Econometría II

Performance

Carlos A. Yanes Guerra

2024-I

Performance de Pronosticos

Modelos de series y testeo

- Las series de tiempo deben aparte de cumplir con una serie de supuestos (sobre todo de consistencia de estimadores). Sus predicciones también deben ser sometidas a ciertas reglas y test con el objeto de ser muy técnicos con esto.

La parte residual

Los errores son contemplados como:

$$\epsilon_t = y_t - \hat{y}_{t+1}$$

Performance de Pronosticos

Mean Absolute Error (MAE)

> La media del valor absoluto del error se contempla como:

$$\text{MAE} = \left| \frac{\sum \epsilon_t}{n} \right|$$

Cuando se comparan métodos de pronósticos aplicados a una sola serie temporal, o a varias series temporales con las mismas unidades, el indicador de MAE es popular porque es fácil de entender y de calcular. Un método de pronósticos que minimice el MAE conducirá a previsiones de la mediana de la serie.

Performance de Pronosticos

Root Mean Square Error (RMSE)

> La raíz del error cuadratico medio se establece como:

$$\text{RMSE} = \sqrt{\frac{\sum \epsilon_t^2}{n}}$$

Tiende a ser un poco mas complejo la interpretación. Sin embargo cuando se tienen varios niveles de pronostico lo mejor es tener el menor de todos ellos. El principio de minimización del error sigue permanente en estas estimaciones.

Performance de Pronosticos

Mean Absolute Percentage Error

> Esta dado por el error porcentual esto es $p_t = 100 \times \frac{\epsilon_t}{y_t}$ y su medida singular se da por:

$$\text{MAPE} = \frac{\sum |p_t|}{n}$$

Tiene algunas desventajas sobre todo cuando $y_t = 0$, o inclusive en un caso particular va a ser infinito o tener valores de la serie muy cerca de cero. Por eso se hace una corrección propuesta por Armstrong (1978) y se establece

$$sMAPE = \text{promedio}\left[\frac{200 \times |y_t - \hat{y}_t|}{(y_t + \hat{y}_t)}\right]$$

Aunque tambien tiene sus desventajas. Se vuelve útil en algunas ocasiones.

Performance de Pronosticos

Scaled Errors

> Es alternativo al test de sMAPE fue propuesto por Hyndman y Koehler (2006). Intenta comparar la precisión del pronostico incluso en series que tienen distintas unidades. Para series no estacionales se propone:

$$q_j = \frac{\epsilon_j}{\frac{1}{T-1}\sum |y_t - y_{t-1}|}$$

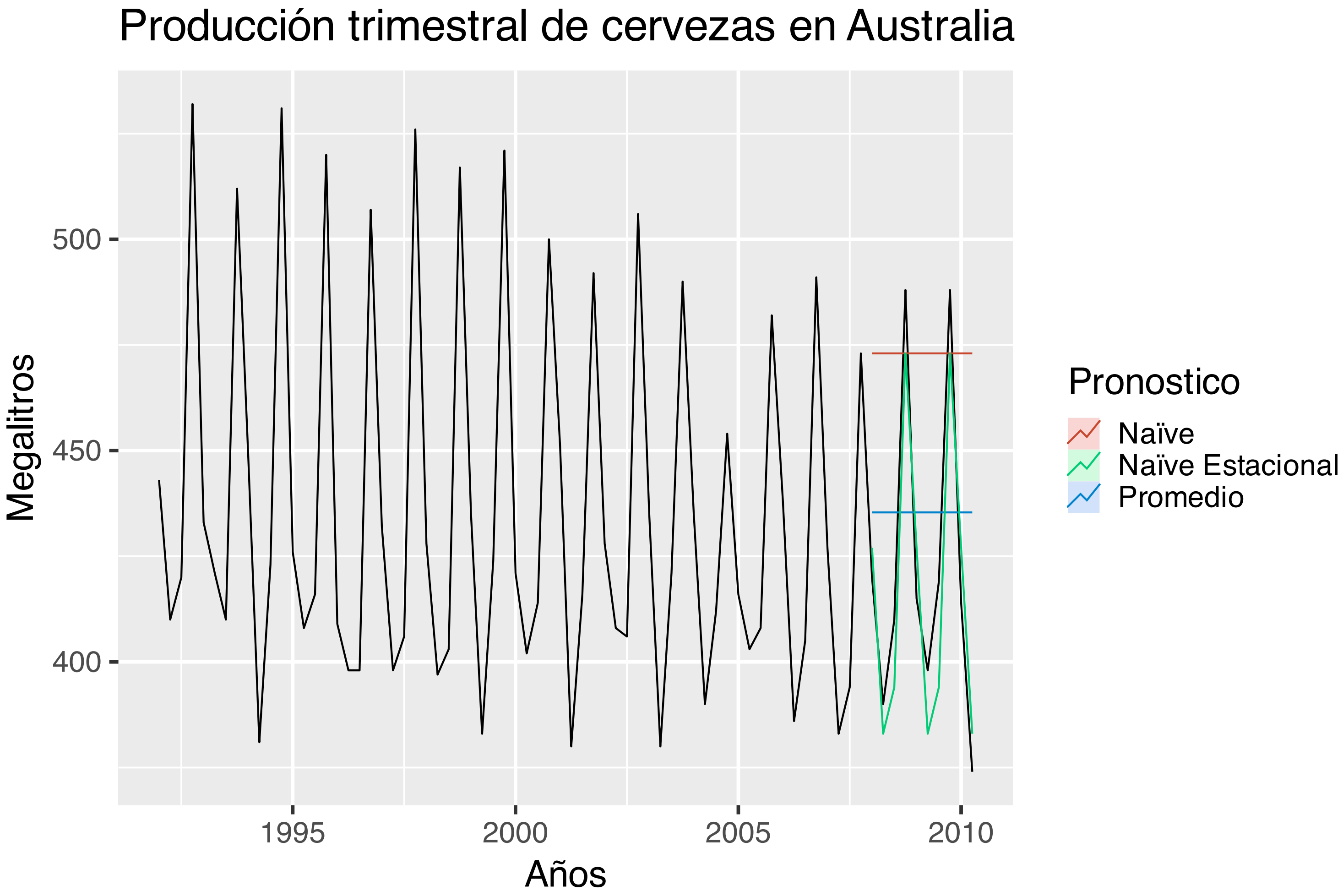
De tal manera que si desea mirar la parte estacional es simplemente:

$$q_j = \frac{\epsilon_j}{\frac{1}{T-m}\sum_{t=m+1} |y_t - y_{t-m}|}$$

Finalmente, el test queda como:

$$MASE = \text{Promedio}\left(|q_j|\right)$$

Performance de Pronosticos



Performance de Pronosticos

Performance de modelos

```
#>                     ME     RMSE      MAE        MPE     MAPE     MASE        ACF1
#> Training set   0.000 43.62858 35.23438 -0.9365102 7.886776 2.463942 -0.10915105
#> Test set     -13.775 38.44724 34.82500 -3.9698659 8.283390 2.435315 -0.06905715
#>              Theil's U
#> Training set        NA
#> Test set      0.801254
```

```
#>                         ME     RMSE      MAE         MPE     MAPE     MASE
#> Training set   0.4761905 65.31511 54.73016  -0.9162496 12.16415 3.827284
#> Test set     -51.4000000 62.69290 57.40000 -12.9549160 14.18442 4.013986
#>                     ACF1 Theil's U
#> Training set -0.24098292        NA
#> Test set     -0.06905715  1.254009
```

```
#>                     ME     RMSE  MAE        MPE     MAPE      MASE       ACF1
#> Training set -2.133333 16.78193 14.3 -0.5537713 3.313685 1.0000000 -0.2876333
#> Test set      5.200000 14.31084 13.4  1.1475536 3.168503 0.9370629  0.1318407
#>              Theil's U
#> Training set        NA
#> Test set      0.298728
```

Performance de Pronosticos## Performance de modelos

```
beer3 <- window(ausbeer, start=2008)
accuracy(beerfit1, beer3) # Modelo Promedio
accuracy(beerfit2, beer3) # Naive
accuracy(beerfit3, beer3) # Naive Estacional
```

La función `accuracy` nos muestra el resumen de cada uno de los modelos utilizando los criterios anteriores.

Performance de Pronosticos

Coeficiente de Theil

> Así como es funcional para desigualdad, tambien lo es para métodos de pronosticos. Queremos que si la serie original se comporta de cierta manera, la serie predicha tambien haga lo mismo. Su estipulacion va con raices de medias.

$$\text{Coeficiente Theil} = \frac{\sqrt{promedio\ \epsilon_t^2}}{\sqrt{promedio\ y_t} + \sqrt{promedio\ \hat{y}_t}}$$

Como en desigualdad, si Theil se hace (1) es lo peor en distribución. Queremos que nuesto modelo de estimación sea cercano a (0) para tener un muy buen ajuste.

Modelos univariados autoregresivos

Modelos univariados

Redefiniendo lo del operador Rezago o "Lag"" es representado por la letra (L).

$$
\begin{aligned}
Ly_t &= y_{t-1} & Lc &= c \\
L^n y_t &= y_{t-n} & L^0 y_t &= y_t \\
L^2 y_t &= y_{t-2} & L^k L^j &= L^{k+j}
\end{aligned}
$$

$$L^{-1}(Ly_t) = Ly_{t-1} = y_t$$

¿Cómo sería un modelo de $y_t = \phi y_{t-1} + \epsilon_t$, expresado en términos de rezago?

R./

$$y_t = \phi L y_t + \epsilon_t$$

Ahora uno como $y_t = \phi y_{t-1} + \phi y_{t-7} + \epsilon_t$

R./

$$y_t = \phi_1 L y_t + \phi_7 L^7 y_t + \epsilon_t$$

Modelos univariados

Operador rezago en AR(1)

$$\begin{aligned} y_t &= \phi y_{t-1} + \epsilon_t \\ y_t - \phi y_{t-1} &= \epsilon_t \\ y_t - \phi L y_t &= \epsilon_t \end{aligned}$$

esto nos da que:

$$\boxed{y_t - \phi L y_t = \epsilon_t}$$

Recuerde por un momento la formula de Taylor

$$1 + \rho + \rho^2 + \rho^3 + \cdots = \sum_{i=1}^{\infty} \rho^i = \frac{1}{1-\rho}$$

Modelos univariados

Regresando al caso

$$\begin{aligned} y_t - \phi L y_t =& \epsilon_t \\ y_t =& \frac{1}{1-\phi L}\epsilon_t \end{aligned}$$

Acá tenemos un par de condiciones y son:

Si $|\phi| < 1$, entonces $(1-\phi L)^{-1}$ existe por eso de:

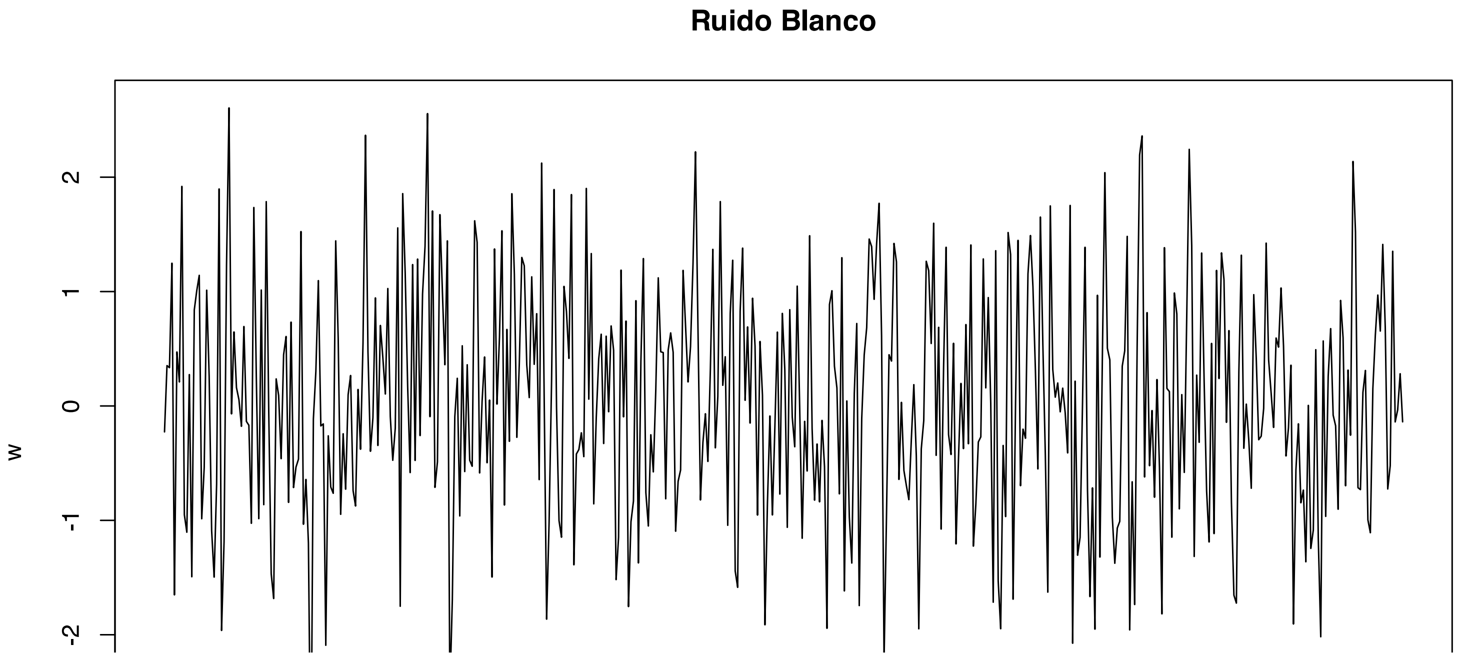
$$(1-\phi L)^{-1} = \frac{1}{(1-\phi L)} = 1 + \phi L + \phi^2 L^2 + \phi^3 L^3 = \sum_{i=1}^{\infty} \phi^i L^i$$

$$\begin{aligned} y_t - \phi L y_t =& \epsilon_t \\ y_t(1-\phi L) =& \epsilon_t \\ y_t =& (1-\phi L)^{-1}\epsilon_t \\ y_t =& (1 + \phi L + \phi^2 L^2 + \phi^3 L^3 + \cdots)\epsilon_t \\ y_t =& \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \phi^3\epsilon_{t-3} + \cdots \end{aligned}$$

Modelos univariados

Ruido blanco

> Un proceso estocástico (lo mas independiente) se considera aleatorio, posee una característica o estructura no discernible, su proceso cambia a través del tiempo. Ejemplo: El Baloto electrónico.



Modelos univariados

Ruido blanco

$Y_t = \epsilon_t, t = 1, 2, 3 \dots T$, es ruido blanco si y solo si:

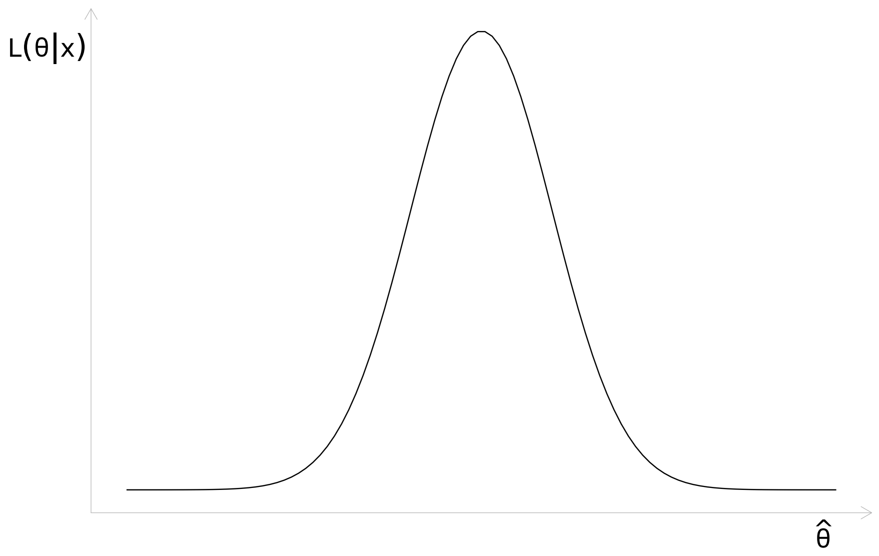
- Media cero: $E(Y_t) = 0 \ \forall t$.
- Varianza constante: $Var(Y_t) = \sigma^2$ y este es $< \infty$.
- Covarianza cero: $Cov(Y_i, Y_j) = 0 \ \forall \ i \neq j$.

Cuando hace lanzamientos con un dado. La media es de 3.5 (21/6), La probabilidad de que salga un valor es de 1/6 y el evento (i) que ocurre al lanzarlo es independiente de (j), es decir, el nuevo lanzamiento no depende del anterior ni tampoco de su futuro.

Con respecto a Máxima verosimilitud

Máxima verosimilitud

El método de máxima verosimilitud (Maximum Likelihood Estimation) es un enfoque estadístico utilizado para estimar los parámetros de un modelo probabilístico a partir de un conjunto de observaciones o datos. El objetivo del método es encontrar los valores de los parámetros que maximizan la probabilidad de observar los datos que tenemos, asumiendo que los datos siguen una cierta distribución de probabilidad.



Máxima verosimilitud

Se debe empezar desde la composición de un vector de característica aleatoria y con una distribución que depende de un parámetro desconocido como (Θ), por tanto se tiene que $X \in \{x_1, x_2, x_3, \cdots, x_n\}$. Por tanto la función de verosimilitud de este vector vendrá a ser dada como:

$$L(\Theta) = (fx_1, fx_2, fx_3, \cdots, fx_n)(x_1, x_2, x_3, \cdots, x_n|\Theta)$$

Cuando las variables sean independientes (explicativas) entonces se procede a establecer la función de verosimilitud como:

$$L(\Theta) = fx_1(x_1;\Theta), fx_2(x_2;\Theta), fx_3(x_3;\Theta) \cdots, fx_n(x_n;\Theta)$$

Si dado el caso, estas variables resultan ser idénticamente distribuidas, entonces se tendrá:

$$L(\Theta) = f(x_1;\Theta), f(x_2;\Theta), f(x_3;\Theta) \cdots, f(x_n;\Theta)$$

Que sería el caso de una muestra aleatoria.

Máxima verosimilitud

Entonces, para obtener el valor de (Θ) que maximiza a la función de verosimilitud se debe establecer la estimación de $L(\Theta)$ o estimador verosímil. La razón principal de calculo, debe ser encontrar un valor numérico observable $(x_1, x_2, x_3, \cdots, x_n)$, de la muestra aleatoria tenga probabilidad máxima.

Sea una muestra aleatoria (m.a), con valores $X_1, \ldots, X_n \sim f(X|\theta)$, se debe encontrar el estimador $\theta =?$ que maximiza la función.

- Primer paso: es plantear la función de máxima verosimilitud:

  $L\left(\theta|x\right) = \prod_{i=1}^{n} f\left(x_i|\theta\right)$

- Segundo paso: es tratar de encontrar la referencia del estimador que es:

$$L\left(\theta_1|x\right) > L\left(\theta_2|x\right)$$

Si lo anterior ocurre $\Rightarrow \theta_1 = \theta$ y será mas verosímil que $\theta_2 = \theta$

- Tercer paso: es escoger ese mejor estimador (mas creíble), es decir, $\hat{\theta} \in \Theta$.

Máxima verosimilitud

Entender lo anterior no es tan trivial, se hace necesario conocer que la estimación máxima verosimilitud (E.M.V) este en función de los valores provistos:

$$\hat{\theta} = E.M.V\left(\theta|X_1,\ldots,X_n\right) = f\left(X_1,\ldots,X_n\right)$$

Lo que en mejores términos vendría a ser:

$$L = \left(\hat{\theta}|X_1,\ldots,X_n\right) = \underset{\{\theta \in \Theta\}}{max}\, L\left(\theta|X_1,\ldots,X_n\right)$$

Veamos un ejemplo 🤓

Ejemplo: Máxima Verosimilitud

Sea la siguiente función definida como:

$$f(x,\theta) = \frac{1}{\theta} e^{\frac{-x}{\theta}} \; ; \; x > 0 \; ; \; \theta > 0$$

Halle el estimador θ

Debemos plantear la función de densidad de cada una de las variables de una m.a y esto es: Tenemos que las variables son $(x_1, \ldots, x_n)$ y la función de cada una de ellas vendrá a ser:

$$f(x_1,\theta) = \frac{1}{\theta} e^{\frac{-x_1}{\theta}} \; ; \; f(x_2,\theta) = \frac{1}{\theta} e^{\frac{-x_2}{\theta}} \; ; \; f(x_n,\theta) = \frac{1}{\theta} e^{\frac{-x_n}{\theta}}$$

La idea es resolver el producto o multiplicación de las funciones usando la formula del logaritmo de verosimililtud y esto resulta:

$$L\left(\hat{\theta} \mid X_1, \ldots, X_n\right) = \frac{1}{\theta} e^{\frac{-x_1}{\theta}} * \frac{1}{\theta} e^{\frac{-x_2}{\theta}} * \cdots * \frac{1}{\theta} e^{\frac{-x_n}{\theta}}$$

Ejemplo: Máxima Verosimilitud

Para lo cual, simplificamos la expresión (lo mas que se pueda\footnote{Acá es útil utilizar todas las herramientas de calculo básico y álgebra.})

$$L\left(\hat{\theta} \mid X_1, \ldots, X_n\right) = \frac{1}{\theta^n} e^{\left\{\frac{-x_1}{\theta} + \frac{-x_2}{\theta} + \cdots + \frac{-x_n}{\theta}\right\}}$$

Obteniendo de forma mas simple:

$$L\left(\hat{\theta} \mid X_1, \ldots, X_n\right) = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\{x_1 + x_2 + \cdots + x_n\}}$$

Que haciendo mas simple la ecuación puede ser reemplazada usando el termino de la sumatoria:

$$L\left(\hat{\theta} \mid X_1, \ldots, X_n\right) = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum x_i}$$

El siguiente proceso será derivar.

Ejemplo: Máxima Verosimilitud

Toda la expresión que ha quedado simplificada y de ahí aplicar el despeje como tal, resultando:

Estableciendo la condición de primer orden:

$$\frac{\partial L(x_1, \ldots, x_n, \theta)}{\partial \theta} = 0$$

Tomar la expresión tal cual se encuentra situada seria algo complejo. Una forma de linearizar es aplicando logaritmos a la expresión de M.V y de ahí si derivar.

$$lnL = ln\left(\frac{1}{\theta^n}\right) + Lne^{-\frac{1}{\theta}\sum x_i}$$

Se aplican todas las propiedades de Logaritmo.

$$lnL = ln(1) - n\; ln\theta - \frac{1}{\theta}\sum x_i Ln(e)$$

Ejemplo: Máxima Verosimilitud

Conocemos que el logaritmo de (1) es cero y que el Ln de (e) es 1, por ende ahora nos encontramos con:

$$lnL = -n\; ln\theta - \frac{1}{\theta}\sum x_i$$

Derivando la expresión con respecto a $\theta \Rightarrow$

$$\frac{\partial LnL}{\partial \theta} = -\frac{n}{\theta} + \frac{\sum x_i}{\theta^2} = 0$$

Despejando θ:

$$\frac{\sum x_i}{\theta^2} = \frac{n}{\theta}$$

Dando como resultado:

$$\theta = \frac{\sum x_i}{n}$$

Para este caso el estimador $\theta = \overline{X}$.

Bibliografía

Chatfield, C. (2000). *Time-series forecasting*. CRC press.

Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia.

Righetti, N., (2022). *Time Series Analysis With R*. Bookdown.

Shumway, R., & Stoffer, D. (2019). *Time series: a data analysis approach using R*. CRC Press.

¡Gracias!

Del contorno de series

Seguimos aprendiendo

Syllabus/ Curso

@keynes37

cayanes@uninorte.edu.co