



Pontificia Universidad
JAVERIANA
Bogotá

MAESTRÍA EN 
EPIDEMIOLOGÍA
CLÍNICA

BIOESTADÍSTICA AVANZADA

MÓDULO I

Regresión lineal

Examen parcial: Bloque 1

Indicaciones generales:

Se presenta un escenario de aplicación con su base de datos correspondiente. El estudiante debe desarrollar, con base en esta información, los cuatro ejercicios que se indican a continuación. Debe escribir un documento en un archivo de MS Word con sus respuestas siguiendo estas indicaciones:

- Cuando sea necesario, trabaje con un nivel de significancia del $\alpha = 0.05$.
- No incluya código del programa estadístico que utilizó.
- No incluya valores de resultados sin su interpretación.

La evaluación tiene una duración de máximo 3 horas. No se recibirán archivos de respuestas posterior a la hora de cierre de la evaluación.

El escenario:

En una muestra de 255 sujetos se evaluaron las siguientes variables: peso (kg), talla (cm), edad (años), sexo (*hombre* = 0, *mujer* = 1) y actividad (número de horas de actividad física por semana). Las variables talla y edad fueron centradas restando a cada valor observado el promedio general de los 255 sujetos de cada variable respectivamente. Los datos obtenidos se encuentran en el archivo adjunto: "datos.xlsx". Con base en esta información, desarrolle los siguientes ejercicios.

Ejercicio 1 (1.5 puntos)

Ajuste un modelo de regresión lineal múltiple tomando como variable desenlace el *peso* y como variables independientes la *talla*, *edad*, *sexo* y *actividad*. Sobre los resultados del ajuste de este modelo, interprete:

- a. Los coeficientes β_{vs} del modelo.
- b. Las pruebas F parciales de cada uno de los coeficientes.
- c. El coeficiente de determinación ajustado R_{adj}^2 .
- d. La evaluación de la hipótesis nula $H_0: \beta_{edad} = \beta_{sexo} = 0$.

Ejercicio 2 (1.5 puntos)

Con relación al diagnóstico del modelo de regresión del peso en función de la talla, edad, sexo y actividad física ajustado en el punto anterior, responder:

- a. Basado en la **distancia de Cook**, ¿cuáles sujetos presentan datos atípicos?

Retire de la base de datos los sujetos con observaciones atípicas identificadas por la distancia de Cook. Ajuste nuevamente el modelo y continúe respondiendo los siguientes ítems basado en este nuevo modelo.

- b. ¿Se cumple el supuesto de normalidad sobre los residuales?
- c. ¿Se cumple el supuesto de Homocedasticidad?
- d. ¿Se pueden presentar problemas de colinealidad?, ¿entre que variables?

Ejercicio 3 (1 punto)

En el contexto de la construcción de un modelo de predicción, aplique el método de selección de variables hacia atrás (backward) al modelo final ajustado en el punto anterior. ¿Qué variables se incluyen en el modelo final?

Ejercicio 4 (1 punto)

En el contexto del ajuste de un modelo para evaluar la asociación entre una exposición x_e y un desenlace y , responder las siguientes preguntas:

- a. ¿La variable *sexo* es una variable modificadora del efecto de la relación entre la $x_e = talla$ y el desenlace $y = peso$? (No considere el resto de variables independientes en este análisis). Justifique su respuesta.

Para el siguiente ítem, asuma que las variables edad, sexo y actividad no son variables modificadoras del efecto entre la talla y el peso.

- b. ¿Las variables *edad*, *sexo* y *actividad*, son variables confusoras de la relación entre las $x_e = talla$ y el $y = peso$? Justifique su respuesta.