parametrów β

Poznaliśmy już własności estymatorów $\hat{\beta}$ i s^2:

$$\hat{\beta} \sim N(\beta, \sigma^2(\mathbb{X}'\mathbb{X})^{-1}); \quad s^2 = \frac{1}{n-p}\|e\|_2^2$$

Przy ich pomocy możemy testować istotność dowolnej kombinacji liniowej elemetów wektora parametrów β. Niech $c = (c_0, ..., c_{p-1})' \in \mathbb{R}^p$ i $d \in \mathbb{R}$. Wówczas:

$$H_0 : c'\beta - d = 0 \quad vs \quad H_1 : c'\beta - d \neq 0$$

Statystyka testowa dla powyższego zagadnienia ma postać:

$$T = \frac{c'\hat{\beta} - d}{s(c'\hat{\beta})} \text{ gdzie } s^2(c'\hat{\beta}) = s^2 c'(\mathbb{X}'\mathbb{X})^{-1}c$$

Statystyka T przy załozeniu prawdziwosci hipotezy H_0, ma rozkład studenta z $n-p$ stopniami swobody. W konsekwencji będziemy odrzucać H_0 na poziomie istotności α, jeżeli $|T| > t_c$ gdzie $t_c = t^*(1-\alpha/2, n-p)$ jest kwantylem rzędu $1-\alpha/2$ z rozkładu studenta z $n-p$ stopniami swobody.
p-wartość dla tego problemu:
$p = P(|z| > |T|) = 2(1 - F_z(|T|))$, gdzie $z \sim t(n-p)$ i F_z – dystrybuanta zmiennej los. z.

Ważne przykłady

Do czego nam się przydaje taki "egzotyczny" test?

Ważne przykłady

Do czego nam się przydaje taki "egzotyczny" test?

Przy jego pomocy unifikujemy kilka poznanych wcześniej testów:

Ważne przykłady

Do czego nam się przydaje taki "egzotyczny" test?

Przy jego pomocy unifikujemy kilka poznanych wcześniej testów:

Przykład 1 (Testowanie istotności parametru β_i)
Jeżeli $c = (0, ..0, 1, 0, ..., 0)'$, gdzie wartość 1 odpowiada parametrowi β_i oraz $d = 0$. Wówczas $c'\beta - d = \beta_i$ i zagadnienie testowe sprowadza się do postaci:

$$H_0 : \beta_i = 0 \quad vs \quad H_1 : \beta_i \neq 0$$

a statystyka testowa ma postać:

$$T_i = \frac{\hat{\beta}_i}{s(\hat{\beta}_i)} \;\; \text{gdzie} \;\; s^2(\hat{\beta}_i) = s^2(\mathbb{X}'\mathbb{X})^{-1}_{i+1 i+1}$$

Ważne przykłady

Przykład 2 (Testowanie dotyczące $E(\tilde{Y}_h)$)
Jeżeli chcemy przeprowadzić testowanie dotyczące wartości oczekiwanej ze zmiennej objaśnianej $E(\tilde{Y}_h) = \tilde{\mathbb{X}}_{h\cdot}\beta = \tilde{\mu}_h$ wystarczy, że za wektor c wstawimy ciąg wartości zmiennych objaśniających stowarzyszonych z $\tilde{Y}_h$, $c = (1, \tilde{X}_{h1}, ..., \tilde{X}_{hp-1})' = \tilde{\mathbb{X}}'_{h\cdot}$. Wówczas $c'\beta - d = \tilde{\mu}_h - d$ i zagadnienie testowe sprowadza się do postaci:

$$H_0 : \tilde{\mu}_h - d = 0 \quad vs \quad H_1 : \tilde{\mu}_h - d \neq 0$$

a statystyka testowa ma postać:

$$T = \frac{\hat{\tilde{\mu}}_h - d}{s(\hat{\tilde{\mu}}_h)} \;\; \text{gdzie} \;\; \hat{\tilde{\mu}}_h = \tilde{\mathbb{X}}_{h\cdot}\hat{\beta} \;\; \text{i} \;\; s^2(\hat{\tilde{\mu}}_h) = s^2\tilde{\mathbb{X}}_{h\cdot}(\mathbb{X}'\mathbb{X})^{-1}\tilde{\mathbb{X}}'_{h\cdot}$$

Uzasadnienie rozkładu statystyki T

Dla regresji liniowej prostej fakt opisujący rozkład statystyki T został podany bez uzasadnienia. Obecnie pokażemy go, w ścisły sposób.

Uzasadnienie rozkładu statystyki T

Dla regresji liniowej prostej fakt opisujący rozkład statystyki T został podany bez uzasadnienia. Obecnie pokażemy go, w ścisły sposób.

Z definicji zmienną losową z rozkładu studenta o k stopniach swobody można wyrazić jako iloraz niezależnych zmiennych losowych: $\tilde{T} = \frac{\tilde{Z}}{\sqrt{\tilde{K}/k}}$ ma rozkład stud. z k st. swob. gdy $\tilde{Z} \sim N(0,1)$; $\tilde{K} \sim \chi^2_k$; $\tilde{Z} \perp \tilde{K}$

Statystykę T możemy wyrazić w następujący sposób:

$$T = \frac{c'\hat{\beta} - d}{\sqrt{s^2 c'(\mathbb{X}'\mathbb{X})^{-1}c}} = \frac{(c'\hat{\beta} - d)/\left(\sigma\sqrt{c'(\mathbb{X}'\mathbb{X})^{-1}c}\right)}{\sqrt{s^2/\sigma^2}}$$

Zauważmy, że przy H_0 licznik i mianownik mają odpowiednie rozkłady:

$$c'\hat{\beta} \sim N(d, \sigma^2 c'(\mathbb{X}'\mathbb{X})^{-1}c) \;\Rightarrow\; (c'\hat{\beta} - d)/\left(\sigma\sqrt{c'(\mathbb{X}'\mathbb{X})^{-1}c}\right) \sim N(0,1)$$

$$s^2/\sigma^2 = \frac{1}{n-p}\|\frac{e}{\sigma}\|_2^2 \;\text{ oraz }\; \|\frac{e}{\sigma}\|_2^2 = K \sim \chi^2_{n-p}$$

Uzasadnienie rozkładu statystyki T

Dla regresji liniowej prostej fakt opisujący rozkład statystyki T został podany bez uzasadnienia. Obecnie pokażemy go, w ścisły sposób.

Z definicji zmienną losową z rozkładu studenta o k stopniach swobody można wyrazić jako iloraz niezależnych zmiennych losowych: $\tilde{T} = \frac{\tilde{Z}}{\sqrt{\tilde{K}/k}}$ ma rozkład stud. z k st. swob. gdy $\tilde{Z} \sim N(0,1)$; $\tilde{K} \sim \chi^2_k$; $\tilde{Z} \perp \tilde{K}$

Statystykę T możemy wyrazić w następujący sposób:

$$T = \frac{c'\hat{\beta} - d}{\sqrt{s^2 c'(\mathbb{X}'\mathbb{X})^{-1}c}} = \frac{(c'\hat{\beta} - d)/\left(\sigma\sqrt{c'(\mathbb{X}'\mathbb{X})^{-1}c}\right)}{\sqrt{s^2/\sigma^2}}$$

Zauważmy, że przy H_0 licznik i mianownik mają odpowiednie rozkłady:

$$c'\hat{\beta} \sim N(d, \sigma^2 c'(\mathbb{X}'\mathbb{X})^{-1}c) \quad \Rightarrow \quad (c'\hat{\beta} - d)/\left(\sigma\sqrt{c'(\mathbb{X}'\mathbb{X})^{-1}c}\right) \sim N(0,1)$$

$$s^2/\sigma^2 = \frac{1}{n-p}\|\frac{e}{\sigma}\|_2^2 \ \text{ oraz } \ \|\frac{e}{\sigma}\|_2^2 = K \sim \chi^2_{n-p}$$

Pozostaje zbadać niezależność. W tym celu wystarczy pokazać niezależność $\hat{\beta}$ i e:

$$Cov(e, \hat{\beta}) = Cov((\mathbb{I} - H)Y, (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'Y) = (\mathbb{I} - H)Cov(Y, Y)\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1} =$$

$$(\mathbb{I} - H)(\sigma^2\mathbb{I})\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1} = \sigma^2(\mathbb{X} - H\mathbb{X})(\mathbb{X}'\mathbb{X})^{-1} = 0$$

Wektory $\hat{\beta}$ i e są nieskorelowane i mają normalne rozkłady (jako przekszt. wekt. Y), stąd są również niezależne. Dodatkowo, własność ta jest dziedziczona na licznik i mianownik jako przekształcenia odpowiednich wektorów (licznik $= f_1(\hat{\beta})$; mianownik $= f_2(e)$).

Przedziały ufności

Na podstawie statystyki T możemy również skonstruować przedział ufności o współczynniku ufności $1-\alpha$ dla kombinacji liniowej $c'\beta$:

$$c'\hat{\beta} \pm t_c s(c'\hat{\beta})$$

gdzie $s^2(\hat{\beta}_i) = s^2 c'(\mathbb{X}'\mathbb{X})^{-1}c$ oraz $t_c = t^*(1-\frac{\alpha}{2}, n-p)$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ z rozkładu studenta z $n-p$ stopniami swobody.

Przedziały ufności

Na podstawie statystyki T możemy również skonstruować przedział ufności o współczynniku ufności $1-\alpha$ dla kombinacji liniowej $c'\beta$:

$$c'\hat{\beta} \pm t_c s(c'\hat{\beta})$$

gdzie $s^2(\hat{\beta}_i) = s^2 c'(\mathbb{X}'\mathbb{X})^{-1}c$ oraz $t_c = t^*(1-\frac{\alpha}{2}, n-p)$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ z rozkładu studenta z $n-p$ stopniami swobody.

Przykład 1 (przedział ufności dla parametru β_i; $c = (0, ..0, 1, 0, ..., 0)'$):

$$\hat{\beta}_i \pm t_c s(\hat{\beta}_i) \text{ gdzie } s^2(\hat{\beta}_i) = s^2(\mathbb{X}'\mathbb{X})^{-1}_{i+1i+1}$$

Przykład 2 (przedział ufności dla $\tilde{\mu}_h$; $c = (1, \tilde{X}_{h1}, ..., \tilde{X}_{hp-1})' = \tilde{\mathbb{X}}'_{h\cdot}$):

$$\hat{\tilde{\mu}}_h \pm t_c s(\hat{\tilde{\mu}}_h) \;\; \text{gdzie} \;\; \hat{\tilde{\mu}}_h = \tilde{\mathbb{X}}_{h\cdot}\hat{\beta} \;\; \text{i} \;\; s^2(\hat{\tilde{\mu}}_h) = s^2\tilde{\mathbb{X}}_{h\cdot}(\mathbb{X}'\mathbb{X})^{-1}\tilde{\mathbb{X}}'_{h\cdot}$$

Przedział predykcyjny dla nowej obserwacji $\tilde{Y}_h$

Wykonując podobne rozumowanie możemy wyznaczyć przedział predykcyjny na poziomie $1-\alpha$ dla nowej obserwacji $\tilde{Y}_h$. Jest on postaci:

$$\hat{\tilde{\mu}}_h \pm t_c s(pred) \quad \text{gdzie} \quad \hat{\tilde{\mu}}_h = \tilde{\mathbb{X}}_{h\cdot}\hat{\beta} \quad \text{i} \quad s^2(pred) = s^2(1 + \tilde{\mathbb{X}}_{h\cdot}(\mathbb{X}'\mathbb{X})^{-1}\tilde{\mathbb{X}}'_{h\cdot})$$

gdzie $t_c = t^*(1-\frac{\alpha}{2}, n-p)$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ z rozkładu studenta z $n-p$ stopniami swobody.

Przedział predykcyjny dla nowej obserwacji $\tilde{Y}_h$

Wykonując podobne rozumowanie możemy wyznaczyć przedział predykcyjny na poziomie $1-\alpha$ dla nowej obserwacji $\tilde{Y}_h$. Jest on postaci:

$$\hat{\hat{\mu}}_h \pm t_c s(pred) \text{ gdzie } \hat{\hat{\mu}}_h = \tilde{\mathbb{X}}_{h.}\hat{\beta} \text{ i } s^2(pred) = s^2(1+\tilde{\mathbb{X}}_{h.}(\mathbb{X}'\mathbb{X})^{-1}\tilde{\mathbb{X}}_{h.}')$$

gdzie $t_c = t^*(1-\frac{\alpha}{2}, n-p)$ jest kwantylem rzędu $1-\frac{\alpha}{2}$ z rozkładu studenta z $n-p$ stopniami swobody.

szkic dowodu

Chcemy pokazać że:

$$P(\hat{\hat{\mu}}_h - t_c s(pred) < \tilde{Y}_h < \hat{\hat{\mu}}_h + t_c s(pred)) = 1-\alpha$$

Łatwo pokazać że:

$$P(\hat{\hat{\mu}}_h - t_c s(pred) < \tilde{Y}_h < \hat{\hat{\mu}}_h + t_c s(pred)) = P\left(\left|\frac{\tilde{Y}_h - \hat{\hat{\mu}}_h}{s(pred)}\right| < t_c\right)$$

Wiemy, że zmienna losowa $\tilde{Y}_h$ ma rozkład $N(\tilde{\mathbb{X}}_{h.}\beta, \sigma^2)$ i jako nowa obserwacja jest niezależna od $\hat{\hat{\mu}}_h \sim N(\tilde{\mathbb{X}}_{h.}\beta, \sigma^2\tilde{\mathbb{X}}_{h.}(\mathbb{X}'\mathbb{X})^{-1}\tilde{\mathbb{X}}_{h.}')$. Stąd:

$$E(\tilde{Y}_h - \hat{\hat{\mu}}_h) = 0$$

$$\sigma^2(pred) = var(\tilde{Y}_h - \hat{\hat{\mu}}_h) = var(\tilde{Y}_h) + var(\hat{\hat{\mu}}_h) = \sigma^2(1+\tilde{\mathbb{X}}_{h.}(\mathbb{X}'\mathbb{X})^{-1}\tilde{\mathbb{X}}_{h.}')$$

Łatwo zauważyć, że $s^2(pred)$ jest estymatorem $\sigma^2(pred)$. Ponadto jest on zmienną losową niezależną od $\tilde{Y}_h - \hat{\hat{\mu}}_h$ ($\hat{\hat{\mu}}_h$ jest funkcją $\hat{\beta}$ a $\tilde{Y}_h$ jest nową nieazleżną obserwacją). Zatem:

$$\frac{\tilde{Y}_h - \hat{\hat{\mu}}_h}{s(pred)} = \frac{(\tilde{Y}_h - \hat{\hat{\mu}}_h)\Big/\sqrt{\sigma^2(1+\tilde{\mathbb{X}}_{h.}(\mathbb{X}'\mathbb{X})^{-1}\tilde{\mathbb{X}}_{h.}')}}{\sqrt{\frac{s^2}{\sigma^2}}} \qquad \begin{array}{c} \sim N(0,1) \\ \sim \sqrt{\chi^2_{n-p}/(n-p)} \end{array}$$

Table of Contents

Analiza wariancji

Analiza wariancji dla regresji liniowej jest metodą uporządkowania zmienności w wektorze odpowiedzi $Y = (Y_1, ..., Y_n)'$. Są dwa źródła zmienności:

- zmienność wynikająca z błędów losowych $\epsilon_i \sim N(0, \sigma^2)$ powiązana z wielkością parametru σ^2,
- zmienność, której źródłem są różnice w wartościach oczekiwanych zmiennych objaśnianych $E(Y_i) = \mu_i = \beta_0 + X_{i1}\beta_1 + ... + X_{ip-1}\beta_{p-1}$.

Analiza wariancji

Do analizy zmienności w wektorze odpowiedzi Y wykorzystywaliśmy w regresji liniowej prostej, poniższe statystyki (dla wartości $p = 2$):

Source	SS	df	MS
Model	$SSM = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$	$dfM = p - 1$	$MSM = SSM/dfM$
Error	$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$	$dfE = n - p$	$MSE = SSE/dfE = s^2$
Total	$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$	$dfT = n - 1$	$MST = SST/dfT$
Relations	SST= SSM+SSE	dfT=dfM+dfE	MST$\neq$MSM+MSE

Całość rozumowania w przypadku regresji liniowej wielorakiej jest taka sama z tą różnicą, że w tym przypadku liczba stopni swobody statystyk SSM i SSE zależy od liczby kolumn p.

Liczba stopni swobody

Uzasadnimy obecnie skąd bierze się kolumna z liczbą stopni swobody. Jest ona równa śladowi macierzy, na którą wykonywany jest rzut ortogonalny wektora Y związany z odpowiednimi sumami kwadratów SSM, SSE oraz SST.

Liczba stopni swobody

Uzasadnimy obecnie skąd bierze się kolumna z liczbą stopni swobody. Jest ona równa śladowi macierzy, na którą wykonywany jest rzut ortogonalny wektora Y związany z odpowiednimi sumami kwadratów SSM, SSE oraz SST.

Zacznijmy od statystyki SSE, gdyż dla niej do pewnego stopnia analiza już została wykonana. Zauważmy, że:

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \|Y - \hat{Y}\|_2^2 = \|(\mathbb{I} - H)Y\|_2^2 = \|e\|_2^2$$

Z wcześniejszej analizy wiemy dodatkowo, że ślad macierzy rzutu ortogonalnego $Tr(\mathbb{I} - H) = n - p$.

Liczba stopni swobody

W analogiczny sposób możemy potraktować statystykę SST. Wystarczy zauważyć, że jest ona tożsama ze statystyką SSE dla szczególnej macierzy rzutu $\tilde{H}$:

$$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \|(\mathbb{I} - \tilde{H})Y\|_2^2 = \|\tilde{e}\|_2^2$$

gdzie $\tilde{H}$ jest rzutem ortogonalnym na przestrzeń rozpiętą na kolumnie jedynek związanej z interceptem $Lin(\mathbb{1} = (1, .., 1)')$ (brak zmiennych objaśniających).

$$\tilde{H} = \mathbb{1}(\mathbb{1}'\mathbb{1})^{-1}\mathbb{1}' = \frac{1}{n}\mathbb{1}\mathbb{1}' = \begin{pmatrix} 1/n & \cdots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \cdots & 1/n \end{pmatrix}; \quad \tilde{H}Y = \begin{pmatrix} \bar{Y} \\ \vdots \\ \bar{Y} \end{pmatrix}$$

Łatwo pokazać, że $Tr(\tilde{H}) = 1$, dlatego $dfT = Tr(\mathbb{I} - \tilde{H}) = n - 1$.

Liczba stopni swobody

W przypadku statystyki SSM wykorzystamy wyniki dla poprzednich punktów

$$SSM = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \|(H - \tilde{H})Y\|_2^2$$

Stąd $dfM = Tr(H - \tilde{H}) = Tr(H) - Tr(\tilde{H}) = p - 1$.
Wypada wykazać że $H - \tilde{H}$ jest macierzą rzutu ortogonalnego. Symetryczność jest oczywista, także pozostaje wykazać idempotentność. Jednakże w pierwszej kolejności należy zauważyć że:

$$H\tilde{H} = \tilde{H} \text{ oraz } \tilde{H}H = \tilde{H}$$

Powyższe równości wynikają bezpośrednio z faktu, iż przestrzeń $Lin(\mathbb{1})$ (na którą rzutuje $\tilde{H}$) jest podprzestrzenią $Lin(\mathbb{X})$ (na którą rzutuje H). Wykorzystując powyższe fakty uzyskujemy idempotentność macierzy $H - \tilde{H}$:

$$(H - \tilde{H})^2 = H^2 - H\tilde{H} - \tilde{H}H + \tilde{H}^2 = H - \tilde{H}$$

Test F

Na podstawie statystyk MSM i MSE skonstruowany jest tzw. test F testujący hipotezę o istotności wpływu wszystkich regresorów (równocześnie):

$$H_0: \ \ \beta_1 = ... = \beta_{p-1} = 0 \qquad vs \qquad H_1: \ \ \beta_i \neq 0 \text{ przynajmniej dla jednego i}$$

Odrzucamy hipotezę zerową, gdy $F = MSM/MSE > F_c$, gdzie $F_c = F^*(1-\alpha, dfM, dfE)$ jest kwantylem rzędu $1-\alpha$ z rozkładu Fishera–Snedecora z dfM i dfE stopniami swobody.

Test F

Na podstawie statystyk *MSM* i *MSE* skonstruowany jest tzw. test F testujący hipotezę o istotności wpływu wszystkich regresorów (równocześnie):

$$H_0: \ \beta_1 = ... = \beta_{p-1} = 0 \qquad vs \qquad H_1: \ \beta_i \neq 0 \text{ przynajmniej dla jednego i}$$

Odrzucamy hipotezę zerową, gdy $F = MSM/MSE > F_c$, gdzie $F_c = F^*(1-\alpha, dfM, dfE)$ jest kwantylem rzędu $1-\alpha$ z rozkładu Fishera–Snedecora z *dfM* i *dfE* stopniami swobody.

Zwykle, wnioskowanie dokonywane jest na podstawie p-wartości: $p = P(z > F) = 1 - \tilde{F}_z(F)$, gdzie $z \sim F(dfM, dfE)$ a $\tilde{F}_z$ jest dystrybuantą zm. z.

Test F

Na podstawie statystyk MSM i MSE skonstruowany jest tzw. test F testujący hipotezę o istotności wpływu wszystkich regresorów (równocześnie):

$$H_0: \ \beta_1 = ... = \beta_{p-1} = 0 \qquad vs \qquad H_1: \ \beta_i \neq 0 \text{ przynajmniej dla jednego i}$$

Odrzucamy hipotezę zerową, gdy $F = MSM/MSE > F_c$, gdzie $F_c = F^*(1-\alpha, dfM, dfE)$ jest kwantylem rzędu $1-\alpha$ z rozkładu Fishera–Snedecora z dfM i dfE stopniami swobody.

Zwykle, wnioskowanie dokonywane jest na podstawie p-wartości: $p = P(z > F) = 1 - \tilde{F}_z(F)$, gdzie $z \sim F(dfM, dfE)$ a $\tilde{F}_z$ jest dystrybuantą zm. z.

Na hipotezy w tescie F można również spojrzeć jak na porównanie dwóch modeli:

$$H_0: \text{ dane pochodzą z modelu } Y_i = \beta_0 + \epsilon_i \text{ (tzw. model zredukowany)}$$

$$H_1: \text{ dane pochodzą z modelu } Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_{p-1} X_{ip-1} \epsilon_i \text{ (tzw. model pełny)}$$

Tabela analizy wariancji (ANOVA Table)

Dane potrzebne do wykonania testu F zwykle przedstawiane są w tzw. tabeli analizy wariancji, o następującej postaci:

	df	SS	MS	F	p-value
Model	dfM	SSM	MSM	$\mathrm{F} = \mathrm{MSM}/\mathrm{MSE}$	$p = P(z > F)$
Error	dfE	SSE	MSE		

Uwagi o teście F:

- gdy zachodzi hipoteza alternatywna statystyka F pochodzi z niecentralnego rozkładu Fishera–Snedecora,
- umożliwia to wyznaczenie funkcji mocy testu,
- gdy $p = 2$ (mamy jedeną zm. objaśniającą) test F i test oparty na statystyce T dla parametru β_1 są równoważne.

Table of Contents

Współczynnik determinacji R²

Współczynnik determinacji R^2 jest miarą jakości dopasowania modelu. Mówi on o tym, jaką część całkowitej zmiennosci w wektorze Y (SST) stanowi zmienność wyjaśniona przez model (SSM).

$$R^2 = SSM/SST = 1 - SSE/SST$$

Przyjmuje on wartości od 0 do 1 (czasami wyrażany jest w skali procentowej).

Współczynnik determinacji R^2

Współczynnik determinacji R^2 jest miarą jakości dopasowania modelu. Mówi on o tym, jaką część całkowitej zmiennosci w wektorze Y (SST) stanowi zmienność wyjaśniona przez model (SSM).

$$R^2 = SSM/SST = 1 - SSE/SST$$

Przyjmuje on wartości od 0 do 1 (czasami wyrażany jest w skali procentowej).

- Przy użyciu statystyki R^2 możemy porównywać modele o tej samej liczbie regresorów (również w regresji wielorakiej),
- czasami gdy porównujemy modele o różnej liczbie regresorów, zamiast R^2 używa się tzw. modyfikowanego współczynnika determinacji:

$$\tilde{R}^2 = 1 - MSE/MST$$

istnieją jednak lepsze metody.