COVID-19 y Modelos No lineales

Sarai E Gómez Ibarra

Universidad Autónoma de Nuevo León

11 de Agosto del 2020

Contenido

Análisis de regresión lineal simple

$$Y_i = \beta_o + \beta_1 x_i + \varepsilon_i \tag{1}$$

El Modelo mencionado anteriormente cumple las siguientes condiciones:

Análisis de regresión lineal simple

$$Y_i = \beta_o + \beta_1 x_i + \varepsilon_i \quad (1)$$

El Modelo mencionado anteriormente cumple las siguientes condiciones:

- ε_i se distribuye Normal$(0, \sigma^2)$

Análisis de regresión lineal simple

$$Y_i = \beta_o + \beta_1 x_i + \varepsilon_i \tag{1}$$

El Modelo mencionado anteriormente cumple las siguientes condiciones:

- ε_i se distribuye Normal$(0, \sigma^2)$
- σ^2 es constante

Análisis de regresión lineal simple

$$Y_i = \beta_o + \beta_1 x_i + \varepsilon_i \tag{1}$$

El Modelo mencionado anteriormente cumple las siguientes condiciones:

- ε_i se distribuye Normal$(0, \sigma^2)$
- σ^2 es constante
- ε_i no esta correlacionado

Estimación de los parámetros

Para ajustar nuestros datos a el modelo (1), tenemos que realizar la estimación de los parámetros desconocidos β_o y β_1. Para esto existen dos métodos muy conocidos

Estimación de los parámetros

Para ajustar nuestros datos a el modelo (1), tenemos que realizar la estimación de los parámetros desconocidos β_o y β_1. Para esto existen dos métodos muy conocidos

- Mínimos Cuadrados
  El objetivo es minimizar la suma de cuadrados de los residuos, i.e.

$$\min_{\beta_o, \beta_1} \sum_{i=1}^{n} (Y_i - (\beta_o + \beta_1 x_i))^2$$

- Máxima Verosimilitud
  El método consiste en encontrar los valores de β_o y β_1 que maximizan $L(\beta_o, \beta_1)$, donde $L(\beta_o, \beta_1)$ esta dada por:

$$L(\beta_o, \beta_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(Y_i - (\beta_o + \beta_1 x_i))^2}{2\sigma^2}}$$

- Máxima Verosimilitud
  El método consiste en encontrar los valores de β_o y β_1 que maximizan $L(\beta_o, \beta_1)$, donde $L(\beta_o, \beta_1)$ esta dada por:

$$L(\beta_o, \beta_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(Y_i-(\beta_o+\beta_1 x_i))^2}{2\sigma^2}}$$

Para el Modelo (1) se tiene que ambos métodos nos arrojan los mismos estimadores de β_o y β_1.

Contenido

Modelo Gompertz

$$Y_i = \alpha exp(-\beta e^{-kx_i}) + \varepsilon_i \tag{2}$$

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

Modelo Gompertz

$$Y_i = \alpha exp(-\beta e^{-kx_i}) + \varepsilon_i \tag{2}$$

donde $\varepsilon_i \sim N(0, \sigma^2)$
La fomra parametrica del Modelo esta dada por

$$N(t) = \alpha exp(-\beta e^{-kx_i}) \tag{3}$$

Modelo Gompertz

$$Y_i = \alpha exp(-\beta e^{-kx_i}) + \varepsilon_i \tag{2}$$

donde $\varepsilon_i \sim N(0, \sigma^2)$
La fomra parametrica del Modelo esta dada por

$$N(t) = \alpha exp(-\beta e^{-kx_i}) \tag{3}$$

donde:

- $t = 0$ es el número inicial de casos
- $N(t)$ es el número acumulados de casos confirmados
- La asíntota α corresponde al número total de casos al final de la epidemia

Contenido

Método de mínimos cuadrados ordinarios

Debido a que el problema es estimar el vector desconosido θ, lo mas natural es elegir el valor de θ que minimiza la distancia entre los valores de $f(x_i, \theta)$ y las observaciones Y. Uno puede elegir el valor que minimiza la suma de cuadrados, definida como:

$$C(\theta) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - f(x_i, \theta))^2$$

Método de mínimos cuadrados ordinarios

Debido a que el problema es estimar el vector desconosido θ, lo mas natural es elegir el valor de θ que minimiza la distancia entre los valores de $f(x_i, \theta)$ y las observaciones Y. Uno puede elegir el valor que minimiza la suma de cuadrados, definida como:

$$C(\theta) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - f(x_i, \theta))^2$$

Sea $\hat{\theta}$ el estimador de minimos cuadrados de θ. Si asumimos que $Var(\varepsilon_{ij}) = \sigma^2$, una estimacion de σ^2 es:

$$\hat{\sigma^2} = \frac{C(\hat{\theta})}{n} \tag{4}$$

Bajo el supuesto de la ec. (), $\hat{\theta}$ es tambien la solucion del conjjunto de ecuaciones p:

Bajo el supuesto de la ec. (), $\hat{\theta}$ es tambien la solucion del conjjunto de ecuaciones p:

$$\sum_{i=1}^{k} \frac{\partial f}{\partial \theta_a}(x_i, \theta) \sum_{j=1}^{n_i} (Y_{ij} - f(x_i, \theta)) = 0 \tag{5}$$

Método de Quasi-likelihood

Considere el modelo de regresión no lineal

$$Y_i = f(x_i, \theta) + \varepsilon_i \tag{6}$$

$$Var(\varepsilon_i) = \sigma^2 g(x_i, \theta, \tau), E(\varepsilon_i) = 0$$

donde ε_i es una variable aleatoria independiente.

Método de Quasi-likelihood

Considere el modelo de regresión no lineal

$$Y_i = f(x_i, \theta) + \varepsilon_i \tag{6}$$

$$Var(\varepsilon_i) = \sigma^2 g(x_i, \theta, \tau), E(\varepsilon_i) = 0$$

donde ε_i es una variable aleatoria independiente.
la función de MLE logarítmica esta dada como:

$$\log L = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}[\log(\sigma^2 g(x_i, \theta, \tau)) + \frac{(Y_i - f(x_i, \theta))^2}{\sigma^2 g(x_i, \theta, \tau)}]$$

Método de Quasi-likelihood

Considere el modelo de regresión no lineal

$$Y_i = f(x_i, \theta) + \varepsilon_i \tag{6}$$

$$Var(\varepsilon_i) = \sigma^2 g(x_i, \theta, \tau), E(\varepsilon_i) = 0$$

donde ε_i es una variable aleatoria independiente.
la función de MLE logarítmica esta dada como:

$$\log L = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}[\log(\sigma^2 g(x_i, \theta, \tau)) + \frac{(Y_i - f(x_i, \theta))^2}{\sigma^2 g(x_i, \theta, \tau)}]$$

El estimador MLE se basa en la suposición de que los errores ε_i se distribuye como variables gaussianas. Sin embargo no siempre esto se cumple, o no es apropiado suponerlo; por tal motivo podemos utilizar el método de quasi-likelihood el cual se basa en el conocimiento de las funciones de regresión y la varianza.

Consideremos que nuestro modelo esta dado de la forma ec. (), sea p la dimensión de θ y q la dimensión de τ, el modelo depende de los parámetros $p + q + 1$.
Tenemos que realizar las estimaciones de θ, *tau* y *sigma*. Las formulas o ecuaciones de quasi-likelihood son:

Consideremos que nuestro modelo esta dado de la forma ec. (), sea p la dimensión de θ y q la dimensión de τ, el modelo depende de los parámetros $p+q+1$.
Tenemos que realizar las estimaciones de θ, *tau* y *sigma*. Las formulas o ecuaciones de quasi-likelihood son:

$$U_a(\theta,\tau)=\sum_{i=1}^{n}\frac{\partial f}{\partial\theta_a}(x_i,\theta)\frac{Y_i-f(x_i,\theta)}{g(x_i,\theta,\tau)} \tag{7}$$

$$U_{b+q}(\theta,\tau)=\sum_{i=1}^{n}\frac{\partial g}{\partial\theta_a}(x_i,\theta,\tau)\frac{Y_i-f(x_i,\theta)-\sigma^2 g(x_i,\theta,\tau}{g^2(x_i,\theta,\tau)} \tag{8}$$

Los estimadores de quasi-likelihood $\hat{\theta}_{QL}, \hat{\tau}_{QL}$ y $\hat{\sigma}^2_{QL}$, satisfacen:

Los estimadores de quasi-likelihood $\hat{\theta}_{QL}, \hat{\tau}_{QL}$ y $\hat{\sigma}^2_{QL}$, satisfacen:

$$U_a(\hat{\theta}_{QL}, \hat{\tau}_{QL}) = 0$$

$$U_{b+q}(\hat{\theta}_{QL}, \hat{\tau}_{QL}) = 0$$

$$\hat{\sigma}^2_{QL} = \frac{1}{n} \sum_{i=1}^{n} \frac{(Y_i - f(x_i, \hat{\theta}_{QL}))^2}{g(x_i, \hat{\theta}_{QL}, \hat{\tau}_{QL})}$$

para $a = 1, ..., p$ y $b = 1, ...b$