

PROYECTO

Objetivos

General

Poner en práctica los conocimientos cubiertos en el curso, desde la estructura de datos hasta la visualización de los mismos.

Específicos

- Manipular distintos tipos de datos.
- Aplicar funciones preestablecidas de R.
- Visualizar datos a través de la librería ggplot2.
- Aplicar lógica de programación para resolver algunos apartados del proyecto.

Descripción

La sepsis ocasiona en el mundo más de mil muertes cada día, muchos de estos pacientes adquieren la infección estando hospitalizados y constituye la complicación intrahospitalaria más frecuente. La infección de pacientes por las manos contaminadas del personal de salud es una de las formas de diseminación de los agentes infecciosos. La higiene de las manos es el factor individual más importante para el control de las infecciones, es difícil entender que un hecho tan rutinario en la práctica de la medicina, como lavarse las manos previo a examinar a un paciente, haya causado tanta controversia en la época de 1800.

A principios de la década de 1840 se descubrió una enfermedad mortal que afectaba a las mujeres que acababan de dar a luz, cerca del 10% de las mujeres morían a causa de esta enfermedad y el origen provenía de las manos contaminadas del personal médico que atendía los partos.

En este proyecto, vamos a analizar los datos que demuestran la importancia del lavado de manos, identificando anomalías en los procedimientos médicos.

1. Conociendo los datos

- 1.1. Cargue el paquete *tidyverse*.
- 1.2. Lea el archivo *deaths_handwashing.csv* usando *read_csv* y asígnelo a la variable *yearly*.
- 1.3. Imprima la variable *yearly*.

2. El alarmante número de muertes

La tabla anterior muestra el número de mujeres que dieron a luz durante los años 1841 a 1846 en una ciudad al noreste de Austria, en dos clínicas en particular. Notará que dar a luz era muy peligroso; un número alarmante de mujeres murió como resultado del parto. Vemos esto con mayor claridad si observamos la proporción de muertes con respecto al número de mujeres que dieron a luz.

- 2.1. Use *mutate* para agregar la columna *proportion_deaths* a *yearly*, esta nueva columna se calculará como la proporción de muertes por número de nacimientos.
- 2.2. Imprima *yearly*.

3. Muerte en las clínicas

Si ahora graficamos la proporción de muertes tanto en la clínica 1 como en la clínica 2, veremos un patrón curioso.

- 3.1. Utilice *ggplot* para hacer una gráfica lineal de *proportion_deaths* por año (columna *year*) con una línea por clínica.
- 3.2. Las líneas deben tener colores distintos.

4. Comienza el lavado de manos

¿Por qué la proporción de muertes es mucho más alta en la Clínica 1? La única diferencia en ambas clínicas era que muchos estudiantes de medicina trabajaban en la Clínica 1, mientras que en la Clínica 2 no había tantos. Otro dato interesante es que los médicos que atendían partos solo se encargaban de esa función, mientras que los estudiantes también pasaban tiempo en las salas de autopsias examinando cadáveres.

Se empezó a sospechar que algo en los cadáveres, propagado de las manos de los estudiantes de medicina, causaba la enfermedad. Entonces, en un intento desesperado por detener las altas tasas de mortalidad, se decretó: ¡Lávese las manos en todo momento!

- 4.1. Cargue los datos mensuales de la Clínica 1 para ver si el lavado de manos tuvo algún efecto, para esto lea el archivo *deaths_clinic_1.csv* y asígnelo a la variable *monthly*.
- 4.2. Agregue la columna *proportion_deaths* a *monthly*, esta nueva columna se calculará como la proporción de muertes por número de nacimientos.
- 4.3. Imprima las primeras filas de *monthly* usando la función *head()*.

5. El efecto del lavado de manos

Con los datos cargados, ahora podemos ver la proporción de muertes a lo largo del tiempo.

- 5.1. Haga una gráfica lineal de *proportion_deaths* por fecha para el data frame *monthly* usando *ggplot*.
- 5.2. Utilice la función *labs* para dar al eje *x* y al eje *y* una estética a las etiquetas.
- 5.3. Use *group_by* y *summarise* para encontrar la media por año y guárdelo en el dataset *death_by_year*.
- 5.4. Utilice el conjunto de datos *death_by_year* para crear un gráfico de barras que muestre la media en cada año.

6. Los resultados del lavado de manos

A partir del año 1847 la proporción de muertes se reduce drásticamente y, sí, fue entonces cuando se hizo obligatorio el lavado de manos.

El efecto del lavado de manos se hace aún más claro si lo resaltamos en el gráfico.

- 6.1. Agregue una columna TRUE/FALSE a *monthly* llamado *handwashing_started*, que será TRUE a partir del año en que se impuso el lavado de manos.
- 6.2. Haga una gráfica lineal de *proportion_deaths* por fecha para el data frame *monthly* usando *ggplot*. Haga que el color de la línea dependa de *handwashing_started*.
- 6.3. Utilice la función *labs* para dar al eje *x* y al eje *y* una estética a las etiquetas.

Dado que la columna `monthly$date` es una fecha, puede compararla con otras fechas utilizando los operadores de comparación (`<`, `>`, `=`, `==`, etc.).

7. ¿Más lavado de manos, menos muertes?

Nuevamente, las estadísticas demuestran que lavarse las manos tuvo un efecto enorme. ¿Cuánto redujo la proporción mensual de muertes en promedio?

- 7.1. Utilice `group_by` y `summarise` para calcular el promedio de `proportion_deaths` antes y después de que se impusiera el lavado de manos.
- 7.2. Asigne el resultado en la variable `month_summary` e imprímala.
- 7.3. El data frame resultante debería verse como el siguiente:

<code>handwashing_started</code>	<code>mean_proportion_deaths</code>
TRUE	0...?
FALSE	0...?

8. Análisis estadístico de los datos de lavado de manos

¡Redujo la proporción de muertes en alrededor de 8 puntos porcentuales! Del 10% en promedio antes del lavado de manos a solo el 2% cuando se hizo cumplir el lavado de manos (que sigue siendo un número alto según los estándares modernos). Para tener una idea de la incertidumbre en torno a cuánto se reduce la mortalidad por lavarse las manos, podríamos observar un intervalo de confianza (aquí calculado mediante una prueba t).

Utilice la función `t.test` para calcular un intervalo de confianza del 95%. Los dos grupos que tenemos son proporción mensual de muertes antes y después de que comenzara el lavado de manos "`proportion_deaths`" y "`handwashing_started`".

Tome de referencia la siguiente sintaxis:

```
t.test(outcome ~ group, data = df)
```

Una prueba t produce muchos números, pero lo que interesa es el intervalo de confianza del 95%.

Restricciones

- Debe crear un repositorio en Git y constantemente hacer publicaciones de acuerdo a los resultados que vaya obteniendo.
- El nombre del repositorio debe ser **Proyecto_r**.
- El repositorio debe contener un solo cuaderno con todos los puntos que se describen en el enunciado.
- Los cambios realizados al repositorio después de la fecha de entrega no se tomarán en cuenta.
- Para dudas concernientes al proyecto se utilizará el foro creado en el aula virtual de manera que todos los estudiantes puedan ver las preguntas y posteriores respuestas.
- Las copias totales o parciales obtendrán una nota de 0 puntos.
- No habrá prórroga.

Entrega

- La entrega será el miércoles **14 de septiembre a las 23:59**.
- La entrega se hará en el aula virtual.
- Debe enviar únicamente el enlace del repositorio de Github.