REGRESIÓN MÚLTIPLE: INTERVALOS DE CONFIANZA

1.4. Intervalos de confianza para los coeficentes de regresión

Sabemos que $Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$

entonces $Y \sim N(X\beta, \sigma^2 I)$ por lo tanto

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

esto implica que la distribución de:

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 C_{jj}) \quad j=0,1,2,\ldots,k$$

C_{jj} es el j-ésimo elemento de la diagonal de $(X'X)^{-1}$

Si la varianza es conocida tenemos:

$$Z = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 C_{jj}}} \sim N(0,1)$$

Si la varianza no es conocida tenemos que la estadística T tiene la distribución t, con n-p grados de libertad, esto es:

$$T = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{ii}}} \quad j = 1,2,\ldots,k \quad \sim t(n-p)$$

Construyendo el estadístico de prueba tenemos:

$$T = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma 2 C_{jj}}} \Big/ (SCE/\sigma^2)/(n-p) \quad \sim t_{(n-p)}$$

Así que, podemos afirmar que:

$$P\left(\frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}^2 C_{ii}}} \leq t_{(\alpha/2,\ n-p)} \right) = 1-\alpha$$

Por lo tanto un intervalo de confianza de 100(1-α)% para los coeficientes de regresión esta dado por:

$$\text{IC}(\beta) = \hat{\beta}_j \pm t_{\left(\frac{\alpha}{2},n-p\right)}\sqrt{\hat{\sigma}^2 C_{jj}}$$

6. Estimación del Intervalo de Confianza de la respuesta media

La estimación del valor medio de la variable respuesta para un conjunto de valores particulares de las variables regresoras para lo cual definimos:
$\boldsymbol{x}_0' = (1, x_{01}, \ldots\ldots\ldots\ldots\ldots, x_{0k})$

El valor ajustado en este punto es

$$\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$$

que es un estimador insesgado de $E(y/x_0)$ porque $E[\hat{y}_0] = \boldsymbol{x}_0'\beta = E(y/x_0)$

Y la varianza de $\hat{y}_0$ es: $\text{var}(\hat{y}_0) = \sigma^2 X'_0 (X'X)^{-1} X'_0$

Por consiguiente, un intervalo de confianza de 100(1-α)% de la respuesta media en el punto $(\boldsymbol{x}_0' = (x_{01}, \ldots\ldots\ldots\ldots\ldots, x_{0k})$ es:

$$\hat{y}_0 - t_{\left(\frac{\alpha}{2},n-p\right)}\hat{\sigma}\sqrt{\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0} < E[y/\boldsymbol{x}_0] < \hat{y}_0 + t_{\left(\frac{\alpha}{2},n-p\right)}\hat{\sigma}\sqrt{\boldsymbol{x}_0'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0}$$

1.5. Predicción de Nuevas Observaciones

Para un valor particular no observado "y_0", su predicción se obtiene mediante la media de su distribución condicional "$\hat{y}_0$";
Para predecir observaciones futuras de y que correspondan a determinados valores de las variables regresoras por ejemplo $x_{01}, \ldots\ldots\ldots\ldots\ldots, x_{0k}$. entonces el estimado puntual de la observación futura y0 en el punto $x_{01}, \ldots\ldots\ldots\ldots\ldots, x_{0k}$ es:

$$\hat{y}_0 = \boldsymbol{x}_0'\hat{\boldsymbol{\beta}}$$

Un intervalo de confianza de 100(1-α)% para una futura observación es:

$$\hat{y}_0 - t_{\left(\frac{\alpha}{2},n-p\right)}\hat{\sigma}\sqrt{1+\boldsymbol{x}_0^{'}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0} < y_0 < \hat{y}_0 + t_{\left(\frac{\alpha}{2},n-p\right)}\hat{\sigma}\sqrt{1+\boldsymbol{x}_0^{'}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_0}$$

.

Intervalos de Confianza para Combinaciones Lineales de los Parámetros q´β

Dado que q´β es una combinación lineal de β
Además

Sabemos que $Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$

Y además que $Y \sim N(X\beta, \sigma^2 I)$ por lo tanto

Como combinaciones lineales de normales tenemos:

$$U\beta^{\wedge} \sim N(q'\beta,\ q'\sigma^2(X'X)^{-1}q)$$

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

estandarizando tenemos

$$Z = (q'\beta^{\wedge} - q'\beta)/\sqrt{\hat{\sigma}^2 q'(\boldsymbol{X}'\boldsymbol{X})^{-1}q} \sim N(0,1)$$

Por otro lado sabemos que

$$SCE/\ \sigma^2 \sim \square_{(n-p)}$$

Con ello podemos construir la estadística T para evaluar el intervalo

$$T = \frac{(q'\beta^{\wedge} - q'\beta)/\sqrt{\hat{\sigma}^2 q'(\boldsymbol{X}'\boldsymbol{X})^{-1}q} \quad \sim N(0,1)}{SCE/\ \sigma^2 \quad \sim \square_{(n-p)}}$$

$$T = \frac{q'\beta^{\wedge} - q'\beta}{\sqrt{\hat{\sigma}^2 q'(\boldsymbol{X}'\boldsymbol{X})^{-1}q}} \sim t_{(n-p)}$$

Un intervalo de confianza de 100(1-α)% para $q'\beta$ es:

$$IC_{q'\beta} = q'\hat{\beta} \pm t_{\left(1-\frac{\alpha}{2},\ n-p\right)} \sqrt{\hat{\sigma}^2 q'(X'X)^{-1}q} \sim N(0,1)$$

INTERVALOS SIMULTANEOS DE CONFIANZA

A menudo necesitamos formar intervalos de confianza para los estimadores de los coeficientes de regresión de un modelo dado y además necesitamos hacer predicciones de varios estimadores a la vez; esto es, formar intervalos de confianza para dos o más estimadores con una confianza establecida de 1−α

Definiremos la región de confianza conjunta para los parámetros del modelo:

$$Y = X\beta + \varepsilon,$$

lo que implica

$$\left\{\left[(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)\right] / \sigma^2\right\} \sim \square_{(p)}$$

Es decir:

$$P\left\{\left[(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)\right] / p x CME \leq \sim F_{(\alpha,p,n-p)}\right\} = 1-\alpha$$

De donde obtenemos una región de confianza conjunta al 100(1-α)% para todos los parámetros en β. La región contendrá aquellos valores de β tales que:

$$\left\{\left[(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)\right] / p \ x\ SCE\right\} \leq p\ x\ CME\ F_{(\alpha,p,n-p)}$$

$$\left[(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)\right] \leq p\ x\ S^2 F_{(\alpha,p,n-p)}$$

Tarea: Desarrollar y presentar el método Bonferroni y su ejemplo por grupos.