

Práctica Calificada 2

Profesores del curso

19/09/2023

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

Pregunta 1 (4.0 puntos)

Un investigador estudia el uso de un molino de viento para generar electricidad. La base de datos `viento.csv` contiene las siguientes variables medidas en varios experimentos con un molino de viento:

| Variable | Descripción |
|-----------------------|--|
| <code>velocity</code> | Velocidad del viento en mph |
| <code>dc</code> | Corriente directa producida en voltios |

- (1.0 punto) Estudie la posible relación lineal de la variable velocidad del viento con la corriente directa producida por el molino. Justifique su respuesta mediante gráficos y medidas apropiadas.
- (1.0 punto) Estime un modelo de regresión para predecir la corriente producida por el molino en función de la velocidad del viento. Interprete los coeficientes de regresión estimados.
- (1.0 punto) Considere ahora que un analista sugiere que se use como variable para predecir la corriente producida por el molino, la inversa de la velocidad del viento ($1/\text{velocity}$) y compárelo con el modelo hallado en el ítem anterior. Indique que modelo sería el más adecuado. Justifique su respuesta en base a sus resultados.
- (1.0 punto) Considerando el mejor modelo encontrado encontrado en c), estime la corriente producida por el molino con una velocidad del viento de 6 mph.

Pregunta 2 (4.0 puntos)

Para esta pregunta se considera los datos de 311 individuos de una muestra realizada para un estudio de obesidad.

Los datos se encuentran disponibles en el archivo `datos_bmi.csv` en donde se recolectaron las siguientes variables

| Variable | Descripción |
|-----------|-------------------------------|
| bmi | Índice de masa corporal (IMC) |
| sexo | Sexo |
| edad | Edad |
| edad.papa | Edad del padre |
| edad.mama | Edad de la madre |
| bmi.mama | IMC de la madre |
| bmi.papa | IMC del padre |

- a) (1.0 punto) Si se seleccionan al azar y sin reemplazo 12 individuos de la muestra, calcule la probabilidad que **exactamente dos de ellos tengan un IMC que se considere atípico entre todos los individuos del estudio**, según el criterio del diagrama de cajas (boxplot).
- b) (1.5 puntos) Si se selecciona al azar y sin reemplazo 8 individuos de la muestra, calcule la probabilidad que se hayan seleccionado más mujeres que hombres.
- c) (1.5 puntos) Si se selecciona al azar y sin reemplazo 10 individuos de la muestra, calcule la probabilidad que se haya seleccionado **una mujer con IMC que sea considerado atípico entre las mujeres del estudio** o que se haya seleccionado **un hombre con IMC considerado atípico entre los hombres del estudio**.