

Práctica Calificada 1

Profesores del curso

11/04/2023

Indicaciones Generales:

- Esta práctica debe ser desarrollada por un grupo de máximo 3 alumnos del mismo horario.
- Justifique claramente sus soluciones, presente los códigos y resultados obtenidos en R, interprete estos resultados y escriba respuestas completas teniendo en cuenta el contexto de los ejercicios propuestos; todo esto influirá en la calificación.
- La presentación, la ortografía y la gramática de los trabajos influirán en la calificación.
- La solución debe ser subida en único archivo PDF en la tarea PC1 de su correspondiente horario del curso. La primera hoja debe presentar el código PUCP, nombre completo y correo electrónico de todos los integrantes del grupo.

Cuestionario

La base de datos `cardio_dataset.csv` contiene las siguientes variables medidas en 68337 pacientes

Al leer este conjunto de datos mediante la función `read.csv()`, utilice el argumento `stringsAsFactors = TRUE`, de manera que las variables cualitativas sean leídas como factores. En este caso la sintaxis sería

```
d=read.csv("cardio_dataset.csv",stringsAsFactors = TRUE)
```

Adicionalmente, ejecute el siguiente código para reordenar el nivel de algunos factores, de manera que se mantenga el orden correcto de las categorías en algunas de las variables ordinales.

Variable	Descripción
id	Identificación de la observación
age	Edad
gender	Género: F = Female, M = Male
height	Altura en centímetros
weight	Peso en kg
ap_hi	Presión sanguínea sistólica en mmHg
ap_lo	Presión sanguínea diastólica en mmHg
cholesterol	Nivel de colesterol: normal, above normal, well above normal
gluc	Nivel de glucosa: normal, above normal, well above normal
smoke	Status de fumador: No, Yes
alco	Consumo de alcohol: No, Yes
active	Actividad física: No, Yes
cardio	Enfermedad cardiovascular: No, Yes
bmi	Índice de masa corporal en kg/m2

```
d$cholesterol = factor(d$cholesterol, levels = levels(d$cholesterol)[c(2,1,3)])  
d$gluc = factor(d$gluc, levels = levels(d$gluc)[c(2,1,3)])
```

Pregunta 1 (3.0 puntos)

Se desea comparar la distribución de frecuencias relativas de fumadores y no fumadores (**smoke**) entre pacientes hombres (**gender=="M"**) y mujeres (**gender=="F"**). Con este fin construya la tabla de distribución de frecuencias para cada grupo y represéntela gráficamente. Interprete sus resultados.

Pregunta 2 (3.0 puntos)

Construya una tabla de frecuencias y un gráfico apropiado para representar la distribución de la variable peso de los pacientes (**weight**). Indique que tipo de asimetría presentan los datos, justifique su respuesta mediante el gráfico realizado y una medida estadística apropiada.

Pregunta 3 (3.0 puntos)

Para un estudio se van a considerar solamente la información de mujeres (**gender=="F"**) cuyo índice de masa corporal (**bmi**) haya estado dentro del 15% más altas de su grupo, y de hombres cuyo índice de masa corporal (**bmi**) haya estado dentro del 10% más altas de su grupo.

¿Qué índice de masa corporal debe superar como mínimo una mujer para que sea incluida en el estudio?
¿Qué índice de masa corporal debe superar como mínimo un hombre para que sea incluida en el estudio?
Justifique sus respuestas.

Pregunta 4 (3.0 puntos)

Elabore un gráfico de boxplot para presión sanguínea sistólica (**ap_hi**) para los niveles de la variable nivel de colesterol (**cholesterol**). A partir de sus resultados, realice una comparación en términos de tendencia central, dispersión y asimetría de la presión sanguínea sistólica por nivel de colesterol.

Pregunta 5 (3.0 puntos)

Calcule las estadísticas necesarias y realice una comparación en términos de tendencia central, dispersión y asimetría de la presión sanguínea diastólica (**ap_lo**) para fumadores y no fumadores (**smoke**). Indique que medidas de tendencia central y de dispersión serían las más adecuadas para realizar una comparación.

Pregunta 6 (5.0 puntos)

Evalúe la veracidad o falsedad de cada una de las siguientes afirmaciones. Justifique su respuesta (presente sus códigos y resultados en R como parte de su justificación)

- (1.0 punto) Para representar la tendencia central de la variable peso (**weight**) es adecuado usar la mediana.
- (1.0 punto) Se puede afirmar que la mediana del nivel de colesterol (**cholesterol**) es **above normal**.
- (1.0 punto) Se aplicó la función **summary** a la variable altura (**height**) y a partir de los resultados obtenidos, se llegó a la conclusión que aproximadamente el 50% de las personas tiene una altura mayor a 164.5 *cm*.

- d) (1.0 punto) Se puede afirmar que la presión sanguínea sistólica (**ap_lo**) mediana es la misma para fumadores y no fumadores (**smoke**).
- e) (1.0 punto) La variable Índice de masa corporal (**bmi**) presenta una menor dispersión que edad (**age**).