

PEC4: Contrastes de Hipótesis.

UOC

```
library(dplyr)
library(ggplot2)

load("PEC4/BBDD_COVID2.RData")
```

El reciente problema de salud pública generado por la COVID-19 ha retado a la comunidad científica a identificar los factores de riesgo para el desarrollo de COVID severo. La base de datos que se os presenta contiene una gran cantidad de información anonimizada relacionada con el paciente hospitalizado por COVID-19, incluidas características sociodemográficas y clínicas previas a la infección por SARS-CoV-2. Los datos públicos proceden del siguiente enlace de Kaggle. En esta práctica se trabaja con una selección aleatoria de pacientes hospitalizados ($n = 3000$) y variables de interés ($p = 10$).

Las variables que se encuentran en el dataset son las siguientes:

- *SEX* : Sexo del paciente (Femenino/Masculino).
- *PATIENT_TYPE* : Hospitalización del paciente (Yes/No).
- *AGE* : Edad del paciente.
- *OBESITY* : Paciente con diagnóstico de obesidad (Yes/No).
- *HIPERTENSION* : Paciente con diagnóstico de hipertensión (Yes/No).
- *DIABETES* : Paciente con diagnóstico de diabetes (Yes/No).
- *PNEUMONIA* : Paciente con inflamación pulmonar (Yes/No).
- *ICU* : Ingreso en unidad de cuidados intensivos (Yes/No).
- *INTUBED* : Requerimiento de intubación (Yes/No).
- *DIED* : fallecimiento (Yes/No).

Os puede ser útil consultar el siguiente material:

- Apuntes de contraste de hipótesis
- Apuntes de contraste de dos muestras

Hay que entregar la práctica en fichero pdf o html (exportando el resultado final a pdf o html por ejemplo). Se recomienda generar el informe con Rmarkdown que genera automáticamente el pdf/html a entregar. Se puede utilizar el fichero .Rmd, que disponéis en la PEC, como plantilla para resolver los ejercicios.

Problema 1

- a) Según estudios previos, la edad media de los pacientes hospitalizados por COVID-19 es 53 años. Realiza el contraste de hipótesis sobre si la media de edad de nuestra cohorte es diferente a la media de edad teórica. ¿Podemos rechazar la hipótesis nula de igualdad a un nivel de significación del 0.05?

- b) Otros estudios estiman que la prevalencia de diabetes en pacientes hospitalizados por COVID-19 es del 30%. ¿Hay evidencia suficiente para considerar que la proporción de diabéticos en la cohorte es diferente a la teórica? Realiza el contraste y razona la respuesta.
- c) Existe evidencia de que los pacientes diabéticos tienen un mayor riesgo de requerimiento de soporte ventilatorio invasivo (INTUBED). ¿Qué contraste de hipótesis plantearías para comparar el riesgo entre grupos?
- d) Presenta la tabla de contingencia entre las variables diabetes y soporte ventilatorio invasivo. Además realiza el contraste de hipótesis planteado en el apartado anterior con un nivel de significación del 0.05. (Nota: usa el parámetro `correct = FALSE`)
- e) De acuerdo con el valor p , ¿Podemos decir que existe aumento estadísticamente significativo en la proporción de pacientes con requerimiento de soporte ventilatorio invasivo en el grupo de pacientes con diabetes? Razona la respuesta. Nota: fijamos el nivel de significación al 0.05.
- f) Es conocido que la edad es un factor determinante en el riesgo de intubación. Realiza una comparativa gráfica de la distribución de la edad de los pacientes en función de si requirieron intubación o no. Discute los resultados. (Nota: puedes realizar la comparativa gráfica mediante diagramas de cajas o histogramas).
- g) Estima mediante un intervalo de confianza al 95% la diferencia de media de edad entre los pacientes que fueron intubados respecto aquellos que no. Comenta el resultado. Nota: asume varianzas iguales.
- h) En clave contraste de hipótesis, ¿podemos rechazar la hipótesis nula de igualdad de medias entre grupos? razona la respuesta.

A partir de este momento se decide estudiar en profundidad a los pacientes que requirieron soporte ventilatorio invasivo (INTUBED). Para ello, nos centraremos únicamente en esta subpoblación.

- i) Entre los pacientes que requirieron soporte ventilatorio invasivo, contrasta si la proporción de mortalidad fue diferente entre los pacientes con diabetes y los que no. Comenta los resultados.
- j) Tras el alta hospitalaria los pacientes supervivientes sufren secuelas pulmonares graves. Está demostrado que la capacidad de difusión del monóxido de carbono (DLCO) es una medida sensible al daño pulmonar y su funcionalidad. Interesa saber si seis meses después del alta hospitalaria los pacientes han mejorado este parámetro. Para ello, se les hace una medición el día del alta hospitalaria y seis meses después. Plantea el contraste de hipótesis que realizarías para contrastar si los pacientes mejoraron significativamente. (nota: Mayor valor de DLCO mejor estado pulmonar).