

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

Nombre: \_\_\_\_\_ Cédula: \_\_\_\_\_

1. **(5 puntos)** En el marco de la transformación digital de los mercados de consumo, las plataformas de delivery de alimentos han revolucionado los patrones de consumo urbano en América Latina. Estas plataformas representan un caso fascinante de economía de dos lados (two-sided markets), donde convergen restaurantes, repartidores y consumidores en un ecosistema digital que genera millones de transacciones diarias. El crecimiento explosivo de este sector, acelerado por la pandemia de COVID-19, ha creado nuevas dinámicas de consumo que requieren un análisis profundo para comprender los determinantes del comportamiento del consumidor y optimizar las estrategias de negocio.

Con el objetivo de analizar los factores que influyen en el comportamiento de consumo en plataformas de delivery, se ha generado una base de datos sintética que representa una muestra diversa de usuarios activos de aplicaciones de entrega de alimentos en ciudades principales de Colombia durante el año 2024. Esta base de datos contiene información demográfica, comportamental, transaccional y contextual que permite explorar las relaciones entre diferentes variables socioeconómicas y los patrones de gasto en estas plataformas digitales.

El análisis de estos datos es crucial para economistas y administradores, ya que permite entender la elasticidad de la demanda en mercados digitales, identificar segmentos de clientes rentables, optimizar estrategias de pricing dinámico, y diseñar programas de fidelización efectivos. Además, proporciona insights valiosos sobre la economía del comportamiento aplicada a decisiones de consumo en tiempo real y bajo diferentes contextos de incentivos.

La base de datos contiene los siguientes campos:

| Variable               | Descripción                         |
|------------------------|-------------------------------------|
| edad                   | Edad del usuario                    |
| genero                 | Género del usuario                  |
| ciudad                 | Ciudad de residencia                |
| estrato                | Estrato socioeconómico              |
| nivel_educativo        | Máximo nivel educativo alcanzado    |
| estado_civil           | Estado civil del usuario            |
| ocupacion              | Tipo de ocupación principal         |
| modalidad_trabajo      | Modalidad de trabajo predominante   |
| tiene_hijos            | Indica si el usuario tiene hijos    |
| num_hijos              | Número de hijos menores de 18 años  |
| ingreso_mensual        | Ingreso mensual del hogar           |
| gasto_mensual_delivery | Gasto total mensual en delivery     |
| frecuencia_pedidos_mes | Número de pedidos realizados al mes |
| ticket_promedio        | Valor promedio por pedido           |
| categoria_preferida    | Categoría más frecuente de pedidos  |
| horario_preferido      | Horario más frecuente de pedidos    |
| dia_semana_frecuente   | Tipo de día con más pedidos         |
| tiempo_usuario_meses   | Antigüedad como usuario de la app   |
| suscripcion_premium    | Tiene suscripción premium/plus      |
| uso_descuentos         | Porcentaje de pedidos con descuento |

| Variable                | Descripción                               |
|-------------------------|---|
| calificacion_promedio   | Calificación promedio que da el usuario   |
| tiempo_entrega_promedio | Tiempo promedio de entrega esperado       |
| distancia_promedio_km   | Distancia promedio del restaurante        |
| metodo_pago_principal   | Método de pago más utilizado              |
| dispositivo_principal   | Dispositivo más usado para pedidos        |
| pedidos_cancelados_mes  | Promedio de pedidos cancelados por mes    |
| diversidad_restaurantes | Número de restaurantes diferentes por mes |
| pedidos_grupo           | Porcentaje de pedidos para 2+ personas    |
| propina_promedio_pct    | Porcentaje de propina sobre el pedido     |
| pedidos_madrugada       | Porcentaje pedidos después de medianoche  |

Antes de realizar cualquier análisis estadístico, los estudiantes deben aplicar AL MENOS 2 filtros sobre variables categóricas para crear una subpoblación específica. Las variables para las que debes establecer un filtro son: genero, suscripcion\_premium.

Esto permitirá análisis más específicos y relevantes para segmentos particulares del mercado.

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

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

cómo plantearía dicha relación desde sus racionalidad?. Seleccione **una variable explicativa**  $X$  y explique el por qué esta variable 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 la **variable explicativa**  $X$  propuestas.
- c) **(0.5 puntos)** Realice el cálculo de los estimadores para los parámetros  $\beta_0$  y  $\beta_1$ , para la **relación propuesta** 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 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)** Construya un intervalo 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) **(1 punto)** Verifique si se cumplen o no los supuestos del modelo de regresión lineal planteado, e interprete los resultados.
- g) **(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.
- h) **(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?
- i) **(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.