**# Medias Móviles{#moving-averages}**

En esta página se tratan dos métodos para calcular y visualizar las medias móviles:

1. Calcular con el paquete **deslizante**
2. Calcular dentro de un comando ggplot() con el paquete **tidyquant**

## Preparación

### Cargar paquetes

Este trozo de código muestra la carga de los paquetes necesarios para los análisis. En este manual destacamos p\_load() de **pacman**, que instala el paquete si es necesario y lo carga para su uso. También puede cargar los paquetes instalados con library() de **.** Consulta la página sobre [los fundamentos de R](#r-basics) para obtener más información sobre los paquetes de R.

### Importar datos

Importamos los datos de casos de una epidemia de ébola simulada. Si quiere seguir el proceso, [clica para descargar linelist "limpia"](https://github.com/epirhandbook/Epi_R_handbook/raw/master/data/case_linelists/linelist_cleaned.rds) (como archivo .rds). Importe los datos con la función import() del paquete **rio** (maneja muchos tipos de archivos como .xlsx, .csv, .rds - vea la página de [importación y exportación](#import-and-export) para más detalles).

A continuación se muestran las primeras 50 filas del listado.

## Calcular con el ****deslizador****

**Utiliza este enfoque para calcular una media móvil en un dataframe antes de trazarla.**

El paquete **slider** proporciona varias funciones de "ventana deslizante" para calcular medias móviles, sumas acumulativas, regresiones móviles, etc. Trata un dataframe como un vector de filas, permitiendo la iteración por filas sobre un dataframe.

Estas son algunas de las funciones más comunes:

* slide\_dbl() - itera a través de una columna numérica (de ahí "\_dbl") realizando una operación mediante una ventana deslizante
  + slide\_sum() - función abreviada de suma rodante para slide\_dbl()
  + slide\_mean() - función abreviada de media móvil para slide\_dbl()
* slide\_index\_dbl() - aplica la ventana rodante en una columna numérica utilizando una columna separada para indexar la progresión de la ventana (útil si se rueda por fecha con algunas fechas ausentes)
  + slide\_index\_sum() - Función abreviada de suma rodante con indexación
  + slide\_index\_mean() - Función de acceso directo a la media móvil con indexación

El paquete **deslizante** tiene muchas otras funciones que se tratan en la sección de Recursos de esta página. Tocamos brevemente las más comunes.

**Argumentos básicos**

* .x, el primer argumento por defecto, es el vector sobre el que iterar y al que aplicar la función
* .i = para las versiones de "índice" de las funciones de **deslizamiento** - proporciona una columna para "indexar" el rollo (véase la sección [siguiente](#roll_index))
* .f =, el segundo argumento por defecto, bien:
  + Una función, escrita sin paréntesis, como media, o
  + Una fórmula, que se convertirá en una función. Por ejemplo ~ .x - mean(.x) devolverá el resultado del valor actual menos la media del valor de la ventana
* Para más detalles, consulta este [material de referencia](https://davisvaughan.github.io/slider/reference/slide.html)

**Tamaño de la ventana**

Especifique el tamaño de la ventana utilizando los argumentos .before, .after o ambos:

* .before = - Proporcionar un número entero
* .after = - Proporcionar un número entero
* .complete = - Ponga este valor a TRUE si sólo quiere que se realicen cálculos en ventanas completas

Por ejemplo, para conseguir una ventana de 7 días que incluya el valor actual y los seis anteriores, utiliza .before = 6. Para conseguir una ventana "centrada" proporcione el mismo número tanto a .before = como a .after =.

Por defecto, .complete = será FALSE por lo que si la ventana completa de filas no existe, las funciones utilizarán las filas disponibles para realizar el cálculo. Si se ajusta a TRUE, los cálculos sólo se realizan en ventanas completas.

**Ventana expansiva**

Para lograr operaciones acumulativas, establezca el argumento .before = en Inf. Esto realizará la operación sobre el valor actual y todos los que vengan antes.

### Enrollar por fecha

El caso más probable de uso de un cálculo rotativo en epidemiología aplicada es examinar una medida a lo largo del tiempo. Por ejemplo, una medición continua de la incidencia de casos, basada en el recuento diario de casos.

Si tiene datos de series temporales limpios con valores para cada fecha, puede estar bien utilizar slide\_dbl(), como se demuestra aquí en la página de [series temporales y detección de brotes](#timeseries_moving).

Sin embargo, en muchas circunstancias de epidemiología aplicada puede haber fechas ausentes en sus datos, donde no hay eventos registrados. En estos casos, es mejor utilizar las versiones "índice" de las funciones **deslizantes**.

### Datos indexados

A continuación, mostramos un ejemplo utilizando slide\_index\_dbl() en la lista de casos. Digamos que nuestro objetivo es calcular una incidencia rodante de 7 días - la suma de casos utilizando una ventana rodante de 7 días. Si está buscando un ejemplo de media rodante, vea la sección de abajo sobre [rodamiento agrupado](#roll_slider_group).

Para empezar, se crean los datos daily\_counts para reflejar los recuentos diarios de casos del listado, calculados con count() de **dplyr**.

Aquí está el dataframe de daily\_counts - hay nrow(daily\_counts) filas, cada día está representado por una fila, pero especialmente al principio de la epidemia algunos días no están presentes (no hubo casos admitidos en esos días).

Es crucial reconocer que una función estándar de balanceo (como slide\_dbl() utilizaría una ventana de 7 filas, no de 7 días. Por lo tanto, si hay fechas ausentes, ¡algunas ventanas se extenderán realmente más de 7 días naturales!

Se puede conseguir una ventana móvil "inteligente" con slide\_index\_dbl(). El "índice" significa que la función utiliza una columna independiente como "índice" para la ventana móvil. La ventana no se basa simplemente en las filas del dataframe.

Si la columna índice es una fecha, tiene la posibilidad añadida de especificar la extensión de la ventana a .before = y/o .after = en unidades de **lubridate** days() o months(). Si hace estas cosas, la función incluirá los días ausentes en las ventanas como si estuvieran allí (como valores NA).

Mostremos una comparación. A continuación, calculamos la incidencia móvil de casos de 7 días con ventanas regulares e indexadas.

Obsérvese cómo en la columna normal de las 7 primeras filas el recuento aumenta constantemente a pesar de que las filas no tienen 7 días de diferencia. La columna adyacente "indexada" tiene en cuenta estos días naturales ausentes, por lo que sus sumas de 7 días son mucho menores, al menos en este periodo de la epidemia en el que los casos están más alejados.

Ahora puede trazar estos datos utilizando ggplot():

### Rodando por el grupo

Si agrupa sus datos antes de utilizar una función **deslizante**, las ventanas deslizantes se aplicarán por grupo. Tenga cuidado de organizar sus filas en el orden deseado por grupo.

Cada vez que se inicia un nuevo grupo, la ventana deslizante se reinicia. Por lo tanto, un matiz a tener en cuenta es que si sus datos están agrupados y ha establecido .complete = TRUE, tendrá valores vacíos en cada transición entre grupos. A medida que la función se desplaza hacia abajo a través de las filas, cada transición en la columna de agrupación reiniciará la acumulación del tamaño mínimo de la ventana para permitir un cálculo.

Consulta la página del manual sobre [Agrupar datos](#grouping-data) para obtener detalles sobre la agrupación de datos.

A continuación, contamos los casos del listado por fecha y por hospital. Luego ordenamos las filas en orden ascendente, primero ordenando por hospital y luego dentro de éste por fecha. A continuación establecemos group\_by(). Entonces podemos crear nuestra nueva media móvil.

Aquí está el nuevo conjunto de datos:

Ahora podemos trazar las medias móviles, mostrando los datos por grupo especificando ~ hospital a facet\_wrap() en ggplot(). Para divertirnos, trazamos dos geometrías: una geom\_col() que muestra los recuentos de casos diarios y una geom\_line() que muestra la media móvil de 7 días.

**PELIGRO:** Si obtiene un error que dice "slide() fue obsoleta en tsibble 0.9.0 y ahora está desaparecida. Por favor, utiliza slider::slide() en su lugar", significa que la función slide() del paquete **tsibble** está enmascarando la función slide() del paquete **slider**. Solucione esto especificando el paquete en el comando, como slider::slide\_dbl().

## Calcular con ****tidyquant**** dentro de ggplot()

El paquete **tidyquant** ofrece otro enfoque para calcular las medias móviles, esta vez desde el propio comando ggplot().

Debajo del listado, los datos se cuentan por fecha de inicio, y esto se traza como una línea descolorida (alfa < 1). Encima hay una línea creada con geom\_ma() del paquete **tidyquant**, con una ventana de 7 días (n = 7) con el color y el grosor especificados.

Por defecto geom\_ma() utiliza una media móvil simple (ma\_fun = "SMA"), pero se pueden especificar otros tipos, como:

* "EMA" - media móvil exponencial (más peso a las observaciones recientes)
* "WMA" - media móvil ponderada (los wts se utilizan para ponderar las observaciones en la media móvil)
* Otros se pueden encontrar en la documentación de la función

Consulta esta [viñeta](https://cran.r-project.org/web/packages/tidyquant/vignettes/TQ04-charting-with-tidyquant.html) para obtener más detalles sobre las opciones disponibles en **tidyquant**.

## Recursos

Vea la útil [viñeta](https://cran.r-project.org/web/packages/slider/vignettes/slider.html) en línea [para elpaquete de**deslizadores**](https://cran.r-project.org/web/packages/slider/vignettes/slider.html)

La [página github del](https://github.com/DavisVaughan/slider) **deslizador**

Una [viñeta](https://davisvaughan.github.io/slider/articles/slider.html) **deslizante**

[viñeta tidyquant](https://cran.r-project.org/web/packages/tidyquant/vignettes/TQ04-charting-with-tidyquant.html)

Si tu caso de uso requiere que te "saltes" los fines de semana e incluso los días festivos, puede que te guste el paquete **almanaque**.