

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 una variable.
- 1.3. Imprima la variable del inciso 1.2.

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 una nueva columna que se calculará como la proporción de muertes por número de nacimientos.
- 2.2. Imprima el dataset con la nueva variable.

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. Tome de referencia la columna creada en el inciso 2.1 (eje y) versus la variable *year* (eje x).
- 3.2. Utilice la variable *clinic* para definir un color distinto en la estética de la gráfica.
- 3.3. Por último, imprima la gráfica y exponga su punto de vista.

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 una nueva variable.
- 4.2. Utilice *ggplot* para hacer una gráfica lineal. La gráfica tendrá la característica de ser dual (dos ejes Y). Para el eje X se utilizará la variable *date* y para Y1 y Y2, serán *births* y *deaths* respectivamente.
- 4.3. Utilice la función *scale_y_continuous* para agregar un título a cada eje Y.
- 4.4. Agregue un título al eje X y un título general para toda la gráfica.
- 4.5. Establezca color verde al comportamiento de la variable *births* y naranja para *deaths*.
- 4.6. Imprima la gráfica resultante.
- 4.7. Construya una opinión con el resultado de la gráfica.

5. El efecto del lavado de manos. Analicemos la proporción de muertes a lo largo del tiempo.

Perspectiva lineal

- 5.1. Agregue una nueva columna a la variable creada en el inciso 4.1. Esta nueva columna se calculará como la proporción de muertes por número de nacimientos. Guarde este cambio en una nueva variable.
- 5.2. Imprima las primeras filas del dataframe del punto anterior. Haga uso de la función *head()*.
- 5.3. Elabore una gráfica lineal utilizando la columna creada en el inciso 5.1 (eje y) versus la columna *date* (eje x).
- 5.4. Utilice la función *labs* para dar al eje x y al eje y una estética a las etiquetas.
- 5.5. Por último, imprima la gráfica y concluya.

Perspectiva de barras

- 5.6. Use *group_by* para la variable *year* y *summarise* para encontrar la media tomando de referencia la variable del inciso 5.1. Guarde el resultado en una nueva variable.
- 5.7. Utilice la variable creada en el inciso anterior para crear un gráfico de barras que muestre la media en cada año.
- 5.8. Imprima la gráfica y concluya.

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 a la variable creada en el inciso 5.1, esta columna será `TRUE` a partir del año en que se impuso el lavado de manos y `FALSE` para años anteriores. Guarde este cambio en una nueva variable.
- 6.2. Construya una gráfica lineal tomando de referencia la variable con la proporción (muertes/nacimientos) para el eje y versus la variable `date` para el eje x. Haga que el color de la línea dependa de la columna creada en el inciso 6.1.
- 6.3. Utilice la función `labs` para dar al eje x y al eje y una estética a las etiquetas.
- 6.4. Imprima la gráfica y concluya.

Dado que la columna `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. Tome de referencia la columna que se agregó en el inciso 6.1 y utilice la función `group_by` para crear un agrupamiento.
- 7.2. En continuidad al inciso anterior, aplique `summarise` para calcular el promedio de la proporción de muertes antes y después de que se impusiera el lavado de manos. Tome de referencia la variable con la proporción (muertes/nacimientos).
- 7.3. Asigne el resultado a una nueva variable e imprímala. El resultado debería verse como el siguiente:

<code><GROUP_BY_INCISO_6.1></code>	<code><SUMMARISE_MEAN_PROPORCION></code>
<code>TRUE</code>	0...?
<code>FALSE</code>	0...?

- 7.4. Utilice `ggplot` para crear un boxplot que tome de referencia la columna creada en el inciso 6.1 (eje x) versus la variable de proporción (muertes/nacimientos).
- 7.5. Imprima la gráfica y concluya en función de los incisos 7.3 y 7.4.

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. Tome de referencia la columna con la proporción (muertes/nacimientos) y la columna creada en el inciso 6.1.

Un ejemplo de utilidad para implementar la función *t.test* es el siguiente:

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

- `outcome`: proporción de muertes/nacimientos.
- `group`: clasificación TRUE/FALSE para antes y después de que se impusiera el lavado de manos.
- `df`: dataset completo.

Interprete los datos resultantes y proporcione una conclusión.

9. Análisis estadístico propio

- 9.1. Aplique funciones como `filter`, `count`, `select` o visualizaciones de tipo histograma o gráfica de puntos, que le permitan descubrir comportamientos o tendencias en los datos. No se limite a estas sugerencias, puede aplicar cualquier otro tipo de función vista en clase.
- 9.2. Para cada análisis estadístico que proporcione concluya en función de los resultados obtenidos.

Restricciones

- El proyecto se debe trabajar individualmente.
- Todos los incisos deben ser desarrollados en un Notebook de R con el nombre **Proyecto.Rmd**. Se recomienda comentar el código como buena práctica para entender la lógica aplicada.
- En el Notebook debe figurar el nombre del estudiante.
- 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 **03 de julio a las 23:59**.
- La entrega se hará en el aula virtual.
- Enviar únicamente el .Rmd.
- La calificación se hará a partir del jueves **04-julio** y finalizará el sábado **06-julio**, para lo cual se compartirá un calendario de calificación.