# # Tests estadísticos sencillos

{#simple-statistical-tests}

Esta página muestra cómo realizar tests estadísticos sencillos utilizando R base, **rstatix** y **gtsummary**.

* Prueba T
* Prueba de Shapiro-Wilk
* Prueba de suma de rangos de Wilcoxon
* Prueba de Kruskal-Wallis
* Prueba de Chi-cuadrado
* Correlaciones entre variables numéricas

...se pueden realizar muchas otras pruebas, pero sólo mostramos estas comunes y enlazamos con más documentación.

Cada uno de los paquetes mencionados aporta ciertas ventajas y desventajas:

* Utiliza las funciones **básicas de** R para imprimir una salida estadística en la consola de R
* Utiliza las funciones **rstatix** para devolver los resultados en un dataframe, o si desea que las pruebas se ejecuten por grupos
* Utiliza **gtsummary** si desea imprimir rápidamente tablas listas para su publicación

## 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). Importa tus datos con la función import() del paquete **rio** (acepta muchos tipos de archivos como .xlsx, .rds, .csv - vea la página de [importación y exportación](#import-and-export) para más detalles).

A continuación se muestran las primeras 50 filas del listado.

## ****base**** R

Puedes utilizar las funciones **básicas de** R para realizar pruebas estadísticas. Los comandos son relativamente sencillos y los resultados se imprimen en la consola de R para su visualización. Sin embargo, las salidas suelen ser listas y, por lo tanto, son más difíciles de manipular si desea utilizar los resultados en operaciones posteriores.

### Pruebas T

Una prueba [t](https://en.wikipedia.org/wiki/Student's_t-test), también llamada "Prueba t de Student", se utiliza normalmente para determinar si existe una diferencia significativa entre las medias de alguna variable numérica entre dos grupos. Aquí mostraremos la sintaxis para hacer esta prueba dependiendo de si las columnas están en el mismo dataframe.

**Sintaxis 1:** Esta es la sintaxis cuando tus columnas numéricas y categóricas están en el mismo dataframe. Proporcione la columna numérica en el lado izquierdo de la ecuación y la columna categórica en el lado derecho. Especifique los datos a data =. Opcionalmente, establezca paired = TRUE, y conf.level = (0.95 por defecto), y alternative = (ya sea "two.sided", "less", o "greater"). Introduzca ?t.test para obtener más detalles.

**Sintaxis 2:** Puedes comparar dos vectores numéricos separados utilizando esta sintaxis alternativa. Por ejemplo, si las dos columnas están en conjuntos de datos diferentes.

También puede utilizar una prueba t para determinar si la media de una muestra es significativamente diferente de algún valor específico. Aquí realizamos una prueba t de una muestra con la media poblacional conocida/hipotética como mu =:

### Prueba de Shapiro-Wilk

La [prueba de Shapiro-Wilk](https://en.wikipedia.org/wiki/Shapiro–Wilk_test) puede utilizarse para determinar si una muestra procede de una población distribuida normalmente (un supuesto de muchas otras pruebas y análisis, como la prueba t). Sin embargo, sólo puede utilizarse en una muestra de entre 3 y 5000 observaciones. Para muestras más grandes puede ser útil un [gráfico de cuantiles](https://ggplot2.tidyverse.org/reference/geom_qq.html).

### Prueba de suma de rangos de Wilcoxon

La prueba de suma de rangos de Wilcoxon, también llamada [prueba U de Mann-Whitney](https://en.wikipedia.org/wiki/Mann–Whitney_U_test), se utiliza a menudo para ayudar a determinar si dos muestras numéricas proceden de la misma distribución cuando tus poblaciones no se distribuyen normalmente o tienen una varianza desigual.

### Prueba de Kruskal-Wallis

La prueba de [Kruskal-Wallis](https://en.wikipedia.org/wiki/Kruskal–Wallis_one-way_analysis_of_variance) es una extensión de la prueba de suma de rangos de Wilcoxon que puede utilizarse para comprobar las diferencias en la distribución de más de dos muestras. Cuando sólo se utilizan dos muestras, los resultados son idénticos a los de la prueba de suma de rangos de Wilcoxon.

### Prueba de Chi-cuadrado

La [prueba de Chi-cuadrado de Pearson](https://en.wikipedia.org/wiki/Chi-squared_test) se utiliza para comprobar las diferencias significativas entre grupos categóricos.

## paquete ****rstatix****

El paquete **rstatix** ofrece la posibilidad de ejecutar pruebas estadísticas y recuperar los resultados en un marco "amigable". Los resultados se encuentran automáticamente en un dataframe para que pueda realizar operaciones posteriores con los resultados. También es fácil agrupar los datos que se pasan a las funciones, de modo que las estadísticas se ejecutan para cada grupo.

### Estadísticas resumidas

La función get\_summary\_stats() es una forma rápida de devolver estadísticas de resumen. Sólo tiene que enviar su conjunto de datos a esta función y proporcionar las columnas que desea analizar. Si no se especifica ninguna columna, las estadísticas se calculan para todas las columnas.

Por defecto, se devuelve una gama completa de estadísticas de resumen: n, max, min, mediana, 25%ile, 75%ile, IQR, desviación absoluta mediana (mad), media, desviación estándar, error estándar y un intervalo de confianza de la media.

Puedes especificar un subconjunto de estadísticas de resumen para devolver proporcionando uno de los siguientes valores a type =: "completo", "común", "robusto", "cinco\_números", "media\_sd", "media\_se", "media\_ci", "mediana\_iqr", "mediana\_mad", "cuantil", "media", "mediana", "min", "max".

También puede utilizarse con datos agrupados, de forma que se devuelva una fila por cada variable de agrupación:

También puede utilizar **rstatix** para realizar pruebas estadísticas:

### Prueba T

Utiliza una sintaxis de fórmula para especificar las columnas numéricas y categóricas:

O utiliza ~ 1 y especifique mu = para una prueba T de una muestra. Esto también puede hacerse por grupo.

Si procede, las pruebas estadísticas pueden realizarse por grupos, como se muestra a continuación:

### Prueba de Shapiro-Wilk

Como ya se ha dicho, el tamaño de la muestra debe estar entre 3 y 5000.

### Prueba de suma de rangos de Wilcoxon

### Prueba de Kruskal-Wallis

También conocida como la prueba U de Mann-Whitney.

### Prueba de Chi-cuadrado

La función de prueba chi-cuadrado acepta una tabla, así que primero creamos una tabulación cruzada. Hay muchas formas de crear una tabulación cruzada (véase [Tablas descriptivas](#descriptive-tables)), pero aquí utilizamos tabyl() de **janitor** y eliminamos la columna más a la izquierda de las etiquetas de valores antes de pasarla a chisq\_test().

Se pueden ejecutar muchas más funciones y pruebas estadísticas con las funciones de rstatix. Consulta la documentación de **rstatix** [en línea aquí](https://github.com/kassambara/rstatix) o introduciendo ?rstatix.

## paquete gtsummary

Utiliza **gtsummary** si desea añadir los resultados de una prueba estadística a una tabla bonita creada con este paquete (como se describe en la sección **gtsummary de la** página [Tablas descriptivas](#tbl_gt)).

La realización de pruebas estadísticas de comparación con tbl\_summary se lleva a cabo añadiendo la función add\_p a una tabla y especificando qué prueba utilizar. Es posible obtener valores p corregidos para pruebas múltiples utilizando la función add\_q. Ejecuta ?tbl\_summary para obtener más detalles.

### Prueba de Chi-cuadrado

Compara las proporciones de una variable categórica en dos grupos. La prueba estadística por defecto de add\_p() cuando se aplica a una variable categórica es realizar una prueba de independencia de chi-cuadrado con corrección de continuidad, pero si algún recuento de llamadas esperado es inferior a 5, se utiliza una prueba exacta de Fisher.

### Pruebas T

Compare la diferencia de medias de una variable continua en dos grupos. Por ejemplo, compare la media de edad por resultado del paciente.

### Prueba de suma de rangos de Wilcoxon

Compara la distribución de una variable continua en dos grupos. Por defecto se utiliza la prueba de suma de rangos de Wilcoxon y la mediana (IQR) cuando se comparan dos grupos. Sin embargo, para datos no distribuidos normalmente o para comparar varios grupos, la prueba de Kruskal-wallis es más apropiada.

### Prueba de Kruskal-Wallis

Comparar la distribución de una variable continua en dos o más grupos, independientemente de que los datos se distribuyan normalmente.

## Correlaciones

La correlación entre variables numéricas puede investigarse con el paquete **tidyverse**  
**corrr**. Permite calcular las correlaciones mediante Pearson, Kendall tau o Spearman rho. El paquete crea una tabla y también tiene una función para trazar automáticamente los valores.

## Recursos

Gran parte de la información de esta página está adaptada de estos recursos y viñetas en línea:

[gtsummary](http://www.danieldsjoberg.com/gtsummary/articles/tbl_summary.html) [dplyr](https://dplyr.tidyverse.org/articles/grouping.html) [corrr](https://corrr.tidymodels.org/articles/using-corrr.html) [sthda correlación](http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r)