

# Series de Tiempo

## Series de Tiempo en la APN

Una API REST es un servicio web que permite hacer consultas a una base de datos o aplicación en línea. Muchas APIs pueden usarse simplemente como una URL configurable / parametrizable en el navegador.

La Administración Pública Nacional dispone de APIs de datos en las que todos los organismos pueden publicar: <https://apis.datos.gob.ar>

Una de ellas permite consultar indicadores con evolución temporal de distintos ministerios (actualmente +20 mil series):

- **Documentación API:** <https://apis.datos.gob.ar/series>
- **Buscador web de series:** <http://datos.gob.ar/series> (permite buscar los ids de las series deseadas)

El buscador permite llevarse una URL a la API que descarga un CSV actualizado de los indicadores elegidos:

- **Tipo de cambio vendedor BNA:** [http://apis.datos.gob.ar/series/api/series/?limit=1000&ids=168.1\\_T\\_CAMBIOR\\_D\\_0\\_0\\_26&format=csv](http://apis.datos.gob.ar/series/api/series/?limit=1000&ids=168.1_T_CAMBIOR_D_0_0_26&format=csv)
- **IPC Nacional. Nivel General:** [http://apis.datos.gob.ar/series/api/series/?limit=1000&ids=148.3\\_INIVELNAL\\_DICI\\_M\\_26&format=csv](http://apis.datos.gob.ar/series/api/series/?limit=1000&ids=148.3_INIVELNAL_DICI_M_26&format=csv)
- **EMAE:** [http://apis.datos.gob.ar/series/api/series/?limit=1000&ids=143.3\\_NO\\_PR\\_2004\\_A\\_21&format=csv](http://apis.datos.gob.ar/series/api/series/?limit=1000&ids=143.3_NO_PR_2004_A_21&format=csv)

Como te podrás imaginar, `limit`, `ids` y `format` son algunos de esos parámetros que te permiten personalizar la consulta:

- `ids`: el parámetro más importante! Permite pedir una lista de series por id, separados por comas.
- `format`: puede ser "csv" o "json".
- `limit`: por default la API devuelve 100 filas, pero se puede extender hasta 1000.

En la mayoría de las APIs REST, los parámetros empiezan después del `?` y se separan por `&`.

## Time series dataframes

### Descargar CSVs de la API de series de tiempo

- Descargar el IPC en un dataframe de R. Hint: `df = read.csv(*)`
- Descargar el Estimador Mensual de la Actividad Económica (EMAE) general, de la Construcción y de la Industria Manufacturera en el mismo dataframe. Hint: `ids=*,*,*`.
- Descargar el tipo de cambio mínimo, promedio y máximo mensual, usando la API (sin programar para eso en R). Hint: `&collapse=month` transforma la llamada en "mensual" y `&ids=:min,:avg` indica cómo agregar los valores de esas series.
- Descargar el IPC, la inflación mensual y la inflación inter-anual en un mismo dataframe, usando la API (sin programar para eso en R) sólo desde 2017 en adelante. Hint: `&ids=:percent_change,:percent_change_a_year_ago / &start_date=*`.

### Buscar relaciones entre series de tiempo

Las series de tiempo tienen la característica distintiva de que todas están relacionadas con un índice de tiempo (son datos estructurados cronológicamente, con una frecuencia determinada). Esto nos permite agregar nuevas posibilidades en la búsqueda de relaciones entre variables.

a. Crear un modelo que explique el IPC nivel general nacional en base a las expectativas de inflación futura desde el 2017 en adelante, y hacer un scatter plot con su recta de regresión.

```
a1. Descargar en un mismo dataframe ambas series. Hint: df = read.csv("https://
a2. Generar el primer modelo de regresión con los primeros 22 valores de c/u. F
a3. Scatter plot de las dos variables y agregar la recta de regresión. Hint: pl
```

b. Existe una correlación! Pero tal vez las expectativas de inflación futura inciden con uno o dos meses de retraso en la inflación mensual real?

Para esto hace falta comparar una regresión entre las dos variables, contra una en la que las expectativas están 1 o 2 meses *adelantadas*. Esto requiere tratar las variables como series de tiempo y aplicarles *lags*. Para comparar el rendimiento del mismo modelo con lags diferentes, vamos a hacer un gráfico partido en 4 para ver cómo se comporta cada uno.

```
b1. Generar 3 modelos: uno sin lags, uno con 1 mes de lag y otro con 2 meses de
b1. Partir la pantalla en 4. Hint: par(mfrow=c(*,*))
b2. Graficar nuevamente el scatter plot con la recta de regresión (como en el p
b3. Agregar un gráfico comparando los R2 de cada modelo contra la cantidad de *
```

```
info0=summary(modelo0)
info1=summary(modelo1)
info2=summary(modelo2)

ajuste=c(info0$r.squared, info1$r.squared, info2$r.squared)

plot(c(0,1,2), ajuste, xlab="Lag", ylab="Bondad del ajuste [R2]", pch=16)
```

b4. Corroborar, de todas formas, que ninguno de los 3 modelos es muy bueno porq

**Bonus track!** Existen dos librerías que facilitan el uso de *lags* en regresiones con series de tiempo llamada *dynlm* y *Hmisc*. Requiere que el dataframe se reconozca como un dataframe *de series de tiempo*, con un índice de tiempo y permite aplicar lags con una función más fácilmente...

```
b1. install.packages("dynlm"); install.packages("Hmisc"); library(Hmisc); libra
b2. library(zoo); df_ts = read.zoo(df, index = 1, tz = "", format = "%Y-%m-%d")
b3. Regresar la inflación mensual contra las expectativas de inflación futura.
b3. Regresar la inflación mensual contra las expectativas de inflación futura c
b4. Comparar los R2 de cada regresión. ¿Mejoró el % de variabilidad de la infla
```

## LASSO

Como dijimos tal vez nos falten variables explicativas, la inflación es un fenómeno multicausal. Probablemente el mejor modelo explicativo sea uno que incluya muchas variables... La base de series de tiempo ofrece 20 mil! Si bien no tiene sentido probar con todas, habría que buscar un método que elija el mejor modelo al enfrentarse a muchas variables. (Ver tutorial de LASSO en R para más detalles: [https://web.stanford.edu/~hastie/glmnet/glmnet\\_alpha.html](https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html))

a. Cargar en un dataframe de series de tiempo la inflación mensual, la tasa de interés, la variación mensual del tipo de cambio nominal, las expectativas de inflación futura y el tipo de cambio real multilateral, desde 2017. Hint: df2 =

```
read.csv("https://apis.datos.gob.ar/series/api/series/?
limit=1000&ids=*:percent_change,*,*:percent_change,*,*&format=csv&start_date=20
```

b. Usar el método lasso (librería *glmnet*) para encontrar el mejor modelo posible entre estas variables.

```
b1. install.packages("glmnet", repos = "http://cran.us.r-project.org"); library
b2. Convertir el dataframe a series de tiempo. Hint: df_ts2 = read.zoo(*, index
```

b3. Remover las filas que tengan algún valor nulo. Hint: `* = *[complete.cases(*`  
b4. Crear variables con lags de 1 mes para analizar efectos desplazados en el t  
b5. Crear una matriz de predictores x (sin la variable a predecir). Hint: `x = c`  
b6. Correr lasso usando cross validation. Hint: `cvfit = cv.glmnet(x, y)`  
b7. Encontrar los coeficientes del modelo que minimiza el error de predicción.

c. Comparar la predicción del modelo con la realidad en un plot. Hint: `y_predict =`  
`predict(cvfit, newx = *)` y `plot(y, *)`