

Departamento de Estadística y Matemáticas  
Facultad de Ciencias Económicas  
Estadística II  
Parcial IV

Nombre: \_\_\_\_\_ Cédula: \_\_\_\_\_

1. **(5 puntos)** Con el objetivo de promover la investigación y la práctica en el análisis de datos socioeconómicos en Colombia, se ha generado una base de datos sintética que representa una muestra diversa de la población colombiana. Esta base de datos contiene una amplia gama de variables demográficas, socioeconómicas y laborales, incluyendo estimaciones de ingresos expresadas en Salarios Mínimos Mensuales Legales Vigentes (SMMLV). Diseñada para ser un recurso educativo y exploratorio, esta base de datos permite a estudiantes, investigadores y analistas practicar técnicas de modelado estadístico, explorar las dinámicas del mercado laboral colombiano y comprender mejor los factores que influyen en los niveles de ingreso de la población en Colombia.

La base de datos contiene los siguientes campos:

Variable Name	Description
Edad	Edad de la persona en años.
Anios_Educacion	Años de educación formal completados.
Horas_Trabajo_Semana	Horas trabajadas por semana.
Experiencia_Laboral	Años de experiencia laboral.
Personas_Hogar	Número de personas que viven en el hogar.
Ninos_Hogar	Número de niños (menores de 18 años) en el hogar.
Deuda_Total	Monto total de deudas en pesos colombianos.
Ahorros	Monto total de ahorros en pesos colombianos.
Estrato_Numerico	Estrato socioeconómico representado numéricamente.
Tiempo_Comunidad	Años viviendo en la comunidad actual.
Distancia_Trabajo_KM	Distancia al lugar de trabajo en kilómetros.
Calificacion_Profesional	Calificación numérica de habilidades profesionales (auto-reportada).
Inversion_Educativa_Anual	Inversión anual en educación (propia o de hijos) en pesos colombianos.
Creditos_Bancarios	Número de créditos bancarios activos.
Servicios_Publicos_Basicos	Índice de acceso a servicios públicos básicos (agua, luz, gas, etc.).
Motivacion_Personal	Variable de ruido representando motivación intrínseca y ambición personal.
Suerte_Personal	Variable de ruido representando eventos aleatorios y suerte personal.
Departamento	Departamento de Colombia donde reside la persona.
Genero	Género auto-reportado de la persona.
Nivel_Educativo	Nivel educativo más alto alcanzado.
Tipo_Empleo	Tipo de empleo o situación laboral actual.
Acceso_Internet	Indica si la persona tiene acceso a internet en el hogar.
Estado_Civil	Estado civil actual de la persona.
Ingresos_SMMLV	Ingresos mensuales estimados en Salarios Mínimos Mensuales Legales Vigentes.

Basado en la información contenida en la base de datos suministrada, al crear una subpoblación de la base de datos mediante la aplicación de **al menos 2 condicionales (Aplique a la base de datos al menos 2 filtros, antes de trabajar con ella el parcial)** .

- a) **(0.5 puntos)** Si tuviera que plantear una relación lineal entre la variable `Distancia_Trabajo_KM` y otra variable de la forma

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

cómo plantearía dicha relación desde sus racionalidad?. Seleccione **tres posibles variables explicativas**  $X$  y explique el por qué estas variables podrían explicar el comportamiento de la variable respuesta  $Y$ .

- b)* **(0.5 puntos)** Basado en el planteamiento que realizó en el inciso anterior, justifique cómo esperaría usted que fuese el signo de los parámetros  $\beta_0$  y  $\beta_1$  del modelo, al usar de forma individual cada una de las **tres variables explicativas**  $X$  propuestas.
- c)* **(0.5 puntos)** Realice el cálculo de los estimadores para los parámetros  $\beta_0$  y  $\beta_1$ , para cada una de las **tres relaciones propuestas** e interprete éstos en el contexto de los datos. Los resultados obtenidos fueron consistentes con lo que esperaba en el inciso anterior?.
- d)* **(0.5 puntos)** Pruebe cada una de las **tres relaciones propuestas** la significancia estadística de los parámetros  $\beta_0$  y  $\beta_1$ , empleando para ello un nivel de significancia del 5 %, e interprete en el contexto de los datos.
- e)* **(0.5 puntos)** Para cada una de las **tres relaciones propuestas**, construya intervalos de confianza para los parámetros  $\beta_0$  y  $\beta_1$ , empleando para ello un nivel de confianza del 95 %, e interprete en el contexto de los datos.
- f)* **(0.5 puntos)** Dados los resultados obtenidos hasta el momento, explique cuál de las tres variable explicativa parece ser la más adecuada para describir la relación con la variable `Distancia_Trabajo_KM`?

**NOTA:** A partir de este punto los análisis solo se tienen que realizar con la variable explicativa que haya considerado como la más adecuada para describir la relación con la variable respuesta `Distancia_Trabajo_KM`.

- g)* **(0.5 puntos)** Verifique si se cumplen o no los supuestos del modelo de regresión lineal planteado, e interprete los resultados.
- h)* **(0.5 puntos)** Pruebe la significancia estadística de la regresión lineal planteada, empleando para ello un nivel de significancia del 5 %, e interprete en el contexto de los datos.
- i)* **(0.5 puntos)** Realice el cálculo del coeficiente de determinación  $R^2 = \frac{SCR}{SCT}$  asociado a la regresión lineal planteada, e interprete el resultado obtenido. Dicho resultado es consistente con lo que se concluyó en el inciso anterior?
- j)* **(0.5 puntos)** Seleccione tres valores para  $x_0$  entre los posibles que considera que puede tomar la variable que escogió como  $X$ , y con éste, construya un intervalo de predicción del 95 % para la variable que escogió como  $Y$  e interprete en el contexto de los datos.