Modelos de Regresión

– Fundamentos–

Juan Carlos Martínez Ovando

juan.martinez.ovando@itam.mx

20 de octubre de 2015

1. **Modelos lineales** Fundamentos

y - variable aleatoria

F - función de distribución para y

$\mathbb{E}_F$ - esperanza, con respecto a la distribución F, i.e.

$$\mathbb{E}_F(y) = \int yF(dy)$$

x - variable explicativa para y, tal que

$$\mathbb{E}_F(y|x) = \alpha + \beta x,$$

ó

$$y = \alpha + \beta x + \varepsilon,$$

con

$$\varepsilon \sim G$$

tal que

$$\mathbb{E}_G(\varepsilon) = 0$$

1. Modelos lineales Fundamentos

¿Quién es F en la relación anterior?

$F|x$ - es en realidad la distribución de probabilidad G desplazada por $\alpha + \beta x$, i.e.

$$F(y|x) = G(y - \alpha - \beta x)$$

Entonces, en realidad estamos interesados en $F(y|x)$ y, sobre todo, en $\mathbb{E}_F(y|x)$ cuando estudiamos los modelos de regresión.

1. Modelos lineales Fundamentos

¿Qué un modelo lineal?

Re: Es un modelo que relaciona la esperanza condicional de y *con* x *en la media, i.e.*

$$\mathbb{E}_F(y|x) = g(x, \theta, \varepsilon),$$

θ es un conjunto de parámetros [e.j., $\theta = (\alpha, \beta)$]

ε es una inovación aleatoria $\varepsilon \sim G$, t.q. $\mathbb{E}_G(\varepsilon) = 0$

g es una función lineal en la inovación y los parámetros, pero multiplicativa en x respecto a los parmámetros

1. Modelos lineales Fundamentos

Ejemplos de modelos lineales:

- $g(x, \theta, \varepsilon) = \alpha + \beta x$

- $g(x, \theta, \varepsilon) = \alpha + \beta \log(x)$

- $g(x, \theta, \varepsilon) = \alpha + \beta \log(x) + \gamma x^2$

- $g(x, \theta, \varepsilon) = \alpha + \beta \log(x) + \sum_{j=1}^{J} \gamma_j \psi(x, \eta_j)$

Ejemplos de modelos no lineales:

- $g(x, \theta, \varepsilon) = \alpha \cdot x^{\beta}$
- $g(x, \theta, \varepsilon) = \alpha \cdot x^{\beta}$
- $g(x, \theta, \varepsilon) = \alpha + \log(\beta x) + \gamma x^{2}$

1. Modelos lineales Fundamentos

Regresión lineal múltiple

- $g(x, \theta, \varepsilon) = \alpha + \beta_1 x_1 + \ldots + \beta_p x_p,$

Alternativamente,

- $g(x, \theta, \varepsilon) = \theta' x,$

  $x = (1, x_1, \ldots, x_p)'$

  $\theta = (\alpha, \beta_1, \ldots, \beta_p)'$

1. Modelos lineales Fundamentos

Inferencia

- $g(x_i, \theta, \varepsilon_i) = \theta' x_i$, para $i = 1, \ldots, n$, ó

$$y_i = \theta' x_i + \varepsilon_i,$$

$x_i = (1, x_{1,i}, \ldots, x_{p,i})'$

$\theta = (\alpha, \beta_1, \ldots, \beta_p)'$

$\varepsilon_i \stackrel{iid}{\sim} G$

La inferencia sobre los modelos lineales descansa fundamentalmente sobre ideas de álgebra lineal y matricial. Así, revisémos algunos conceptos y propiedades útiles.

1. Modelos lineales Fundamentos

Algebra matricial: Notación

$\boldsymbol{A}$ - matriz con elementos (a_{ij})

$\boldsymbol{I}$ - matriz identidad, i.e. $a_{ij} = 1$, si $i = j$, y 0, e.o.c.

$\mathbf{0}$ - matriz nula, i.e. $a_{ij} = 0$

$\boldsymbol{A}'$ - transpuesta de la matriz $\boldsymbol{A}$

$\boldsymbol{A}^{-1}$ - inversa de la matriz, t.q. $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$

1. Modelos lineales Fundamentos

Algebra matricial: Operadores. Sean $\boldsymbol{A} = (a_{ij})$ y $\boldsymbol{B} = (b_{ij})$,

- $$\boldsymbol{A} + \boldsymbol{B} = (a_{ij} + b_{ij}),$$

  iff $dim(\boldsymbol{A}) = dim(\boldsymbol{B})$

- $$\boldsymbol{A} \cdot \boldsymbol{B} = \left(\sum_k a_{ik} b_{kj} \right),$$

  iff $ncol(\boldsymbol{A}) = nrow(\boldsymbol{B})$

Algebra matricial: Propiedades.

- $$\left(\boldsymbol{A}'\right)^{-1} = \left(\boldsymbol{A}^{-1}\right)',$$

  iff $\boldsymbol{A}$ es no singular

- $$(\boldsymbol{A} \cdot \boldsymbol{B})' = \boldsymbol{B}' \cdot \boldsymbol{A}'$$

- $$(\boldsymbol{A} \cdot \boldsymbol{B})^{-1} = \boldsymbol{B}^{-1} \cdot \boldsymbol{A}^{-1}$$

  iff $\boldsymbol{A}$ y $\boldsymbol{B}$ son no singulares

- $$rank(\boldsymbol{A}\boldsymbol{A}') = rank(\boldsymbol{A}'\boldsymbol{A}) = rank(\boldsymbol{A})$$

- Si $\boldsymbol{x}$ y $\boldsymbol{y}$ son dos vectores, y $\boldsymbol{A}$ es no singular, entonces

  $$\boldsymbol{x} = \boldsymbol{A} \cdot \boldsymbol{y} \quad \text{y} \quad \boldsymbol{A}^{-1} \cdot \boldsymbol{x} = \boldsymbol{y}$$

  son equivalentes.

Algebra matricial: Formas cuadráticas.

$\boldsymbol{y}$ - vector de dimensión $n \times 1$

$\boldsymbol{A}$ - matriz de dimensión $n \times n$

- La forma cuadrática de y y A es

$$\boldsymbol{y}' \cdot \boldsymbol{A} \cdot \boldsymbol{y} = \sum_i \sum_j y_i a_{ij} y_j.$$

- $$rank\left(\boldsymbol{y}' \cdot \boldsymbol{A} \cdot \boldsymbol{y}\right) = rank\left(\boldsymbol{A}\right)$$

- $\boldsymbol{y}' \cdot \boldsymbol{A} \cdot \boldsymbol{y}$ es *definida positiva* si $\boldsymbol{y}' \cdot \boldsymbol{A} \cdot \boldsymbol{y} > 0$ para todo $\boldsymbol{y}$ no nulo.

- Si $\boldsymbol{A}$ es t.q. $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}$, se dice que $\boldsymbol{A}$ es *ortogonal*.

1. Modelos lineales Fundamentos

Algebra matricial: Transformaciones y formas cuadráticas.

$\boldsymbol{y}$ - vector de dimensión $n \times 1$

$\boldsymbol{x}$ - vector de dimensión $p \times 1$

$\boldsymbol{P}$ - matriz de dimensión $n \times p$

- t.q.

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{x}.$$

Entonces,

- $$\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} = \boldsymbol{x}'\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x},$$

  con $\boldsymbol{B} = \boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}$, y $\boldsymbol{B}$ es no singular.

Teoremas útiles:

- Si $\boldsymbol{P}$ es no singular y $\boldsymbol{A}$ es definida positiva, entonces $\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}$ es definida positiva.

- $\boldsymbol{A}$, matriz simétrica, es definida positiva iff existe $\boldsymbol{P}$ no singular, t.q.

  $$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{P}'.$$

Determinantes:

$\boldsymbol{A}$ - matriz simétrica de dimensión $p \times p$

- **Raíz característica:** λ escalar, tal que

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x},$$

para todo vector $\boldsymbol{x}$.

$\boldsymbol{x}$ - vector $(p \times 1)$ característico asociado a λ.

- **Polinomio característico:** Se define como,

$$|\boldsymbol{A} - \lambda \boldsymbol{I}|,$$

que es de orden p.

- *Las raices del polinomio caractaerístico corresponden a las raices características de* $\boldsymbol{A}$*.*

Determinantes (algunos teoremas)

- $rank(\boldsymbol{A}) =$ número de raices unitarias distintas de 0

- Si $\boldsymbol{C}$ es una matriz ortogonal, entonces las raices características de $\boldsymbol{A}$ y $\boldsymbol{CAC'}$

- Si $\boldsymbol{A}$ es simétrica, entonces sus raices características son reales.

- *Si $\boldsymbol{A}$ es simétrica, existe una matriz ortogonal $\boldsymbol{C}$, t.q.*

$$\boldsymbol{C'AC} = \boldsymbol{D},$$

donde $\boldsymbol{D}$ es una matriz diagonal cuyas entradas con las raices características de $\boldsymbol{A}$.

- Si $\boldsymbol{C}$ es una matriz ortogonal, entonces

$$|\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}| = |\boldsymbol{A}|.$$

- Si $\boldsymbol{A}$ es singular, entonces $|\boldsymbol{A}| = 0$.

- Si $\boldsymbol{C}$ es ortogonal, entonces $|\boldsymbol{C}| = \pm 1$.

1. **Modelos lineales** Fundamentos

Traza: La traza de una matriz se define como $tr(\boldsymbol{A}) = \sum_{j=1}^{p} a_{jj}$.

- $tr(\boldsymbol{I}) = p$.
- $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$.
- $tr(\boldsymbol{ABC}) = tr(\boldsymbol{CAB}) = tr(\boldsymbol{BCA})$.
- Si $\boldsymbol{C}$ es ortogonal, entonces $tr(\boldsymbol{C}'\boldsymbol{AC}) = tr(\boldsymbol{A})$.

1. Modelos lineales Fundamentos

Lo que sigue:

– *Encontrar distribuciones de probabilidad asociadas a formas cuadráticas de matrices.*

– *Deducir derivadas de formas cuadráticas (para la estimación por máxima verosimilitud).*

1. Modelos lineales Inferencia

$\{y_i, \boldsymbol{x}_i\}$ - colección de n variables y covariables

y_i - variable de respuesta de la i ésima observación

$\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})$ conjunto de p covariables de la i-ésima observación

Relación lineal:

$$\mathbb{E}(y_i) = \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$$

Alternativamente:

$$\begin{aligned} y_i &= \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \\ \varepsilon_i &\overset{iid}{\sim} G, \end{aligned}$$

t.q.

$$\mathbb{E}_G(\varepsilon_i) = 0.$$

Representación matricial:

$$\boldsymbol{y} \;=\; \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

donde

- $\boldsymbol{y} = (y_1, \ldots, y_n)'$
- $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$
- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$

1. Modelos lineales Inferencia

Pregunta: ?́Qué podemos decir acerca de β?

Alternativas:

- Método de momentos*

- Estimación por máxima verosimilitud*

- Mínimos cuadrados*

- Inferencia bayesiana

Método de momentos (MdeM). Supuestos:

- Las observaciones son iid
- No forma distribucional es supuesta para F, sólo los momentos poblacionales (o del modelo),

$$M(k) = \mathbb{E}_F(y^k|\boldsymbol{x}),$$

son requeridos.

Recodemos que $F(y|\boldsymbol{x})$ es en realidad lo que deseamos encontrar.

Procedimiento:

Supongamos que $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ caracterizan a $F(y_i|\boldsymbol{x}_i, \boldsymbol{\theta})$.

- Dada la muestra con n observaciones, el MdeM consiste en empatar los momentos muestrales con los momentos poblacionales (del modelo), los cuales dependen de $\boldsymbol{\theta}$, y resolver las ecuaciones resultantes para $\boldsymbol{\theta}$ (simultáneamente) .

1. Modelos lineales Inferencia

Máxima verosimilitud (MV). Supuestos:

- Las observaciones son iid
- Se supone una forma distribucional para F –usualmente gaussiana– siendo $f(y|\boldsymbol{x},\boldsymbol{\theta})$ la densidad.
- Función de verosimilitud,

$$lik(\boldsymbol{\theta}, \text{datos}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{x}_i, \boldsymbol{\theta})$$

Procedimiento:

- El estimador de $\boldsymbol{\theta}$, denotado por $\widehat{\boldsymbol{\theta}}$, se obtiene como

$$\widehat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} lik(\boldsymbol{\theta}, \text{datos}).$$

- La solución, calcular y resolver para $\boldsymbol{\theta}$,

$$\frac{\partial}{\partial \boldsymbol{\theta}} lik(\boldsymbol{\theta}, \text{datos}) = \mathbf{0}.$$

Mínios cuadrados (MC). Supuestos:

- Las observaciones son iid

- No se supone una forma distribucional para F, solo se require la restricción $\mathbb{E}_G(\varepsilon_i) = 0$.

- Función de cuadrados de los residuos,
$$rss(\boldsymbol{\beta}, \text{datos}) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}})^2,$$
para el estiador $\widehat{\boldsymbol{\beta}}$ objetivo.

- ?́Qué pasa con el estimador de dispersión de ε?

Procedimiento:

- El estimador de $\boldsymbol{\beta}$, denotado por $\widehat{\boldsymbol{\beta}}$, se obtiene como

$$\widehat{\boldsymbol{\beta}} = \text{argmin}_{\boldsymbol{\beta}\in} rss(\boldsymbol{\beta}, \text{datos}).$$

Desviaciones.

- Todos los procedimientos se basan, de alguna forma, en la noción de *desviaciones*. Desviaciones de los datos de respuesta observados (y_i) y de los datos "ajustados", ($\widehat{y}_i$), donde

$$\begin{aligned} \widehat{y}_i &= \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}}. \\ &= \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_p x_{ip}. \end{aligned}$$

1. Modelos lineales Inferencia

Desviaciones.

Desviaciones.jpg

Desviaciones.

- Debemos considerar que la relación entre y y $\boldsymbol{x}$ (inclusive en el caso en que sólo tengamos una sola covariable x) no es simétrica.

- Los residuos cuya suma de cuadrados deseamos minimizar se refieren a las desviaciones verticales.

Supuestos adicionales.

- Los datos observados son observados aleatoriamente.
- Las columnas en $\boldsymbol{X}$ no son colineales (o al menos, no estrictamente colineales, admitiendo cierto grado de colinealidad.)
- Las observaciones en (y_i) son obtenidas con precisión de medición.
- Los residuos (r_i) tienen varianza constante en i y una distribución normal (o gaussiana).

1. **Modelos lineales** Predicción

- Los modelos estadísticos se emplean, comúnmente para predicción o pronóstico.

- Aun cuando β sea el interés final, la capacidad predictiva del modelo le da validez a los resultados.

- Predicción en modelos lineales es relativamente simple.

1. Modelos lineales Predicción

Procedimiento predictivo:

1. Definimos el modelo

$$y = \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon.$$

2. Estimamos $\widehat{\boldsymbol{\beta}}$ y $\widehat{\sigma}$.

3. Obtenemos el modelo ajustado

$$\widehat{y} = \boldsymbol{x}'\widehat{\boldsymbol{\beta}}.$$

4. Predecimos un valor futuro

$$\widehat{y}_f = \boldsymbol{x}'_f \widehat{\boldsymbol{\beta}},$$

con base en covariables futuras $\boldsymbol{x}_f$.

Precisión de las predicciones:

- Intervalos de predicción para

$$\widehat{y}_f \in \boldsymbol{x}'_f\widehat{\boldsymbol{\beta}} \pm \rho\widehat{\sigma},$$

  con base en covariables futuras $\boldsymbol{x}_f$ y en $\widehat{\sigma}$. El parámetro ρ se fija para un cierto grado de credibilidad.

- Más aun, podemos recuperar la distribución predictiva

$$F(y_f|\boldsymbol{x}_f) = N(y_f|\boldsymbol{x}'_f\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2).$$