① Parte 1 Repaso

1 — Un modelo estadistico es una representación simplificada de la relación entre variables. Que intenta predecir/describir sus comportamientos entre sus elementos están:

variable dependiente : Y que depende de X

variable independiente: X influye en Y, pueden ser varias $(X_1, X_2, \ldots, X_p)$

Parametros :

coeficientes β_0 intercepto y pendientes $\beta_1, \ldots$ que miden el cambio en Y por X.

Supuestos: Describen como se distribuyen los errores o datos.

Ej.

RLS : $Y = \beta_0 + \beta_1 X + \varepsilon$

LLM : $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$

Donde ε es la variabilidad no explicada

2- es una cantidad que depende del azar, cambia de valor cada que se repite el experimento.

3- es una formula que se utiliza datos muestrales para aproximar un parametro poblacional.

no sesgado

$(X^T X)^{-1} X^T E[\varepsilon] = E[B] = B$

4- es una función o transformación que preserva/mantiene las operaciones de suma vectorial y multiplicación por escalar. Se cumple:

① $T(u+v) = T(u) + T(v)$

② $T(cu) = cT(u)$.

Preserva la estructura lineal!

8- Es una función de los datos y del parametro, tiene una distribución conocida y que está libre de parámetros desconocidos.

6- Se encuentra una cantidad pivotal con distribución conocida. Se fija una probabilidad central, se despeja el parámetro lo cual da el intervalo de confianza

7- Esperanza, el valor promedio esperado.

$$E[X] = \sum x_i P(x_i)$$

- $E[aX+b] = aE[X]+b$.
- $E[X+Y] = E(X) + E(Y)$
- $E(\sum a_i X_i) = \sum a_i E(x_i)$

Varianza: dispersión respecto a la media.

- $Var(AZ) = A\,Var(Z)A'$
- $Var(A+B) = Var(A) + Var(B) + 2Cov(A,B)$
- $V(aX) = a^2 V(X)$
- $V(aX \pm b^2 Y) = a^2 V(X) + b^2 V(Y) \pm 2ab\,Cov(X,Y)$
- $V(\sum a_i X_i) = \sum a_i^2 V(X_i) + \sum a_i a_j Cov(X_i X_j)$.

8- es la probabilidad de obtener un resultado al menos tan extremo como el observado por los datos.

$p \leq \alpha$ rechaza H_0

$p > \alpha$ no se rechaza H_0

En RLS

$H_0: B = 0$ no hay relación entre x y Y

Se hace estadístico de prueba

Sirve para ver qe variables contribuyen al modelo. La prueba F en ANOVA evalúa si al menos un predictor es útil.

$H_0: \beta_1 = \beta_2 \cdots = 0$

el p-value mide la fuerza de la evidencia contra H_0

9- establece qe la prueba basada en la razón de verosimilitud es la más poderosa entre las que tienen el mismo nivel de significancia

$$\Lambda = \frac{L(\theta_0)}{L(\theta_1)} \Rightarrow \text{Rechaza } H_0 \text{ si}$$

$$\Lambda \leq c_\alpha$$

si los datos son mucho más probables bajo H_1 qe bajo H_0 se rechaza H_0

$H_0: \theta = \theta_0$

$H_1: \theta = \theta_1$

Se usa para construcción de intervalos de confianza

10- Si $Z \sim N(0,1) \Rightarrow Z^2 \sim X_1^2$

de hecho la t-student = $\dfrac{\text{Normal}}{\text{chi cuadrada}}$

las variables aleatorias i.i.d con $N(0,1)$ son variables

$\sum Z_i^2 \sim X_k^2$

que siguen distribución chi cuadrado con k grados libres

11- La X^2 no centrada aparece cuando la media del vector normal no es cero.

$Z \sim N(\mu, I) \Rightarrow Z^T Z \sim X^2(\lambda)$

parametro de no centralidad? $\lambda = ||\mu||^2$

12- Un estimador no sesgado - es el que su valor esperado coincide con el parametro verdadero. En promedio acierta el valor del parametro

$X^T X \ (p \times p)$

Parte 2.

1- RLS $Y_i = B_0 + B_1 X_i + \mathcal{E}_i$

RLM $Y_i = B_0 + B_1 X_{1i} + B_2 X_{2i} + \dots + B_p X_{pi} + \mathcal{E}_i$

Supuestos

$E[\mathcal{E}] = 0$

- errores independientes entre si.

$Var(\mathcal{E}) = \sigma^2$

$Cov(\mathcal{E}_i, \mathcal{E}_j) = 0 \ \forall i \neq j$

Supuesto RLM

$\bar{Y} = X\bar{B} + \bar{\mathcal{E}}_i$

$E(\mathcal{E}) = 0$

$Var(\mathcal{E}) = \sigma^2 I$

- errores independientes

No multicolinealidad.

Ambos explican la relacion entre Y's y X'

Se estiman sus parametros

Comparten supuestos

La RLM es una generalización de RLS

2 -

el error

$$\sum_i \varepsilon_i^2 = \sum_i \left(Y_i - \hat{Y} \right)^2 \qquad \hat{Y} \xrightarrow{\text{estimada}} Y$$

$$\hat{Y} = \beta_0 + \beta_1 X_i$$

$$\Rightarrow \underbrace{\sum_i \left(Y_i - \underbrace{(\beta_0 + \beta_1 X_i)}_{-1}\right)^2}_{} = F(\beta_0, \beta_1)$$

función de perdida

(1) $$\frac{\partial F}{\partial \beta_0} = -2(Y_i - \beta_0 - \beta_1 X_i) = -2\sum_i (Y_i - \beta_0 - \beta_1 X_i)$$

$$= -2\sum_i Y_i + 2n\beta_0 + 2\beta_1 \sum X_i$$

(2) $$\frac{\partial F}{\partial \beta_1} = -2\sum_i (Y_i - (\beta_0 + \beta_1 X_i))X_i$$

$$= -2\sum_i X_i Y_i + 2\beta_0 \sum X_i + 2\beta_1 \sum X_i^2$$

- (1) $* \frac{1}{n}\sum X_i \Rightarrow -\frac{2}{n}\sum Y_i \sum X_i + 2\beta_0 \sum X_i$
$$+ \frac{2}{n}\beta_1 \left(\sum X_i\right)^2 = 0$$

- (2) − (1):

$$-2\sum X_i Y_i + 2\beta_0 \sum X_i + 2\beta_1 \sum X_i^2 + \frac{2}{n}\sum Y_i \sum X_i - 2\beta_0 \sum X_i$$
$$- \frac{2}{n}\beta_0 \left(\sum X_i\right)^2 = 0$$

$$B_1 \left[\sum x_i^2 - \frac{1}{n}\frac{n}{n}\left(\sum x_i\right)^2 \right] = \sum x_i Y_i - \frac{n}{n^2}\sum x_i \sum y_i$$

$$B_1 \left[\sum X_i^2 - n\bar{x}^2 \right] = \sum x_i Y_i - n\bar{x}\bar{Y}$$

$$B_1 = \frac{\sum_i X_i Y_i - n\bar{x}\bar{Y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum_i (x_i - \bar{x}) Y_i}{\sum_i (x_i - \bar{x})^2}$$

muestra que $\hat{B_1}$ mide la covarianza entre X y Y dividida por la varianza de X

$$\hat{\beta}_1 = \frac{Cov(X,Y)}{Var(X)}$$

3- $B\phi = Y_i - B1\, X_i$

$$B\phi = \bar{Y} - B_1 \bar{X} \Rightarrow \hat{B\phi} = \bar{Y} - \hat{B}_1 \bar{X}$$

4- Varianza $\hat{\beta_1}$, $\hat{\beta}_1 = \sum c_i Y_i$

$$V(B_1) = V\left(\sum c_i Y_i\right) \qquad c_i = \frac{x_i - \bar{x}}{(x_i - \bar{x})^2} = \frac{1}{x_i - \bar{x}}$$

$$V\left(\sum c_i Y_i\right) = \sum c_i^2 V(Y_i) + \sum_{ij} c_i c_j \cancel{cov(Y_i Y_j)}^{0}$$

$$= \sum c_i^2 V(Y_i) = \sum c_i^2 \sigma^2 = \sigma^2 \sum c_i^2$$

$$= \sigma^2 \sum_i \left[\frac{1}{x_i - \bar{x}} \right]^2 = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2} = \frac{\sigma^2}{Sxx}$$

5 -

$$V(\hat{\beta}_0) = V(\bar{Y} - \hat{\beta}_1 \bar{x})$$

$$= V(\bar{Y}) + \bar{x}^2 V(\hat{\beta}_1) - 2\bar{x}\,Cov(\bar{Y}, \hat{\beta}_1)$$

$$= V(\bar{Y}) + \bar{x}^2 \frac{\sigma^2}{Sxx}$$

$$= \frac{\sigma^2}{n} + \bar{x}^2 \frac{\sigma^2}{Sxx}$$

6 - el estimador no debería ser dependiente del modelo. se construye con varias observaciones de y para x_0 o con conocimiento a priori de σ^2

Cuando No, σ^2 se obtiene de los residuos o suma del cuadrados del error

$$SSE = \sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$$

Se propone como estimador σ^2

$$\hat{\sigma}^2 = \frac{SSE}{n-2} \rightarrow \text{grados de libertad}$$

Su distribución dada

$$\sum_i \left[\frac{\hat{Y}_i - E(\hat{Y}_i)}{\sigma}\right]^2 = \sum_i \left[\frac{\hat{Y}_i - Y_i}{\sigma}\right]^2$$

$$= \frac{(n-2)\,SSE}{(n-2)\,\sigma^2} = \frac{(n-2)}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{(n-2)}$$

La suma de los cuadrados de variables normales independientes sigue una distribución χ^2

Esto sirve para construir los intervalos de confianza.

7- No sesgo para $\hat{\beta}_0$ y $\hat{\beta}_1$

nos interesa para construir el modelo ajustado.

$$\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{\varepsilon} \; ; \; \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$$

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

$$= \beta_0 + \beta_1\bar{X} + 0 - \bar{X}\beta_1 = \beta_0$$

- $E(\hat{\beta}_1) = E\left(\sum c_i Y_i\right)$

$$= \sum_i c_i E(Y_i)$$

$$= \sum_i c_i \, E(\beta_0 + \beta_1 X_i)$$

$$= \sum_i c_i(\beta_0 + \beta_1 X_i) = \underbrace{\sum c_i}_{0} \beta_0 \sum c_i X_i \beta_1$$

$$= 0 + \beta_1 = \beta_1$$

8- Intervalos de confianza - método pivotal

la cantidad pivotal para $\hat{\beta}_1$

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_\emptyset}{SE(\hat{\beta})} \qquad T = \frac{\hat{\beta}_1 - \beta}{SE(\beta_1)}$$

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{Sxx}}} \sim N(0,1) \quad \text{no conocemos } \sigma\,?$$

$$T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{Sxx}}} \sim t_{n-2} \qquad y$$

$$T = \frac{\hat{\beta}_\emptyset - \beta_0}{SE(\hat{\beta}_\emptyset)} = \frac{\beta_\emptyset - \beta_\emptyset}{\sqrt{\hat{\sigma}^2\left[\frac{1}{n} + \frac{\bar{x}^2}{Sxx}\right]}} \sim t_{n-2}$$

Intervalo de confianza para $\hat{\beta}_1$

$$\beta_1 \in \left(\hat{\beta}_1 \pm t_{1-\frac{\alpha}{2}, n-2} \cdot \sqrt{\frac{\hat{\sigma}^2}{Sxx}}\right)$$

$$\beta_\emptyset \in \left(\hat{\beta}_\emptyset \pm t_{(1-\frac{\alpha}{2}), n-2} \cdot \sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{Sxx}\right)}\right)$$

Pruebas de Hipotesis

$H_\emptyset : \beta_i = \beta_i^*$; $i = 0, 1$

$$H_a : \begin{cases} \beta_i \neq \beta_i^* \\ \beta_i > \beta_i^* \\ \beta_i < \beta_i^* \end{cases}$$

9- RLM

Modelo

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_K X_{iK} + \varepsilon_i \quad i = 1, 2, \dots, n.$$

En matricial

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

$$\underline{Y}_{(n\times1)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} ; \quad X_{n\times p} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1K} \\ 1 & X_{21} & X_{22} & \dots & X_{2K} \\ \vdots & & & \ddots & \\ 1 & X_{n1} & X_{n2} & \dots & X_{nK} \end{pmatrix}$$

$p = K+1$

$$\underline{\beta}_{p\times1} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_K \end{pmatrix} ; \quad \varepsilon_{n\times1} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Estructura del error

$$E(\underline{\varepsilon}) = \underline{0} \; ; \quad \mathbb{V}(\underline{\varepsilon}) = \sigma^2 \mathbb{I}_{n\times n}$$

Propiedades

- $$\mathbb{V}(X) = \begin{bmatrix} V(X_1) & \dots & Cov(X_1 X_p) \\ Cov(X_2 X_1) & & \vdots \\ Cov(X_p, X_1) & & \mathbb{V}(X_p) \end{bmatrix}$$

- $$E[X] = \begin{bmatrix} E[X_1] \\ \vdots \\ E[X_p] \end{bmatrix}$$

10 - Supuestos Gauss-Markov

supon que $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$

con $\mathbb{E}(\underline{\varepsilon}) = 0$

$\mathbb{V}(\underline{\varepsilon}) = \sigma^2 \mathbb{I}$

el estimador de mínimos cuadrados es el mejor estimador linal insesgado

11 -

+ $\varepsilon_i \sim N(0, \sigma^2)$

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

$Y_i \sim N(\underline{X}_i \underline{\beta}, \sigma^2)$ $i = 1, \dots, n$

con X_i el vector de covariables del i-ésimo

+ $\underline{\hat{\beta}} \sim N_p(\underline{\beta}, \sigma^2 (X^T X)^{-1})$

+ $\hat{\beta}_j \sim N(\beta_j, \sigma^2 C_{jj})$; $j = 0, \dots, p$

C_{jj} el j-esimo elemento en la diagonal de la matriz $(X^T X)^{-1}$

12- $\beta^T: 1 \times p$ $\quad X^T X \; (p \times p)$

$X^T: p \times n$

$(X^T X)^{-1}: p \times p$

$\beta = X^T X = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$ $\quad (X^T X)^{-1} = \frac{1}{n S_{xx}} \begin{pmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix}$

$X^T \underline{Y} = \begin{pmatrix} n\bar{y} \\ \sum x_i y_i \end{pmatrix}$

$\hat{\beta} = \frac{1}{n S_{xx}} \begin{pmatrix} \sum x_i^2 & -n\bar{x} \\ n\bar{x} & n \end{pmatrix} \begin{pmatrix} n\bar{y} \\ \sum_i x_i y_i \end{pmatrix}$

$= \frac{1}{n \sum (x_i - \bar{x})^2} \begin{pmatrix} \sum_i y_i \sum x_i^2 - n\bar{x} \sum_i x_i y_i \\ -n^2 \bar{x}\bar{y} + n \sum_i x_i y_i \end{pmatrix}$

$\beta_1 = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum (x_i - \bar{x})^2}$

$\beta_0 = \frac{\sum_i y_i \sum x_i^2 - \sum x_i \sum_i x_i y_i}{\left(\sum x_i \bar{x}\right)^2} = \bar{y} - \hat{\beta}_2 x$

$\hat{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \bar{y} - \beta_2 \bar{x} \\ \frac{\sum (x_i y - n\bar{x}\bar{x})}{\sum (x_i - \bar{x})^2} \end{pmatrix}$

(14) $\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y$

(13) $E[\hat{\beta}] = E[(X^T X)^{-1} X^T \underline{Y}] =$
$E[(X^T X)^{-1} X^T (X\beta + \varepsilon)] = E[\beta] +$
$(X^T X)^{-1} X^T E[\varepsilon] = E[\beta] = \beta$ no sesgado

(16) Dem $Var[\hat{\beta}] = \sigma^2 (X^T X)^{-1}$

$Var[\hat{\beta}] = \mathbb{E}[(\hat{\beta} - \mathbb{E}[\hat{\beta}])(\hat{\beta} - \mathbb{E}[\hat{\beta}])^T]$

$= \mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] = \mathbb{E}[((X^T X)^{-1} X^T Y - \beta)$

$((X^T X)^{-1} X^T Y - \beta)^T] =$

$\mathbb{E}[((X^T X)^{-1} X^T (X\beta + \varepsilon) - \beta) = \sigma^2 (X^T X)^{-1}$

(17) Establece bajo los supuestos de regresion lineal a el estimador de minimos cuadrados como los mejores estimadores lineales Insesgados

(18)

$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \quad \underline{\varepsilon} \sim N(0, \sigma^2 \mathbb{I}_{n\times n})$

$\mathbb{E}[\underline{Y}] = \mathbb{E}[X\beta + \underline{\varepsilon}] = X\underline{\beta} + \mathbb{E}[\underline{\varepsilon}]$

$= X\beta$

$Var[Y] = Var[X\underline{\beta} + \underline{\varepsilon}] = Var[\underline{\varepsilon}] =$

$\sigma^2 \mathbb{I}_{n\times n}$

(19) $\underline{Y} \sim N[X\underline{\beta}, \sigma^2 \mathbb{I}_{n\times n})$

(20) $L(\hat{\beta}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}}, \hat{\sigma}^2 = SSE$

$L(\beta, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}} \quad \lambda = \frac{(2\pi\hat{\sigma}^2)^{-\frac{n}{2}}}{(2\pi\hat{\sigma}^2)^{-\frac{n}{2}}}$

27) el num representa SSR suma de cuadrados explicados
el denominador SSR suma de cuadrados residuales

sigue una razon de varianza (estadistico Fisher.

28) ANOVA
el RLM se pueden utilizar los coeficientes como medias poblacionales
para estimar un estadistico que representa su igualdad o desigualdad)

29) Fo

$$f = \frac{X^2 v_1 / v_1}{X^2 v_2 / v_2} \sim F_{v_1 / v_2}$$

$$F_0 = \frac{SSR/(p-1)}{SSE/(n-p)} \sim F_{p-1,n-p}$$

34) Tabla

$SSR = \Sigma(\hat{y}_i - \bar{y})^2$ $p-1$ $\frac{SSR}{p-1}$

SSE $SSE\Sigma(y_i - \hat{y}_i)^2$ $n-p$ $\frac{SSE}{n-p}$

$SST = y^T y - n\bar{y}^2$ $n-1$