Modele liniowe

Michał Kos

Uniwersytet Wrocławski

Plan wykładu

1. Wstęp
2. Regresja liniowa prosta (Simple Linear Regression)
3. Metody estymacji
4. Teoretyczne własności estymatorów
5. Estymacja $E(Y_h) = \mu_h$ i predykcja Y_h
6. Pasmo ufności dla prostej regresji
7. Podsumowanie

Table of Contents

Ważne uwagi:

1. Modele liniowe to <u>ważny</u> przedmiot.
   - metody analizy danych i ich uogólnienia są powszechnie stosowane w praktycznych zagadnieniach,
   - przynajmniej 4 kontynuacje przedmiotu w IM,

Ważne uwagi:

1. Modele liniowe to <u>ważny</u> przedmiot.
    - metody analizy danych i ich uogólnienia są powszechnie stosowane w praktycznych zagadnieniach,
    - przynajmniej 4 kontynuacje przedmiotu w IM,
2. <u>Biegła</u> znajomość środowiska R.
    - jeden z podstawowych języków używanych przez statystyków na świecie.
    - rozbudowana baza pakietów statystycznych,
    - Podstawowe narzędzie na wykładzie i na kontynuacjach,

Modele liniowe a test studenta

O modelach liniowych można myśleć jako o uogólnieniu testu studenta.
Test studenta:

- Populacji podzielona ze względu na przynależność do podgrup:
  $Y_1, ..., Y_{n_1}$ – i.i.d. $N(\mu_1, \sigma^2)$, n_1 obserwacji cechy Y w grupie 1,
  $Y_{n_1+1}, ..., Y_{n_1+n_2}$ – i.i.d. $N(\mu_2, \sigma^2)$, n_2 obserwacji cechy Y w grupie 2,

Modele liniowe a test studenta

O modelach liniowych można myśleć jako o uogólnieniu testu studenta.
Test studenta:

- Populacji podzielona ze względu na przynależność do podgrup:
  $Y_1, ..., Y_{n_1}$ – i.i.d. $N(\mu_1, \sigma^2)$, n_1 obserwacji cechy Y w grupie 1,
  $Y_{n_1+1}, ..., Y_{n_1+n_2}$ – i.i.d. $N(\mu_2, \sigma^2)$, n_2 obserwacji cechy Y w grupie 2,
- porównania średnich w dwóch próbach $H_0 : \mu_1 = \mu_2$,
  przynależność do grupy jest istotna = dodatkowa zmienna X opisująca przynależnosć do grupy ($X_i \in \{1, 2\}$), wpływa na zmienną Y,
  przynależność do grupy jest nieistotna = zmienna Y jest niezależna od podziału na grupy (zmienna X) = wszystkie Y-ki są z tego samego rozkładu,

Modele liniowe a test studenta

O modelach liniowych można myśleć jako o uogólnieniu testu studenta.
Test studenta:

- Populacji podzielona ze względu na przynależność do podgrup:
  $Y_1, ..., Y_{n_1}$ – i.i.d. $N(\mu_1, \sigma^2)$, n_1 obserwacji cechy Y w grupie 1,
  $Y_{n_1+1}, ..., Y_{n_1+n_2}$ – i.i.d. $N(\mu_2, \sigma^2)$, n_2 obserwacji cechy Y w grupie 2,
- porównania średnich w dwóch próbach $H_0 : \mu_1 = \mu_2$,
  przynależność do grupy jest istotna $=$ dodatkowa zmienna X opisująca przynależnosć do grupy ($X_i \in \{1, 2\}$), wpływa na zmienną Y,
  przynależność do grupy jest nieistotna $=$ zmienna Y jest niezależna od podziału na grupy (zmienna X) $=$ wszystkie Y-ki są z tego samego rozkładu,
- Przykład. Badanie wpływu leku na anemię:
  $Y_1, ..., Y_{n_1}$ – poziom hemoglobiny u przyjmujących lek
  $Y_{n_1+1}, ..., Y_{n_1+n_2}$ – poziom hemoglobiny u przyjmujących placebo
  Czy grupa kontrolna ma średnio niższy poziom hemoglobiny?

Modele liniowe a test studenta

W ogólności, w teście studenta, każda obserwacja $Y_1, ..., Y_{n_1+n_2}$ ma postać:

$$Y_i = f(X_i) + \epsilon_i$$

gdzie:

$$f(X_i) = \begin{cases} \mu_1 & \text{gdy} \quad X_i = 1 \\ \mu_2 & \text{gdy} \quad X_i = 2 \end{cases}$$

oraz $\epsilon_1, ..., \epsilon_{n_1+n_2}$ są i.i.d. $N(0, \sigma^2)$.

Modele liniowe a test studenta

W ogólności, w teście studenta, każda obserwacja $Y_1, ..., Y_{n_1+n_2}$ ma postać:

$$Y_i = f(X_i) + \epsilon_i$$

gdzie:

$$f(X_i) = \begin{cases} \mu_1 & \text{gdy} \quad X_i = 1 \\ \mu_2 & \text{gdy} \quad X_i = 2 \end{cases}$$

oraz $\epsilon_1, ..., \epsilon_{n_1+n_2}$ są i.i.d. $N(0, \sigma^2)$.
Wówczas np. dla $X_i = 1$:

$$Y_i = f(X_i) + \epsilon_i = \mu_1 + \epsilon_i \sim N(\mu_1, \sigma^2)$$

Modele liniowe a test studenta

W ogólności, w teście studenta, każda obserwacja $Y_1, ..., Y_{n_1+n_2}$ ma postać:

$$Y_i = f(X_i) + \epsilon_i$$

gdzie:

$$f(X_i) = \begin{cases} \mu_1 & \text{gdy} \quad X_i = 1 \\ \mu_2 & \text{gdy} \quad X_i = 2 \end{cases}$$

oraz $\epsilon_1, ..., \epsilon_{n_1+n_2}$ są i.i.d. $N(0, \sigma^2)$.
Wówczas np. dla $X_i = 1$:

$$Y_i = f(X_i) + \epsilon_i = \mu_1 + \epsilon_i \sim N(\mu_1, \sigma^2)$$

Badamy zaś czy funkcja $f(\cdot)$ jest stała ($\mu_1 = \mu_2$).

Modele liniowe a test studenta

W przypadku <u>testu studenta zmienna X</u> przyjmuje wyłącznie <u>dwa stany</u>.

Modele liniowe a test studenta

W przypadku testu studenta zmienna X przyjmuje wyłącznie dwa stany.

W Modelach liniowych obserwacje $Y_1, ..., Y_n$ mają tą samą postać $Y_i = f(X_i) + \epsilon_i$. Modele liniowe idą jednak o krok dalej, w tym sensie, że zmienna X może przyjmować wiele stanów, a w szczególnosci może być zmienną ciągłą np. waga, wzrost itp.

Modele liniowe a test studenta

W przypadku testu studenta zmienna X przyjmuje wyłącznie dwa stany.

W Modelach liniowych obserwacje $Y_1, ..., Y_n$ mają tą samą postać $Y_i = f(X_i) + \epsilon_i$. Modele liniowe idą jednak o krok dalej, w tym sensie, że zmienna X może przyjmować wiele stanów, a w szczególnosci może być zmienną ciągłą np. waga, wzrost itp.

Ponadto umożliwiają badanie własnosci zmiennej Y w zależnosci od kilku zmiennych $X^{(1)}, X^{(2)}..., X^{(p)}$:

$$Y_i = f(X_i^{(1)}, X_i^{(2)}, ..., X_i^{(p)}) + \epsilon_i$$

gdzie $X_i^{(j)}$ to wartość j-tej cechy dla i-tej obserwacji.

Modele liniowe a test studenta

W przypadku testu studenta zmienna X przyjmuje wyłącznie dwa stany.

W Modelach liniowych obserwacje $Y_1, ..., Y_n$ mają tą samą postać $Y_i = f(X_i) + \epsilon_i$. Modele liniowe idą jednak o krok dalej, w tym sensie, że zmienna X może przyjmować wiele stanów, a w szczególnosci może być zmienną ciągłą np. waga, wzrost itp.

Ponadto umożliwiają badanie własnosci zmiennej Y w zależnosci od kilku zmiennych $X^{(1)}, X^{(2)} ..., X^{(p)}$:

$$Y_i = f(X_i^{(1)}, X_i^{(2)}, ..., X_i^{(p)}) + \epsilon_i$$

gdzie $X_i^{(j)}$ to wartość j-tej cechy dla i-tej obserwacji.

W dużej ogólności można powiedzieć, że przedmiotem badania Modeli Liniowych są własności funkcji $f(\cdot)$, o której zakładamy że jest funkcją liniową: $f(X_i^{(1)}, ..., X_i^{(p)}) = b_0 + b_1 X_i^{(1)} + ... + b_p X_i^{(p)}$.

Modele liniowe – przykłady

Przykład 1

Przyjmijmy, że badamy ciśnienie pacjenta (zmienna Y).
Wiemy, że za ludzkie ciśnienie odpowiada w pewnym stopniu pula genów pacjenta (geny $X^{(1)}, X^{(2)}..., X^{(p)}$). Jeżeli znamy zależność pomiędzy ciśnieniem pacjenta i pulą genetyczną, to jesteśmy w stanie przewidzieć ryzyko wystąpienia nadciśnienia u pacjenta w wieku 40 lat zaraz po jego narodzinach (już wtedy pula genetyczna pacjenta jest ustalona).

Modele liniowe – przykłady

Przykład 1

Przyjmijmy, że badamy ciśnienie pacjenta (zmienna Y).
Wiemy, że za ludzkie ciśnienie odpowiada w pewnym stopniu pula genów pacjenta (geny $X^{(1)}, X^{(2)} \dots, X^{(p)}$). Jeżeli znamy zależność pomiędzy ciśnieniem pacjenta i pulą genetyczną, to jesteśmy w stanie przewidzieć ryzyko wystąpienia nadciśnienia u pacjenta w wieku 40 lat zaraz po jego narodzinach (już wtedy pula genetyczna pacjenta jest ustalona).

Przykład 2

Podobny mechanizm wykorzystywany jest w ekonomii. Banki modelują ryzyko niespłacenia przez klienta kredytu (Y) na podstawie danych o kliencie w chwili przyznawania kredytu np. wysokość zarobków, wykształcenie, oszczędności itp. $(X^{(1)}, X^{(2)} \dots, X^{(p)})$.

Table of Contents

Pochylenie krzywej wieży w Pizie

Najprostszym modelem liniowym jest tzw. regresja liniowa prosta, w której mamy jedną zmienną odpowiedzi Y (zm. wynikową, objaśnianą, zależną) i jedną zmienną objaśniającą X (zm. niezależną, regresor, predyktor).

Pochylenie krzywej wieży w Pizie

Najprostszym modelem liniowym jest tzw. regresja liniowa prosta, w której mamy jedną zmienną odpowiedzi Y (zm. wynikową, objaśnianą, zależną) i jedną zmienną objaśniającą X (zm. niezależną, regresor, predyktor).

Przykład 3

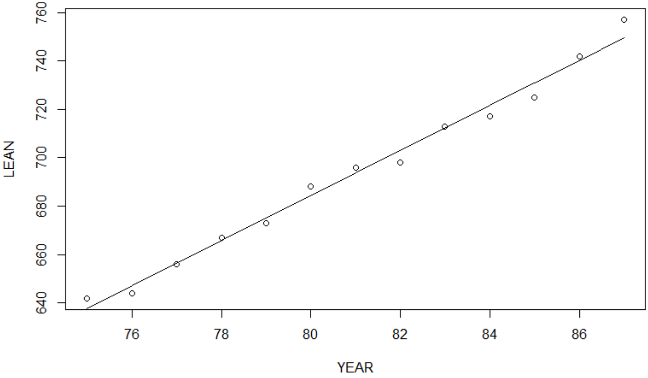
Pochylenie krzywej wieży:

- Pochylenie (Y) – zm. objaśniana,
- Czas (X) – zm. objaśniająca,
- narysowanie zależności na podstawie dotychczasowych danych (R: plot),
- dopasowanie prostej (R: lm),
- predykcja na podstawie dopasowanej prostej (R: predict).

Rysunek: By Softeis - Praca własna, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=25812

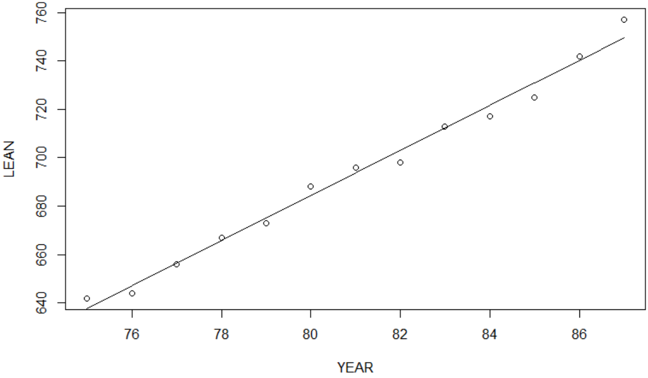
Pochylenie krzywej wieży w Pizie

Dane: piza.txt, Kod R: piza.r



Pochylenie krzywej wieży w Pizie

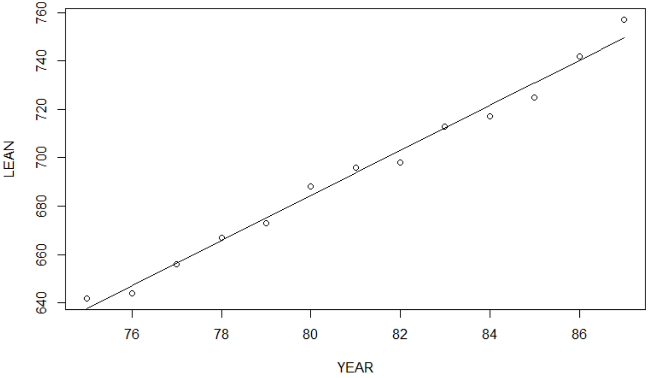
Dane: piza.txt, Kod R: piza.r



Jakie jest przewidywane pochylenie dla roku 2000 (pkt. 100)?

Pochylenie krzywej wieży w Pizie

Dane: piza.txt, Kod R: piza.r



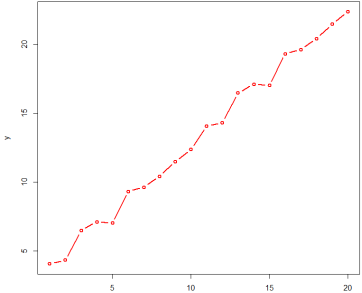
Jakie jest przewidywane pochylenie dla roku 2000 (pkt. 100)? ≈ 870.7

Dane w regresji liniowej prostej

Zakładamy, że dysponujemy zbiorem par $(X_i, Y_i)_{i=1}^n$ oraz, że związek między nimi jest w przybliżeniu liniowy.

$Y_1, ..., Y_n$ – wartości zmiennej zależnej,
$X_1, ..., X_n$ – wartości zmiennej niezależnej.
Od tej pory przez $Y = (Y_1, ..., Y_n)'$ oraz $X = (X_1, ..., X_n)'$ będziemy oznaczać (odpowiednio) wektor odpowiedzi (Y) oraz wektor wartości zmiennej niezależnej (X).

X	Y
1	4.0
2	4.3
3	6.5
⋮	⋮
19	21.5
20	22.4



Teoretyczny model regresji liniowej prostej

Teoretyczny model regresji liniowej prostej

Zakładamy, że związek między zmiennymi zależnymi i odpowiadającymi im wartościami zmiennych niezależnych jest postaci:

$$Y_i = f(X_i) + \epsilon_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad i = 1, ..., n$$

gdzie:
β_0 – Intercept (wyraz wolny) – deterministyczny parametr regresji,
β_1 – Slope (współczynnik kierunkowy) – deterministyczny parametr regrsji,
ϵ_i – błąd zwiazany z i-tym pomiarem, zmienna losowa $N(0, \sigma^2)$
Dodatkowo zakładamy, że $\epsilon_1, ..., \epsilon_n$ są niezależnymi zmiennymi losowymi.

Uwaga

Na tym wykładzie zakładamy, że wartości zmiennej niezależnej $X_1, ..., X_n$ są deterministyczne (nie są zmiennymi losowymi) np. moment pomiaru pochylenia krzywej wieży.

Własności wynikające z teoretycznego modelu

Bezpośrednio z założeń modelu oraz relacji

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad i = 1, ..., n$$

wynika kilka faktów dotyczących zmiennych wynikowych $Y_1, ..., Y_n$:

1. $(\forall i)$ Y_i jako suma dwóch składników: deterministycznego $\beta_0 + \beta_1 X_i$ oraz losowego ϵ_i, jest zmienną losową z rozkładu normalnego. $Y_1, ..., Y_n$ są stochast. niezależne.

Własności wynikające z teoretycznego modelu

Bezpośrednio z założeń modelu oraz relacji

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad i = 1, ..., n$$

wynika kilka faktów dotyczących zmiennych wynikowych $Y_1, ..., Y_n$:

1. $(\forall i)$ Y_i jako suma dwóch składników: deterministycznego $\beta_0 + \beta_1 X_i$ oraz losowego ϵ_i, jest zmienną losową z rozkładu normalnego. $Y_1, ..., Y_n$ są stochast. niezależne.
2. Wartość oczekiwana Y_i dana jest wzorem:

   $$E(Y_i) = E(\beta_0 + \beta_1 X_i) + E(\epsilon_i) = \beta_0 + \beta_1 X_i$$

   Zatem ściśle liniowa relacja zachodzi pomiędzy wartością oczekiwana $E(Y_i)$ a X_i, natomiast relacja pomiędzy Y_i i X_i jest tylko w przybliżeniu liniowa. Z tego względu ϵ_i jest traktowany jako błąd losowy, który może wynikać np. z ograniczonej dokładności pomiarowej urządzenia badawczego.

Własności wynikające z teoretycznego modelu

Bezpośrednio z założeń modelu oraz relacji

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad i = 1, ..., n$$

wynika kilka faktów dotyczących zmiennych wynikowych $Y_1, ..., Y_n$:

1. $(\forall i)$ Y_i jako suma dwóch składników: deterministycznego $\beta_0 + \beta_1 X_i$ oraz losowego ϵ_i, jest zmienną losową z rozkładu normalnego. $Y_1, ..., Y_n$ są stochast. niezależne.
2. Wartość oczekiwana Y_i dana jest wzorem:

$$E(Y_i) = E(\beta_0 + \beta_1 X_i) + E(\epsilon_i) = \beta_0 + \beta_1 X_i$$

   Zatem ściśle liniowa relacja zachodzi pomiędzy wartością oczekiwana $E(Y_i)$ a X_i, natomiast relacja pomiędzy Y_i i X_i jest tylko w przybliżeniu liniowa. Z tego względu ϵ_i jest traktowany jako błąd losowy, który może wynikać np. z ograniczonej dokładności pomiarowej urządzenia badawczego.
3. Dla każdego i, wariancja Y_i dana jest wzorem:

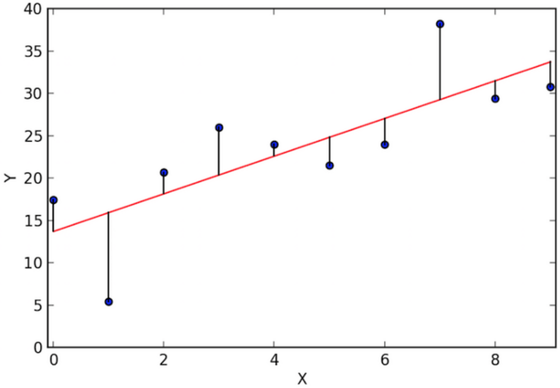
$$Var(Y_i) = Var(\beta_0 + \beta_1 X_i + \epsilon_i) = Var(\epsilon_i) = \sigma^2$$

   Wariancja Y_i (lub ϵ_i) jest zatem dokładnie taka sama dla każdego i. Jest to relatywnie naturalne założenie w sytuacji, w której dokonujemy pomiarów za pomocą tego samego urządzenia badawczego (np. linijka z centymetrową podziałką). Można wówczas zakładać, że błędy pomiarowe będą miały taką samą wariancję.

Graficzna ilustracja teoretycznego modelu

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \qquad i = 1, ..., 10$$

1. Niebiesko–czarne kropki odpowiadają punktom (X_i, Y_i),
2. Czerwona linia przedstawia zależność: $E(Y) = \beta_0 + \beta_1 X$,
3. Czarne odcinki opisują wielkość losowych błędów ϵ_i



Parametry w teoretycznym modelu

W teoretycznym modelu regresji liniowej występują 3 parametry:

- β_0 – Intercept,
- β_1 – Slope,
- σ^2 – wariancja błędów ϵ_i

Znajomość tych parametrów pozwala na generowanie wartości zmiennej odpowiedzi Y_i dla dowolnych wartości regresora X_i. Umożliwia to przeprowadzanie symulacji i empiryczne badanie własności modelu liniowego.

Parametry w teoretycznym modelu

W teoretycznym modelu regresji liniowej występują 3 parametry:

- β_0 – Intercept,
- β_1 – Slope,
- σ^2 – wariancja błędów ϵ_i

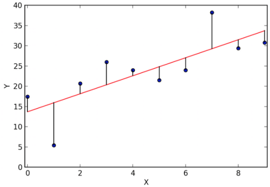
Znajomość tych parametrów pozwala na generowanie wartości zmiennej odpowiedzi Y_i dla dowolnych wartości regresora X_i. Umożliwia to przeprowadzanie symulacji i empiryczne badanie własności modelu liniowego.

Jednakże w realnych problemach zwykle dysponujemy wyłącznie zbiorem par $(X_i, Y_i)_{i=1}^n$, a parametry są nieznane. Dlatego ważnym zagadnieniem jest estymacja parametrów modelu teoretycznego.

Table of Contents

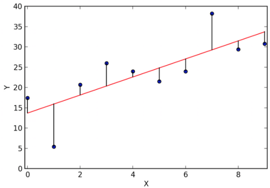
Metoda najmniejszych kwadratów (Ordinary Least Squares)

Ze względu na stochastyczną niezależność, stałą wariancję i symetryczność błędów losowych ϵ_i, naturalnym estymatorem dla prostej $E(Y) = \beta_0 + \beta_1 X$ jest prosta leżąca w jakimś sensie możliwie blisko zbioru danych $(X_i, Y_i)_{i=1}^{n}$.



Metoda najmniejszych kwadratów (Ordinary Least Squares)

Ze względu na stochastyczną niezależność, stałą wariancję i symetryczność błędów losowych ϵ_i, naturalnym estymatorem dla prostej $E(Y) = \beta_0 + \beta_1 X$ jest prosta leżąca w jakimś sensie możliwie blisko zbioru danych $(X_i, Y_i)_{i=1}^n$.



Wiemy że dowolną funkcję liniową w $\mathbb{R}^2$ możemy przedstawić w postaci

$$f(b_0, b_1|X) = b_0 + b_1 X$$

Odległość takiej prostej od punktu (X_i, Y_i) (dla odciętej X_i) wynosi

$$e_i(b_0, b_1) = Y_i - f(b_0, b_1|X_i) = Y_i - (b_0 + b_1 X_i)$$

i nazywana jest wartością resztową lub residuum dla punktu (X_i, Y_i) i parametrów (b_0, b_1).

Metoda najmniejszych kwadratów

Estymator otrzymany metodą najmniejszych kwadratów (Ordinary Least Squares OLS) to wektor $(\hat{\beta}_0, \hat{\beta}_1)'$ dobrany w taki sposób, aby zminimalizować sumę kwadratów residów:

$$(\hat{\beta}_0, \hat{\beta}_1)' = \underset{(b_0,b_1)'\in\mathbb{R}^2}{argmin} \sum_{i=1}^{n} e_i^2(b_0, b_1) = \underset{(b_0,b_1)'\in\mathbb{R}^2}{argmin} \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2$$

Metoda najmniejszych kwadratów

Estymator otrzymany metodą najmniejszych kwadratów (Ordinary Least Squares OLS) to wektor $(\hat{\beta}_0, \hat{\beta}_1)'$ dobrany w taki sposób, aby zminimalizować sumę kwadratów residów:

$$(\hat{\beta}_0, \hat{\beta}_1)' = \underset{(b_0,b_1)' \in \mathbb{R}^2}{argmin} \sum_{i=1}^{n} e_i^2(b_0, b_1) = \underset{(b_0,b_1)' \in \mathbb{R}^2}{argmin} \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2$$

Estymator $(\hat{\beta}_0, \hat{\beta}_1)'$ można wyznaczyć standardowymi metodami analizy matematycznej (pochodne z $\sum_{i=1}^{n} e_i^2(b_0, b_1)$ po b_0 i b_1 przyrównać do zera = ekstremum; dodatkowo sprawdzić warunki na minimum, np. warunki na hesjan).

Metoda najmniejszych kwadratów

Estymator otrzymany metodą najmniejszych kwadratów (Ordinary Least Squares OLS) to wektor $(\hat{\beta}_0, \hat{\beta}_1)'$ dobrany w taki sposób, aby zminimalizować sumę kwadratów residów:

$$(\hat{\beta}_0, \hat{\beta}_1)' = \underset{(b_0,b_1)' \in \mathbb{R}^2}{argmin} \sum_{i=1}^{n} e_i^2(b_0, b_1) = \underset{(b_0,b_1)' \in \mathbb{R}^2}{argmin} \sum_{i=1}^{n} (Y_i - (b_0 + b_1 X_i))^2$$

Estymator $(\hat{\beta}_0, \hat{\beta}_1)'$ można wyznaczyć standardowymi metodami analizy matematycznej (pochodne z $\sum_{i=1}^{n} e_i^2(b_0, b_1)$ po b_0 i b_1 przyrównać do zera = ekstremum; dodatkowo sprawdzić warunki na minimum, np. warunki na hesjan).
Wzory na estymator $\hat{\beta}_0$ oraz $\hat{\beta}_1$ dane są następującymi formułami:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Warto zauważyć, że wzór na $\hat{\beta}_1$ jest ilorazem próbkowej kowariancji pomiędzy zmiennymi X i Y przez próbkową wariancję zmiennej X.

Metoda największej wiarogodności (Maximum Likelihood)

Estymator otrzymany metodą Największej Wiarogodności (ML) to wektor $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)'$ dobrany w taki sposób, aby maksymalizował funkcję wiarogodności dla zbioru zmiennych odpowiedzi $Y_1, ..., Y_n$:

$$L(b_0, b_1, z^2|Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi z^2}} \exp\left(-\frac{1}{2z^2}\left(Y_i - b_0 - b_1 X_i\right)^2\right)$$

Metoda największej wiarogodności (Maximum Likelihood)

Estymator otrzymany metodą Największej Wiarogodności (ML) to wektor $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)'$ dobrany w taki sposób, aby maksymalizował funkcję wiarogodności dla zbioru zmiennych odpowiedzi $Y_1, ..., Y_n$:

$$L(b_0, b_1, z^2|Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi z^2}} \exp\left(-\frac{1}{2z^2}\left(Y_i - b_0 - b_1 X_i\right)^2\right)$$

Ze względu na to, że szukanie maksimum funkcji wiarogodności jest uciążliwe, zwykle analizie podlega jej logarytm:

$$\log(L(b_0, b_1, z^2|Y)) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log z^2 - \frac{1}{2z^2}\sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_i\right)^2$$

Metoda największej wiarogodności (Maximum Likelihood)

Estymator otrzymany metodą Największej Wiarogodności (ML) to wektor $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)'$ dobrany w taki sposób, aby maksymalizował funkcję wiarogodności dla zbioru zmiennych odpowiedzi $Y_1, ..., Y_n$:

$$L(b_0, b_1, z^2|Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi z^2}} \exp\left(-\frac{1}{2z^2}\left(Y_i - b_0 - b_1 X_i\right)^2\right)$$

Ze względu na to, że szukanie maksimum funkcji wiarogodności jest uciążliwe, zwykle analizie podlega jej logarytm:

$$\log(L(b_0, b_1, z^2|Y)) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log z^2 - \frac{1}{2z^2}\sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_i\right)^2$$

Z powyższej postaci widać, że dla dowolne wartości z^2, maksymalizacja po b_0 oraz b_1 jest równoważna minimalizacji wyrażenia:

$$\sum_{i=1}^{n}\left(Y_i - b_0 - b_1 X_i\right)^2$$

czyli minimalizacji sumy kwadratów residów. Wynika stąd, że estymatory $\hat{\beta}_0$ oraz $\hat{\beta}_1$ 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:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2$$

Można pokazać, że jest on obciążony ($E(\hat{\sigma}^2) < \sigma^2$). Nieobciążony estymator parametru σ^2 (zwykle stosowany) ma postać:

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2$$

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

Table of Contents

Teoretyczne własności estymatora $\hat{\beta}_1$

Postać estymatora: $\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$

Rozkład

Estymator $\hat{\beta}_1$ jako liniowe przekształcenie wektora odpowiedzi $Y = (Y_1, ..., Y_n)'$ pochodzi z rozkładu normalnego:

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2(\hat{\beta}_1))$$

gdzie $\sigma^2(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$.

Wartość oczekiwana:

$$E(Y_i - \bar{Y}) = \beta_0 + \beta_1 X_i - \frac{1}{n}\sum_{j=1}^n (\beta_0 + \beta_1 X_j) = \beta_1 (X_i - \bar{X})$$

$$E(\hat{\beta}_1) = \frac{\sum_{i=1}^n (X_i - \bar{X}) E(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \beta_1 \frac{\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \beta_1$$

Teoretyczne własności estymatora $\hat{\beta}_1$ (2)

Wariancja:

$$Var(\hat{\beta}_1) = cov\left(\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \frac{\sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}\right) =$$

$$= \left(\frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2}\right)^2 \sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X})(X_j - \bar{X}) cov(Y_i - \bar{Y}, Y_j - \bar{Y}) = \mathbb{A}$$

$$cov(Y_i - \bar{Y}, Y_j - \bar{Y}) = cov(Y_i, Y_j) - cov(Y_i, \bar{Y}) - cov(\bar{Y}, Y_j) + cov(\bar{Y}, \bar{Y}) =$$

$$= \sigma^2 \mathbb{I}(i = j) - \frac{1}{n}\sigma^2 - \frac{1}{n}\sigma^2 + \frac{1}{n}\sigma^2 = \sigma^2 \mathbb{I}(i = j) - \frac{1}{n}\sigma^2$$

$$\sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X})(X_j - \bar{X}) = \sum_{i=1}^n (X_i - \bar{X}) \sum_{j=1}^n (X_j - \bar{X}) = 0$$

$$\mathbb{A} = \left(\frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2}\right)^2 \sum_{i=1}^n \sum_{j=1}^n (X_i - \bar{X})(X_j - \bar{X}) \sigma^2 \mathbb{I}(i = j) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Przedział ufności dla β_1

Statystyka T

Oznaczmy przez T następującą statystykę:

$$T = \frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)} \quad \text{gdzie} \quad s^2(\hat{\beta}_1) = \frac{s^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Wówczas statystyka T pochodzi z rozkładu studenta z $n-2$ stopniami swobody $T \sim t(n-2)$

Przedział ufności dla β_1

Na podstawie statystyki T możemy skonstruować przedział ufności o współczynniku ufności $1-\alpha$ dla parametru β_1:

$$\hat{\beta}_1 \pm t_c s(\hat{\beta}_1)$$

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

Test istotności dla β_1

Test istotności dla β_1

Na podstawie statystyki T możemy testować czy β_1 jest różna od 0:

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

Statystyka testowa jest postaci:

$$T = \frac{\hat{\beta}_1 - 0}{s(\hat{\beta}_1)} \qquad \text{gdzie} \quad s^2(\hat{\beta}_1) = \frac{s^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Odrzucamy hipotezę zerową, gdy $|T| > t_c$ gdzie $t_c = t^*(1 - \frac{\alpha}{2}, n-2)$ jest kwantylem rzędu $1 - \frac{\alpha}{2}$ z rozkładu studenta z $n-2$ stopniami swobody.
p-wartość dla tego problemu:
$p = P(|z| > |T|)$, gdzie $z \sim t(n-2)$

Łatwo zmodyfikować powyższy test do badania hipotez jednostronnych.

Teoretyczne własności estymatora $\hat{\beta}_0$

Postać estymatora: $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

Rozkład

Estymator $\hat{\beta}_0$ jako liniowe przekształcenie wektora odpowiedzi $Y = (Y_1, ..., Y_n)'$ pochodzi z rozkładu normalnego:

$$\hat{\beta}_0 \sim N(\beta_0, \sigma^2(\hat{\beta}_0))$$

gdzie $\sigma^2(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right)$.

Wartość oczekiwana:

$$E(\hat{\beta}_0) = E(\bar{Y}) - E(\hat{\beta}_1)\bar{X} = (\beta_0 + \beta_1 \bar{X}) - \beta_1 \bar{X} = \beta_0$$

Wariancja:

$$Var(\hat{\beta}_0) = Var(\bar{Y} - \hat{\beta}_1 \bar{X}) = Var(\bar{Y}) + Var(\hat{\beta}_1)(\bar{X})^2 - 2cov(\bar{Y}, \hat{\beta}_1 \bar{X})$$

Teoretyczne własności estymatora $\hat{\beta}_0$ (2)

Wariancja:

$$Var(\hat{\beta}_0) = Var(\bar{Y} - \hat{\beta}_1 \bar{X}) = Var(\bar{Y}) + \bar{X}^2 Var(\hat{\beta}_1) - 2\bar{X} cov(\bar{Y}, \hat{\beta}_1)$$

$$Var(\bar{Y}) = \frac{\sigma^2}{n}$$

$$\bar{X}^2 Var(\hat{\beta}_1) = \sigma^2 \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$cov(\bar{Y}, \hat{\beta}_1) = \frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} \sum_{j=1}^n (X_j - \bar{X}) cov(\bar{Y}, Y_j - \bar{Y}) =$$

$$= \frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2} \sum_{j=1}^n (X_j - \bar{X}) \left(\frac{\sigma^2}{n} - \frac{\sigma^2}{n} \right) = 0$$

Ostatecznie:

$$Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

Przedział ufności dla β_0

Statystyka T

Oznaczmy przez T następującą statystykę:

$$T = \frac{\hat{\beta}_0 - \beta_0}{s(\hat{\beta}_0)} \quad \text{gdzie} \quad s^2(\hat{\beta}_0) = s^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

Wówczas statystyka T pochodzi z rozkładu studenta z $n-2$ stopniami swobody $T \sim t(n-2)$

Przedział ufności dla β_0

Na podstawie statystyki T możemy skonstruować przedział ufności o współczynniku ufności $1-\alpha$ dla parametru β_0:

$$\hat{\beta}_0 \pm t_c s(\hat{\beta}_0)$$

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

Test istotności dla β_0

Test istotności dla β_0

Na podstawie statystyki T możemy testować czy β_0 jest równa pewnej dowolnej stałej β_{00}:

$$H_0: \quad \beta_0 = \beta_{00} \qquad vs \qquad H_1: \quad \beta_0 \neq \beta_{00}$$

Statystyka testowa jest postaci:

$$T = \frac{\hat{\beta}_0 - \beta_{00}}{s(\hat{\beta}_0)} \qquad \text{gdzie} \quad s^2(\hat{\beta}_0) = s^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)$$

Odrzucamy hipotezę zerową, gdy $|T| > t_c$ gdzie $t_c = t^*(1 - \frac{\alpha}{2}, n-2)$ jest kwantylem rzędu $1 - \frac{\alpha}{2}$ z rozkładu studenta z $n-2$ stopniami swobody.
p-wartość dla tego problemu:
$p = P(|z| > |T|)$, gdzie $z \sim t(n-2)$

Łatwo zmodyfikować powyższy test do badania hipotez jednostronnych.

Uwagi

- W praktycznych zagadnieniach, zwykle istotne są przede wszystkim pytania dotyczące parametru β_1 (np.: $\beta_1 = 0$). Jeżeli $\beta_1 = 0$, to $Y_i = \beta_0 + \epsilon_i \sim N(\beta_0, \sigma^2)$. W konsekwencji mamy brak związku pomiędzy Y_i a X_i,

Uwagi

- W praktycznych zagadnieniach, zwykle istotne są przede wszystkim pytania dotyczące parametru β_1 (np.: $\beta_1 = 0$). Jeżeli $\beta_1 = 0$, to $Y_i = \beta_0 + \epsilon_i \sim N(\beta_0, \sigma^2)$. W konsekwencji mamy brak związku pomiędzy Y_i a X_i,
- Wariancja estymatora $\hat{\beta}_1$ dana jest wzorem $\sigma^2(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$.
  Jej postać daje wskazówkę jak konstruować eksperyment, aby $\sigma^2(\hat{\beta}_1)$ była mała (tak dobrać wektor $X = (X_1, ..., X_n)$ aby mianownik był możliwie duży),

Uwagi

- W praktycznych zagadnieniach, zwykle istotne są przede wszystkim pytania dotyczące parametru β_1 (np.: $\beta_1 = 0$). Jeżeli $\beta_1 = 0$, to $Y_i = \beta_0 + \epsilon_i \sim N(\beta_0, \sigma^2)$. W konsekwencji mamy brak związku pomiędzy Y_i a X_i,
- Wariancja estymatora $\hat{\beta}_1$ dana jest wzorem $\sigma^2(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$. Jej postać daje wskazówkę jak konstruować eksperyment, aby $\sigma^2(\hat{\beta}_1)$ była mała (tak dobrać wektor $X = (X_1, ..., X_n)$ aby mianownik był możliwie duży),
- Jeżeli ϵ_i nie pochodzą z rozkładu normalnego, tylko z rozkładu o podobnych własnościach (np. symetryczność, stała wariancja) to przedziały ufności i testy zwykle mają dobre własności.

Komendy w R

Konstrukcja modelu na podstawie próby, testy i przedziały ufności w R:

- $reg1 < -lm(LEAN \sim YEAR, piza)$ – funkcja konstruująca model (na podstawie próby),
- $reg1$ – model + estymatory β_0, β_1,
- $summary(reg1)$ – funkcja zwracająca podsumowanie własności modelu,
- $reg1\$residuals$ – wektor wartości resztowych,
- $confint(reg1)$ – funkcja zwracająca 95% przedziały ufności dla β_0 oraz β_1,

Funkcja mocy testu

Moc testu binarnego (o dwoch hipotezach prostych) to prawdopodobieństwo odrzucenia H_0, gdy prawdziwa jest hipoteza H_1 (pr. uniknięcia błędu II rodzaju).

Na przykład dla testu S badającego następujący problem:

$$H_0 \quad : \quad \theta = 0 \qquad vs \qquad H_1 \quad : \quad \theta = a$$

moc jest równa

$$\pi(a) = P_{\theta=a}(\text{test S odrzucił hipotezę } H_0)$$

Funkcja mocy testu

Moc testu binarnego (o dwoch hipotezach prostych) to prawdopodobieństwo odrzucenia H_0, gdy prawdziwa jest hipoteza H_1 (pr. uniknięcia błędu II rodzaju).

Na przykład dla testu S badającego następujący problem:

$$H_0 \quad : \quad \theta = 0 \qquad vs \qquad H_1 \quad : \quad \theta = a$$

moc jest równa

$$\pi(a) = P_{\theta=a}(\text{test S odrzucił hipotezę } H_0)$$

Łatwo zauważyć, że moc zależy od rozpatrywanej alternatywy. Dlatego funkcja $\pi(a) : \Theta_1 \rightarrow [0, 1]$, z przestrzeni możliwych alternatyw (Θ_1) w odcinek $[0, 1]$, nazywana jest funkcją mocy testu.

Funkcja mocy testu dla β_1

Testujemy nastepujacy problem $H_0: \ \beta_1 = 0 \qquad vs \qquad H_1: \ \beta_1 \neq 0.$
Statystyka testowa oraz wartość krytyczna są postaci:

$$T = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)}, \qquad t_c = t^*(1 - \frac{\alpha}{2}, n-2)$$

Odrzucamy hipotezę zerową, gdy $|T| > t_c$.

Funkcja mocy testu dla β_1

Testujemy nastepujacy problem $H_0: \quad \beta_1 = 0 \qquad vs \qquad H_1: \quad \beta_1 \neq 0.$
Statystyka testowa oraz wartość krytyczna są postaci:

$$T = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)}, \qquad t_c = t^*(1 - \frac{\alpha}{2}, n-2)$$

Odrzucamy hipotezę zerową, gdy $|T| > t_c$.

Aby wyznaczyć funkcję mocy testu musimy obliczyć
$\pi(a) = P_{\beta_1 = a}(|T| > t_c)$ dla dowolnej wartości $a \in \mathbb{R} \setminus \{0\}$ (Co gdy $a = 0$?)

Funkcja mocy testu dla β_1

Testujemy nastepujacy problem $H_0: \quad \beta_1 = 0 \qquad vs \qquad H_1: \quad \beta_1 \neq 0.$
Statystyka testowa oraz wartość krytyczna są postaci:

$$T = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)}, \qquad t_c = t^*(1 - \frac{\alpha}{2}, n-2)$$

Odrzucamy hipotezę zerową, gdy $|T| > t_c$.

Aby wyznaczyć funkcję mocy testu musimy obliczyć
$\pi(a) = P_{\beta_1 = a}(|T| > t_c)$ dla dowolnej wartości $a \in \mathbb{R} \setminus \{0\}$ (Co gdy $a = 0$?)
Przy założeniu że $\beta_1 = a$ statystyka T ma niecentralny rozkład studenta z $n-2$ stopniami swobody i parametrem niecentralności $\delta = \beta_1 / \sigma(\hat{\beta}_1)$ ($T \sim t(n-2, \delta)$)

Funkcja mocy testu dla β_1

Testujemy nastepujacy problem $H_0: \ \beta_1 = 0 \quad vs \quad H_1: \ \beta_1 \neq 0.$
Statystyka testowa oraz wartość krytyczna są postaci:

$$T = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)}, \qquad t_c = t^*(1 - \frac{\alpha}{2}, n-2)$$

Odrzucamy hipotezę zerową, gdy $|T| > t_c$.

Aby wyznaczyć funkcję mocy testu musimy obliczyć
$\pi(a) = P_{\beta_1 = a}(|T| > t_c)$ dla dowolnej wartości $a \in \mathbb{R} \setminus \{0\}$ (Co gdy $a = 0$?)
Przy założeniu że $\beta_1 = a$ statystyka T ma niecentralny rozkład studenta z $n-2$ stopniami swobody i parametrem niecentralności $\delta = \beta_1/\sigma(\hat{\beta}_1)$ ($T \sim t(n-2, \delta)$)

Aby móc obliczyć moc dla $\beta_1 = a$, musimy znać następujące wartości:

$$n, \sum_{i=1}^{n}(X_i - \bar{X})^2, \sigma^2$$

Moc testu dla $\beta_1 = 1.5$ - Przykład

Załóżmy, że $n = 25, \sum_{i=1}^{n}(X_i - \bar{X})^2 = 19800, \sigma^2 = 2500.$

Wówczas $\sigma^2(\hat{\beta}_1) = \sigma^2 / \sum_{i=1}^{n}(X_i - \bar{X})^2 = 0.1263.$

Przyjmijmy dodatkowo, że $\beta_1 = 1.5$, możemy wtedy wyznaczyć parametr niecentralności $\delta = \beta_1/\sigma(\hat{\beta}_1) = 4.22.$

Ponieważ wiemy że statystyka $T \sim t(23, 4.22)$, zatem (np. przy pomocy funkcji w R) możemy wyznaczyć moc

$$\pi(1.5) = P_{\beta_1=1.5}(|T| > t_c) = P_{\beta_1=1.5}(T < -t_c) + P_{\beta_1=1.5}(T > t_c) =$$

$$= F_{\beta_1=1.5}(-t_c) + 1 - F_{\beta_1=1.5}(t_c) = 0.98$$

Funkcja mocy testu dla β_1 - Przykład

Funkcja mocy testu dla β_1 z przedziału $[-2, 2]$:

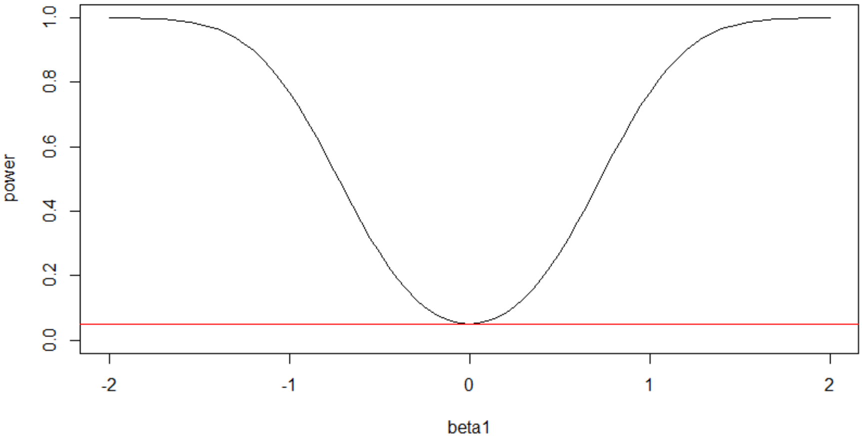


Table of Contents

Estymacja wartości oczekiwanej Y_h ($E(Y_h) = \mu_h$)

Do tej pory opisane zostały własności estymatorów parametrów w modelu teoretycznym $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)$. Jednakże, pierwotnym zagadnieniem była analiza zachowania zmiennej objaśnianej Y_h w zależności od zmiennej objaśniającej X_h.

Estymacja wartości oczekiwanej Y_h ($E(Y_h) = \mu_h$)

Do tej pory opisane zostały własności estymatorów parametrów w modelu teoretycznym ($\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2$). Jednakże, pierwotnym zagadnieniem była analiza zachowania zmiennej objaśnianej Y_h w zależności od zmiennej objaśniającej X_h.

Skoncentrujmy się na własnościach wartości oczekiwanej ze zmiennej zależnej $E(Y_h) = \mu_h = \beta_0 + \beta_1 X_h$ dla podpopulacji o wartości zmiennej niezależnej równej X_h.

Estymator μ_h dany jest następującą zależnością: $\hat{\mu}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h$

Rozkład

Estymator $\hat{\mu}_h$ jako liniowe przekształcenie wektora odpowiedzi $Y = (Y_1, ..., Y_n)'$ pochodzi z rozkładu normalnego:

$$\hat{\mu}_h \sim N(\mu_h, \sigma^2(\hat{\mu}_h))$$

gdzie $\sigma^2(\hat{\mu}_h) = \sigma^2 \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right)$.

Własności teoretyczne $E(Y_h)$

Wartość oczekiwana:

$$E(\hat{\mu}_h) = E(\hat{\beta}_0) + E(\hat{\beta}_1)X_h = \beta_0 + \beta_1 X_h = \mu_h$$

Wariancja:

$$\sigma^2(\hat{\mu}_h) = Var(\hat{\beta}_0 + \hat{\beta}_1 X_h) = Var(\bar{Y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_h) = Var(\bar{Y} + \hat{\beta}_1(X_h - \bar{X})) =$$

$$= Var(\bar{Y}) + Var(\hat{\beta}_1)(X_h - \bar{X})^2 + 2(X_h - \bar{X})\underbrace{cov(\bar{Y}, \hat{\beta}_1)}_{=0} =$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

Przedział ufności oraz testowanie dla $E(Y_h)$

Statystyka T

Oznaczmy przez T następującą statystykę:

$$T = \frac{\hat{\mu}_h - E(\hat{\mu}_h)}{s(\hat{\mu}_h)} \quad \text{gdzie} \quad s^2(\hat{\mu}_h) = s^2 \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \right)$$

Wówczas statystyka T pochodzi z rozkładu studenta z $n-2$ stopniami swobody $T \sim t(n-2)$

Na podstawie statystyki T możemy skonstruować przedział ufności o współczynniku ufności $1-\alpha$ dla parametru $E(Y_h)$:

$$\hat{\mu}_h \pm t_c s(\hat{\mu}_h)$$

Możemy również wykonać testowanie analogicznie jak dla $\hat{\beta}_0$ i $\hat{\beta}_1$.

Predykcja punktowa dla nowej wartości zmiennej odpowiedzi Y_h

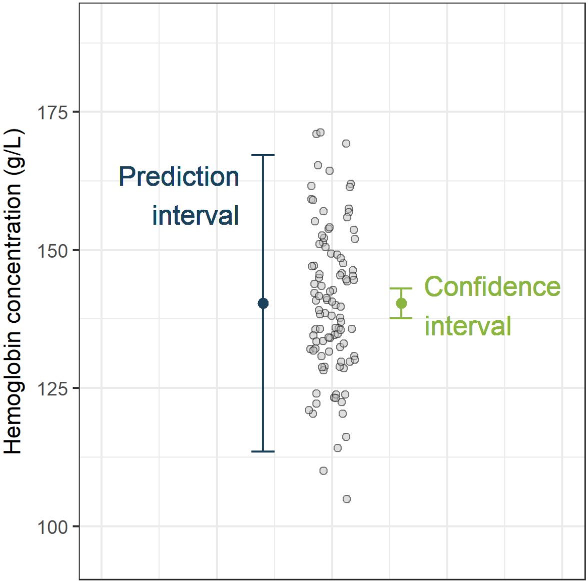
Przed momentem rozważyliśmy własności estymatora wartości oczekiwanej ze zmiennej objaśnianej μ_h. Co jednak zrobić gdy chcemy przewidzieć wartość nowej zmiennej zależnej Y_h.
$Y_h = \beta_0 + \beta_1 X_h + \epsilon$ – wartość nowej zmiennej zależnej dla zmiennej niezależnej o wartości X_h.

Predykcja punktowa dla nowej wartości zmiennej odpowiedzi Y_h

Przed momentem rozważyliśmy własności estymatora wartości oczekiwanej ze zmiennej objaśnianej μ_h. Co jednak zrobić gdy chcemy przewidzieć wartość nowej zmiennej zależnej Y_h.
$Y_h = \beta_0 + \beta_1 X_h + \epsilon$ – wartość nowej zmiennej zależnej dla zmiennej niezależnej o wartości X_h.

Predykcja punktowa Y_h jest postaci $\hat{Y}_h = \hat{\mu}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h$. Ma on tę samą postać co przed momentem rozważany estymator $E(Y_h)$. Nie jest to zaskakujące, gdyż z założenia nie potrafimy modelować zachowania losowego błędu ϵ, który ma symetryczny rozkład. Jednakże, ze względu na ten sam błąd ϵ zmieniają się własności wariancji predykcji.

Przedział ufności vs. przedział predykcyjny



Przedział predykcyjny dla nowej wartości zmiennej odpowiedzi Y_h

Wariancja błędu predykcji jest postaci:

$$Var(Y_h - \hat{\mu}_h) = Var(Y_h) + Var(\hat{\mu}_h) = \sigma^2 + \sigma^2 \left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) =$$

$$= \sigma^2 \left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

Widzimy zatem, że jest ona większa niż wariancja estymatora $E(Y_h)$. Przekłada się to na szerokość przedziału predykcyjnego, która jest większa od szerokości przedziału ufności dla $E(Y_h)$.

$$T = \frac{Y_h - \hat{\mu}_h}{s(pred)} \sim t(n-2) \quad \text{gdzie} \quad s^2(pred) = s^2 \left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

Przedział predykcyjny dla Y_h jest postaci:

$$\hat{\mu}_h \pm t_c s(pred)$$

Table of Contents

Pasmo ufności dla prostej regresji $E(Y|X) = \beta_0 + \beta_1 X$

Na poprzednich slajdach opisano teorię dotyczącą wartości oczekiwanej $E(Y_h) = \mu_h$ dla konkretnej podpopulacji o wartości zmiennej zależnej X_h. W praktycznych zagadnieniach pytamy czasami o równoczesne przedziały ufności dla kilku wartości średniej μ, np. $\mu_{1'}, ..., \mu_{d'}$, lub o tak zwane pasmo ufności dla prostej regresji (równoczesne przedziały ufności dla wszystkich możliwych wartości μ_h).

Pasmo ufności dla prostej regresji $E(Y|X) = \beta_0 + \beta_1 X$

Na poprzednich slajdach opisano teorię dotyczącą wartości oczekiwanej $E(Y_h) = \mu_h$ dla konkretnej podpopulacji o wartości zmiennej zależnej X_h. W praktycznych zagadnieniach pytamy czasami o równoczesne przedziały ufności dla kilku wartości średniej μ, np. $\mu_{1'}, ..., \mu_{d'}$, lub o tak zwane pasmo ufności dla prostej regresji (równoczesne przedziały ufności dla wszystkich możliwych wartości μ_h).

Na podstawie teorii opisującej łączny rozkład estymatorów $(\hat{\beta}_0, \hat{\beta}_1)'$, pasmo ufności (na poziomie α) dla prostej regresji (w dowolnym punkcie X_h) ma postać:

$$\hat{\mu}_h \pm Ws(\hat{\mu}_h)$$

gdzie $W^2 = 2F(1-\alpha, 2, n-2)$ jest podwojonym kwantylem rzędu $1-\alpha$ z rozkładu Fishera-Snedecora z 2 i $n-2$ stopniami swobody.

Pasmo ufności dla prostej regresji $E(Y|X) = \beta_0 + \beta_1 X$

Na poprzednich slajdach opisano teorię dotyczącą wartości oczekiwanej $E(Y_h) = \mu_h$ dla konkretnej podpopulacji o wartości zmiennej zależnej X_h. W praktycznych zagadnieniach pytamy czasami o równoczesne przedziały ufności dla kilku wartości średniej μ, np. $\mu_{1'}, ..., \mu_{d'}$, lub o tak zwane pasmo ufności dla prostej regresji (równoczesne przedziały ufności dla wszystkich możliwych wartości μ_h).

Na podstawie teorii opisującej łączny rozkład estymatorów $(\hat{\beta}_0, \hat{\beta}_1)'$, pasmo ufności (na poziomie α) dla prostej regresji (w dowolnym punkcie X_h) ma postać:

$$\hat{\mu}_h \pm W s(\hat{\mu}_h)$$

gdzie $W^2 = 2F(1-\alpha, 2, n-2)$ jest podwojonym kwantylem rzędu $1-\alpha$ z rozkładu Fishera-Snedecora z 2 i $n-2$ stopniami swobody.

Powyższy wzór daje równoczesne przedziały ufności dla wszystkich wartości X_h.

Table of Contents

Podsumowanie

Dotychczas poznaliśmy teoretyczny model regresji liniowej prostej

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad i = 1, ..., n$$

gdzie ϵ_i – niezależne zmienne losowe z rozkładu $N(0, \sigma^2)$,
oraz:

Podsumowanie

Dotychczas poznaliśmy teoretyczny model regresji liniowej prostej

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad i = 1, ..., n$$

gdzie ϵ_i – niezależne zmienne losowe z rozkładu $N(0, \sigma^2)$,
oraz:

- metody estymacji (OLS i ML) parametrów w modelu β_0, β_1 i σ^2,
- własności teoretyczne estymatorów $\hat{\beta}_0, \hat{\beta}_1$ i $\hat{\sigma}^2$
- teorię dotyczącą przedziałów ufności, testowania hipotez oraz mocy testów dla $\hat{\beta}_0$ i $\hat{\beta}_1$,

Podsumowanie

Dotychczas poznaliśmy teoretyczny model regresji liniowej prostej

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad i = 1, ..., n$$

gdzie ϵ_i – niezależne zmienne losowe z rozkładu $N(0, \sigma^2)$,
oraz:

- metody estymacji (OLS i ML) parametrów w modelu β_0, β_1 i σ^2,
- własności teoretyczne estymatorów $\hat{\beta}_0, \hat{\beta}_1$ i $\hat{\sigma}^2$
- teorię dotyczącą przedziałów ufności, testowania hipotez oraz mocy testów dla $\hat{\beta}_0$ i $\hat{\beta}_1$,
- teorię dotyczącą estymacji wartości oczekiwanej $E(Y_h) = \mu_h$, testowania hipotez oraz przedziałów ufności dla podpopulacji o wartości zmiennej niezależnej równej X_h,
- teorię dotyczącą pasma ufności dla prostej regresji. Innymi słowy równoczesnych przedziałów ufności dla wszystkich wartości średnich $E(Y_h) = \mu_h$,

Podsumowanie

Dotychczas poznaliśmy teoretyczny model regresji liniowej prostej

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad i = 1, ..., n$$

gdzie ϵ_i – niezależne zmienne losowe z rozkładu $N(0, \sigma^2)$,
oraz:

- metody estymacji (OLS i ML) parametrów w modelu β_0, β_1 i σ^2,
- własności teoretyczne estymatorów $\hat{\beta}_0, \hat{\beta}_1$ i $\hat{\sigma}^2$
- teorię dotyczącą przedziałów ufności, testowania hipotez oraz mocy testów dla $\hat{\beta}_0$ i $\hat{\beta}_1$,
- teorię dotyczącą estymacji wartości oczekiwanej $E(Y_h) = \mu_h$, testowania hipotez oraz przedziałów ufności dla podpopulacji o wartości zmiennej niezależnej równej X_h,
- teorię dotyczącą pasma ufności dla prostej regresji. Innymi słowy równoczesnych przedziałów ufności dla wszystkich wartości średnich $E(Y_h) = \mu_h$,
- teorię dotyczącą przedziałów predykcyjnych dla nowych wartości zmienne zależnej Y_h.

Podsumowanie

W zależności od pytania badacza koncentrujemy się na różnych elementach poznanej teorii.

- Gdy badamy siłę relacji pomiędzy zmienną objaśnianą i zmienną objaśniającą, koncentrujemy się na własnościach parametru β_1.
- Gdy interesuje nas średnie zachowanie zmiennej objaśnianej dla konkretnej podpopulacji (X_h), koncentrujemy się na własnościach wartości oczekiwanej μ_h – np. przedziały ufności,
- Gdy koncentrujemy się na predykcji zachowania nowych wartości zmiennej wynikowej, korzystamy z teorii dotyczącej przedziałów predykcyjnych,
- Gdy interesuje nas średnie zachowanie zmiennej objaśnianej dla dowolnej wartości X_h, korzystamy z teorii dotyczącej pasma ufności.