Econometría Básica

Capítulo 04: Modelo de Regresión Lineal Múltiple-Inferencia

José Valderrama & Freddy Rojas
jtvalderrama@gmail.com & frojasca@gmail.com
Universidad Católica Santo Toribio de Mogrovejo

Septiembre de 2021

Contenido

USAT Universidad Católica Santo Toribio de Mogrovejo

Contenido

USAT Universidad Católica Santo Toribio de Mogrovejo

Supuesto para la inferencia

- Suposición RLM6: Normalidad. Con el fin de probar hipótesis se necesita agregar un supuesto adicional a los cinco supuestos ya vistos: $u \sim N(o, \sigma^2)$ El error poblacional se distribuye como un normal con media cero y varianza σ^2
- Lo anterior implica que

$$y/x \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k, \sigma^2)$$

- También $\hat{\beta}_j = N(\beta_j, Var(\hat{\beta}_j))$, por lo que estandarizando se tiene:

$$\frac{\hat{\beta}_j - \beta_j}{Es(\hat{\beta}_j)} \sim N(0, 1)$$

Supuesto para la inferencia

- Suposición RLM6: Normalidad. Con el fin de probar hipótesis se necesita agregar un supuesto adicional a los cinco supuestos ya vistos: $u \sim N(o, \sigma^2)$ El error poblacional se distribuye como un normal con media cero y varianza σ^2
- Lo anterior implica que

$$y/x \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k, \sigma^2)$$

- También $\hat{\beta}_j = N(\beta_j, Var(\hat{\beta}_j))$, por lo que estandarizando se tiene:

$$\frac{\hat{\beta}_j - \beta_j}{Es(\hat{\beta}_j)} \sim N(0, 1)$$

Supuesto para la inferencia

- Suposición RLM6: Normalidad. Con el fin de probar hipótesis se necesita agregar un supuesto adicional a los cinco supuestos ya vistos: $u \sim N(o, \sigma^2)$ El error poblacional se distribuye como un normal con media cero y varianza σ^2
- Lo anterior implica que

$$y/x \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k, \sigma^2)$$

- También $\hat{\beta}_j = N(\beta_j, Var(\hat{\beta}_j))$, por lo que estandarizando se tiene:

$$\frac{\hat{\beta}_j - \beta_j}{Es(\hat{\beta}_j)} \sim N(0, 1)$$

Supuesto para la inferencia

- Suposición RLM6: Normalidad. Con el fin de probar hipótesis se necesita agregar un supuesto adicional a los cinco supuestos ya vistos: $u \sim N(o, \sigma^2)$ El error poblacional se distribuye como un normal con media cero y varianza σ^2
- Lo anterior implica que

$$y/x \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k, \sigma^2)$$

- También $\hat{\beta}_j = N(\beta_j, Var(\hat{\beta}_j))$, por lo que estandarizando se tiene:

$$\frac{\hat{\beta}_j - \beta_j}{Es(\hat{\beta}_j)} \sim N(0,1)$$

Supuesto para la inferencia

- Suposición RLM6: Normalidad. Con el fin de probar hipótesis se necesita agregar un supuesto adicional a los cinco supuestos ya vistos: $u \sim N(o, \sigma^2)$ El error poblacional se distribuye como un normal con media cero y varianza σ^2
- Lo anterior implica que

$$y/x \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k, \sigma^2)$$

- También $\hat{\beta}_j = N(\beta_j, Var(\hat{\beta}_j))$, por lo que estandarizando se tiene:

$$\frac{\hat{\beta}_j - \beta_j}{Es(\hat{\beta}_j)} \sim N(0, 1)$$

Contenido

Contenido

Inferencia 1: prueba de hipótesis para una única combinación lineal de coeficientes

Consideremos nuevamente el siguiente modelo (o PGD) con una constante y más de un regresor

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

La inferencia es el proceso de realizar una prueba estadística; nos interesa la importancia de los efectos sobre la variable dependiente. En particular, el coeficiente de interés es el efecto de X_2 sobre Y. El efecto de X_1 e Y es necesario porque eso nos permite controlar otros efectos. Por ejemplo

$$H_0 : \beta_2 = 0$$
$$H_a : \beta_2 \neq 0$$

por tanto, podemos centrarnos sólo en algún parámetro de interés, en este caso β_2. *[Pregunta: ¿Por qué deberíamos mantener en el análisis la variable X_1?. Son controles].*

Antes de la inferencia

Cuando configuramos un modelo, tenemos los siguientes tipos de regresores:

- Regresores o variables de interés (relacionadas con la política).
- Variables de control.

Si alguna variable no se puede clasificar en las categorías anteriores, no la considere en su análisis.

¿Z o t-student?

Ya hemos discutido este tema con respecto a Z o t-student. Básicamente, debemos elegir la estadística según el número de observaciones bajo análisis. Una regla simple es: una muestra pequeña tiene menos de 150 observaciones; en ese caso nuestro análisis considera el t-student. La estadística es:

$$t_{cal} = \frac{\text{Estimador - Prueba de hipotesis}}{\text{Error estandar del estimador}}$$

por ejemplo.

$$t_{cal} = \frac{\widehat{\beta}_2 - 0}{S.E(\widehat{\beta}_2)}$$

luego contrastamos esa t *realmente calculada"* con valores críticos o simplemente calculamos el *p-value* asociado a esa prueba; es decir, $2Pr(t > |t_{cal}|)$

T-STUDENT

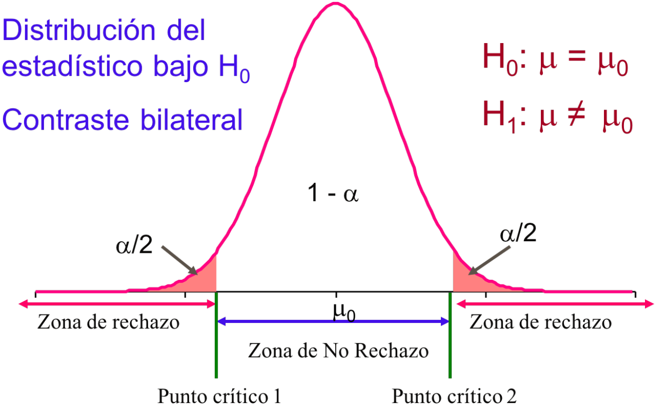
- Aunque el $\widehat{\beta}_j$ teóricamente sigue una distribución normal, al depender de σ^2 (parámetro no conocido) hace que el uso del estimador $\widehat{\sigma}^2 = SRC/(n-k-1)$ cambie la distribución del parámetro.
- Conociendo la distribución se pueden probar diversas hipótesis como $H_0 : \beta_j = 0$ en cuyo caso, si se prueba esa hipótesis la variable x_j no tiene efecto sobre y controlando por las otras $x's$
- El t estadístico para dicha hipótesis sería $t_{\beta_j} \equiv \frac{\widehat{\beta}_j - 0}{Se(\widehat{\beta}_j)}$ que sigue una distribución T de student con $n-k-1$ grados de libertad.
- Aunque con muestras grandes la T de student se comporta como una normal. Específicamente cuando $n-k-1 > 120$ se puede usar una normal para los valores críticos.

T-STUDENT

- Además de la hipótesis nula (H_0) para completar la prueba estadística se necesita de una hipótesis alternativa (H_1) y de un nivel de significancia.
- H_1 puede ser de unilateral o de un lado o bilateral o de dos lados
- Unilateral:$\beta_j > 0$ o $\beta_j < 0$
- Bilateral:$\beta_j \neq 0$
- El nivel de significancia es la probabilidad de cometer un error de tipo I, los valores frecuentemente elegidos son de 1, 5 y 10 por ciento.

T-STUDENT

$H_0 : \beta_j = 0$ y $H_1 : \beta_j \neq 0$
Si $t_{\beta_j} > |Crit|$, entonces se rechaza la hipótesis nula H_0. Donde $Crit$ se obtiene de tablas estadísticas necesitando para ello el nivel de significancia a utilizar (α) y el número de grados de libertad $n - k - 1$



T-STUDENT

- Si se rechaza la H_0 lo que típicamente se dice es que la variable x_j es estadísticamente significativa al nivel de significancia de α
- Si no se rechaza la H_0 lo que típicamente se dice es que la variable x_j es estadísticamente no significativa al nivel de significancia de α
- Aunque se vio el caso de $H_0 : \beta_j = 0$ en general se puede probar la hipótesis: $H_0 : \beta_j = a_j$, en cuyo caso el t-estadístico quedaría definido por: $t = \frac{\hat{\beta}_j - a_j}{Se(\hat{\beta}_j)}$

T-STUDENT: EJEMPLO

Podemos incluir en nuestro análisis de inferencia alguna combinación lineal de parámetros. Por ejemplo, consideremos una función de producción para la economía de EE. UU.

$$Q = e^{\alpha_0} K^{\alpha_k} L^{\alpha_l}$$

donde $\widehat{\alpha}_k = 0{,}632$ y $\widehat{\alpha}_l = 0{,}452$. Adicionalmente, tenemos la siguiente información: $cov(\widehat{\alpha}_k, \widehat{\alpha}_l) = 0{,}055$, $std(\widehat{\alpha}_k) = 0{,}257$ y $std(\widehat{\alpha}_l) = 0{,}219$. Evaluamos is $\widehat{\alpha}_k$ son idñenticos en términos estadísticos

$$t_{cal} = \frac{\widehat{\alpha}_k - \widehat{\alpha}_l}{S.E(\widehat{\alpha}_k - \widehat{\alpha}_l)}$$

o si existe evidencia de retornos constantes a escala

$$t_{cal} = \frac{\widehat{\alpha}_k + \widehat{\alpha}_l - 1}{S.E(\widehat{\alpha}_k + \widehat{\alpha}_l - 1)}$$

luego contrastamos esa t *realmente calculada"* con valores críticos o simplemente calculamos el *p-value* asociado a esa prueba; es decir, $2Pr(t > |t_{cal}|)$

T-STUDENT: RESUMEN DE PASOS

- Calcular las estimaciones de MCO.
- Calcular el error estándar de las combinaciones lineales de parámetros.
- Calcular la estadística t-student o t "*realmente calculada*".
- Calcular el p-value o simplemente concluya utilizando valores críticos.
- La interpretación es muy importante, ¡no te pierdas ese paso !.

Contenido

P-Value

- Además de la prueba T y el intervalo de confianza existe un tercer enfoque para evaluar la significancia individual de los parámetros
- El P-value o valor P es el nivel exacto de significancia y esta definido como el nivel más bajo de significancia al cual puede rechazarse una hipótesis nula
- Como regla práctica si el P-value es menor a 0.05 entonces se dice que se rechaza la nula a un nivel de significancia del 5 por ciento
- La ventaja del uso de este indicador es que no requiere observar ninguna tabla estadística, sólo comparar el P-value con el nivel de significancia deseado.

Contenido

Intervalos de confianza

- Otra forma de de probar una hipótesis es construyendo intervalos de confianza usando el mismo valor crítico empleado para la prueba
- $\hat{\beta}_j \pm Crit * Se(\hat{\beta}_j)$,
- Donde $Crit$ es el percentil $(1 - \frac{\alpha}{2})$ en una distribución t_{n-k-1} bilateral

Un intertvalo de confianza

- Se construye un intervalo de confianza alrededor de la estimación $\widehat{\beta}$ y ese conjunto indica todos los valores posibles donde se encuentra el valor real de la media poblacional o del parámetro (o estimación insesgada); por supuesto con alguna probabilidad de ocurrencia.
- Las probabilidades estándar son 99, 95 y 90 %.
- Un intervalo de confianza no significa precisión, significa confianza donde se encuentra el verdadero parámetro.
- Un enunciado "*Un 95 % de intervalo de confianza bilateral para β es un intervalo construido de modo que contenga el valor verdadero de β en el 95 % de todas las muestras aleatorias posibles*".

Una notación alternativa para intervalos de confianza

Aquí una alternativa y asumiendo normalidad

$$90\,\% \text{ I.C. para} \beta_2 = [\widehat{\beta}_2 \pm t_{gl,0,10/2} \cdot \widehat{\sigma}_{\widehat{\beta}_2}] \equiv [\overline{\beta_2}; \underline{\beta_2}]_{90\,\%}$$

$$95\,\% \text{ I.C. para} \beta_2 = [\widehat{\beta}_2 \pm t_{gl,0,05/2} \cdot \widehat{\sigma}_{\widehat{\beta}_2}] \equiv [\overline{\beta_2}; \underline{\beta_2}]_{95\,\%}$$

$$99\,\% \text{ I.C. para} \beta_2 = [\widehat{\beta}_2 \pm t_{gl,0,01/2} \cdot \widehat{\sigma}_{\widehat{\beta}_2}] \equiv [\overline{\beta_2}; \underline{\beta_2}]_{99\,\%}$$

Consulte aquí las tablas estadísticas. Tenga en cuenta que necesitamos tener los grados de libertad para el cálculo de los intervalos de confianza $(gl = N - k)$

Contenido

Medidas de Ajuste

Surgen algunas preguntas sobre el rendimiento del modelo.

- ¿Qué tan bien esa línea de regresión describe los datos?
- ¿Los regresores explican mucha o poca variación en la variable dependiente?
- ¿Están las observaciones dispersas o lejos de la línea de regresión?

El R^2 y el error estándar de la regresión miden qué tan bien la línea de regresión MCO ... coincide con los datos. Estudiante: ¿Qué es el R^2?

- el R^2 mide la fracción de la varianza de y que se explica por los regresores
- $0 < R^2 < 1$

Medidas de Ajuste

Consideremos el siguiente modelo

$$Y_i = \widehat{Y}_i + \widehat{u}_i$$

reexpresamos arriba en términos de varianza

$$Var(Y_i) = Var(\widehat{Y}_i) + Var(\widehat{u}_i)$$

equivalentemente

$$\frac{Var(\widehat{Y}_i)}{Var(Y_i)} + \frac{Var(\widehat{u})}{Var(Y_i)} = 1$$

o

$$\frac{Var(\widehat{Y}_i)}{Var(Y_i)} = 1 - \frac{Var(\widehat{u}_i)}{Var(Y_i)}$$

Medidas de Ajuste

Por lo tanto

$$R^2 = \frac{Var(\widehat{u}_i)}{Var(Y_i)} = 1 - \frac{SCR}{SCT}$$

Ahora, necestimos definir la **Suma Cuadrado de la Regresión** (SCR) y **Suma Cuadrada del Total** (SCT).

$$SCT = \sum_{i=1}^{n} (Y_I - \overline{Y}_i)^2 \tag{SCT}$$

$$SCR = \sum_{i^1}^{n} \widehat{\mu}_i^2 \tag{SCR}$$

Por lo tanto, si la varianza del error es igual a la varianza de Y, eso significa que nuestro modelo no tiene poder de predicción. Eso implica $R^2 = 0$. Si nuestro modelo puede explicar la mayoría de los cambios de y, entonces $R^2 \to 1$

Medidas de Ajuste: Ejemplo I

1. En el ejemplo relativo a la estimación de una función de producción para Perú.

$$Q = e^{\alpha_0} K^{\alpha_k} L^{\alpha_l}$$

el $R^2 = 0,80$. Eso significa que el 80 % de los cambios en la producción se deben a cambios en el capital y el trabajo.

2. Tenemos los siguientes datos $y = [40\ 55\ 25\ 5\ 10]'$. Su querido profesor quiere que estimen el siguiente modelo $y_t = \beta_0 + u_t$ utilizando el estimador MCO. Específicamente;
    - ¿Cuál es la estimación de β_0
    - ¿Cuál es el R^2?

Medidas de Ajuste: Ejemplo II

1. De las cuentas nacionales, tenemos la siguiente identidad $Y = C + I + G + XM$. Un estudiante que reprobó la econometría propuso el siguiente modelo para evaluar el impacto del consumo en el PIB (Y)

$$Y_t = \beta_0 + \beta_1 C_t + \ldots + \beta_5 M_t + u_t$$

   - ¿Cuál es la media de $\widehat{u}^2$?
   - ¿Cuál es el R^2?

2. el siguiente modelos

$$M_1 : Y_i = \beta_0 + \beta_1 X_{1i} + u_i$$
$$M_2 : Y_i = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + u_i$$

El R^2 de M_1 y M_2 son 0.701 y 0.80 respectivamente.
¿Qué modelo es mejor?

Una advertencia sobre R2

1. El R^2 aumenta cuando se agrega una nueva variable, por lo tanto, un aumento en el R^2 no significa que agregar una nueva variable realmente mejora el ajuste del modelo.
2. De hecho, aunque la nueva variable no es significativa, R^2 aumenta.
3. Un mal investigador puede inflar el ajuste del modelo agregando variables irrelevantes al modelo empírico.
4. R^2 es útil para mostrar el poder de explicación de su modelo, pero no para comparar.

Una alternativa: R2 ajustado o R2 adj

Tenemos una alternativa: el R^2 ajustado

$$R^2 = 1 \frac{N-1}{N-k} \frac{SCR}{SCT}$$

Algunas observaciones sobre esta medida de ajuste o ajuste

1. El R^2-ajustado siempre es más bajo que R^2
2. Agregar un regresor tiene dos efectos opuestos en R^2-ajustado
3. El R^2-ajustado puede ser negativo

Una advertencia con respecto a esta medida de ajuste.

1. La medida de R^2-ajustado aumenta (disminuye) cuando el cuadrado del valor t realmente calculado, relacionado con la variable nueva o adicional, es mayor (menor) que uno.
2. Al final, R^2-ajustado también es útil para mostrar el poder de explicación de su modelo y eso tiene cierto poder de comparación, pero también es débil.

Contenido

Contenido

Inferencia 2: prueba de hipótesis conjunta

Consideremos nuevamente el siguiente modelo (o PGD) con una constante y más de un regresor

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

en particular, podemos estar interesados en probar la importancia conjunta de β_1 y β_2. Por ejemplo

$$H_0 : \beta_1 = 0; \beta_2 = 0$$
$$H_a : \beta_j \neq 0 \text{ al menos un } j$$

¿Por qué podemos estar interesado en probar la significancia de coeficiente smúltiples al mismo tiempo?

Inferencia 2: prueba de hipótesis conjunta

Aquí la respuesta:

- Variables de interés; ocurrencia conjunta de variables aleatorias.
- Importancia del modelo. En el caso de una prueba de significancia global: un modelo con constante versus un modelo con regresores.
- Una prueba de significancia multivariante teniendo en cuenta la correlación entre variables aleatorias.

Prueba F

- La prueba F permite determinar, si en conjunto, todas las variables incluidas en el modelo explican a la variable analizada.
- Más concretamente, dada la regresión:
  $y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k + \hat{u}$
- La Hipótesis nula que se plantea es $\beta_1 = \beta_2 = ... = \beta_k = 0$
- Es decir, que todos los parámetros poblacionales, excepto el intercepto, tienen un valor poblacional de cero.
- En términos sencillos lo que se pregunta la prueba F es si incluir un conjunto de variables explicativas le gana a una regresión ingenua $(y = \hat{\beta}_0 + \hat{u})$

Inferencia II:¿Z o t-student? ¡NO!, ¡Distribución F!

En este caso, usamos la distribución F. El estadístico es:

$$F_{cal} = \frac{(SCR_R - SCR_{NR})}{SCR_{NR}} \frac{N-k}{q}$$

Por ejemplo, en el caso de una prueba de significancia global, el modelo restringido tiene en cuenta la restricción bajo la hipotesis nula

$$Y_i = \widehat{\gamma}_0 + \widehat{\epsilon}_i \qquad \text{(restringido)}$$

El modelo no restringido es el original

$$Y_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} \widehat{u}_i \qquad \text{(no restringido)}$$

$k$es el número de restricciones impuestas en el modelo original. Luego contrastamos esa F realmente calculadaçon valores críticos o simplemente calculamos el valor p asociado a esa prueba (es decir, $PR(F > F_{cal})$)

Inferencia II: expresión alternativa

En el caso de una prueba de hipotesis global, podemos re-expresar la expresión de la prueba F como lo siguiente:

$$F_{cal} = \frac{R_{NR}^2 - R_r^2}{1 - R_{NR}^2} \cdot \frac{N-k}{q}$$

En el caso de probar todos los coeficientes que aceptan la constante son cero (en términos estadísticos), tenemos la siguiente expresión

$$F_{cal} = \frac{R_{NR}^2}{1 - R_{NR}^2} \cdot \frac{N-k}{q}$$

Ejemplo I

Tenemos el siguiente modelo

$$T = \widehat{\beta}_0 + \widehat{\beta}_1 STR + \widehat{\beta}_2 EXP + \widehat{\beta}_3 PctEL + \widehat{u} \quad R^2 = 0{,}4366$$

Evaluamos la siguiente hipotesis

$$H_0 : \beta_1 = 0, \beta_2 = 0$$
$$H_1 : \beta_j \neq 0 \text{ almenos un } j$$

Eso implica estimar el siguiente modelo restringido:

$$TS = \widehat{\gamma}_0 + \widehat{\gamma}_3 PctEl + \widehat{\epsilon} \quad R^2 = 0{,}4149$$

Ejemplo I

Calculamos el F_{cal}

$$
\begin{aligned}
F_{cal} &= \frac{0{,}4366 - 0{,}4149}{1 - 0{,}4366} \cdot \frac{420 - 4}{2} \\
&= 8{,}0114
\end{aligned}
$$

luego contrastamos que F *"calculado realmente"* con valores críticos, en este caso el valor crítico del 5 % viene dado por $F(n_1, n_2) \equiv F(q, N - k) = F(2, 438) = 3{,}00$. Por tanto, rechazamos la hipótesis nula de que ambos coeficientes son cero en términos estadísticos.

Ejemplo II

Tenemos la siguiente estimación de una función de producción

$$lnQ_i = \widehat{\beta}_0 + \widehat{\beta}_1 lnK_i + \widehat{\beta}_2 lnL_i + \widehat{u}_i \quad R^2 = 0{,}8$$

realizamos una prueba de significancia conjunta sobre las elasticidades del capital y el trabajo, es decir, $H_0 : \widehat{\beta}_1 = \widehat{\beta}_2 = 0$ y $H_1 : \widehat{\beta}_j \neq 0$ al menos una j. Calculamos el estadístico F

$$\begin{aligned} F_{cal} &= \frac{0{,}8}{1-0{,}8} \cdot \frac{51-3}{2} \\ &= 96 \end{aligned}$$

luego contrastamos que F *"calculado realmente"* con valores críticos, en este caso el valor crítico del 1 % viene dado por $F(n_1, n_2) \equiv F(q, N-k) = F(2, 48) \approx 5{,}20$. Por tanto, rechazamos la hipótesis nula de que ambos coeficientes son cero en términos estadísticos.

prueba F: resumen de pasos

- Calcular las estimaciones de MCO.
- Enuncie la hipótesis a probar.
- Calcular el SSR o R^2 de modelos restringidos y no restringidos.
- Calcular la estadística F o F. çalculada realmente".
- Calcular el valor P o simplemente concluya utilizando valores críticos en algún nivel de significancia.
- La interpretación es muy importante, ¡no te pierdas ese paso !.

Contenido

P-VALUE

- Así como el estadístico T, la prueba F también tiene asociado un P-Value, el cual tiene la misma interpretación práctica: si es menor que 0.05 se rechaza la hipótesis nula de que las variables explicativas no sirven para explicar a y

Contenido

Referencias

- Referencia 1
- Referencia 2. *colocar alguna referencia*
- Referencia 3. *colocar alguna referencia*

Referencias

- Referencia 1
- Referencia 2. *colocar alguna referencia*
- Referencia 3. *colocar alguna referencia*

- Agregar alguna nota

Econometría Básica

Capítulo 04: Modelo de Regresión Lineal Múltiple-Inferencia

José Valderrama & Freddy Rojas
jtvalderrama@gmail.com & frojasca@gmail.com
Universidad Católica Santo Toribio de Mogrovejo

Septiembre de 2021