

**FACULTAD DE INGENIERÍA Y ARQUITECTURA**

**CARRERA DE INGENIERÍA DE SISTEMAS**

**ESTADÍSTICA APLICADA**

**2023 – 1**

**Guía de ejercicios**

**Introducción a los datos – Asociación entre variables**

1. Se encuestó a 245 estudiantes de la universidad “Santander”. Se sabe que en esta universidad el 72% del alumnado pertenece al área de Humanidades y el resto al área de Ciencias. Después de procesar la información muestral se obtuvo que:

* 56% eran mujeres,
* el 33.2% prefiere el rock como género musical,
* el 45.7% gusta de la emisora 104.7 VIVA F.M.
* La media del número de veces que un oyente llamó a la radio en el último mes fue de 2.86 llamadas y la moda de esa misma variable fue igual a uno.

Para la información identifique las variables de estudio, el tipo de cada variable y su gráfico respectivo.

1. Determinar si los siguientes enunciados son V o F.
2. La característica método de pago (débito, transferencia, efectivo) es de escala nominal.
3. Si al realizar un diagrama de dispersión entre dos variables cuantitativas se observa una relación exponencial, entonces es mejor calcular el coeficiente de correlación de Spearman.
4. Al calcular la media y mediana de la variable ingresos de un empleado se tiene que la media es mucho mayor que la mediana. Entonces, los ingresos de un empleado presentan una distribución asimétrica negativa.
5. Para relacionar el método de pago (débito, transferencia, efectivo) con la zona donde reside el cliente (norte, sur, este, oeste) una de las medidas estadísticas más adecuadas de usar es el estadístico Cramer’s V.
6. Para comparar el grado de homogeneidad entre el ingreso de un empleado y sus respectivas edades se emplea la varianza.
7. El encargado de una tienda registró el tiempo (minutos) que los clientes estuvieron en la tienda desde el momento en que tomaron un número hasta que se fueron. Se tienen los siguientes datos para una muestra de 16 clientes:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 15 | 14 | 16 | 14 | 14 | 14 | 13 | 8 |
| 12 | 9 | 7 | 17 | 10 | 15 | 16 | 16 |

1. ¿Qué gráfico o gráficos exploratorios permiten describir estos datos?
2. Calcular e interpretar las medidas de tendencia central (media, mediana y moda).
3. Calcular e interpretar las medidas de dispersión rango intercuartil y desviación estándar.
4. ¿Cuánto debería ser el tiempo mínimo que pasa un cliente en la tienda para pertenecer al 18% de los clientes que pasan mayor tiempo en la tienda?
5. El tiempo máximo que pasa un cliente en la tienda para pertenecer al 12% de los clientes que pasan menor tiempo en la tienda será:
6. ¿Qué tipo de distribución presenta esta característica? *Considerar calcular el coeficiente de asimetría y curtosis.*
7. El área de recursos humanos de una pequeña empresa considera que existe una relación entre el ausentismo (días) y la edad (años) del empleado. Para verificar esto, selecciona una muestra aleatoria de 8 empleados y obtiene los siguientes datos:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Días ausente** | 15 | 6 | 10 | 18 | 9 | 7 | 14 | 11 |
| **Edad** | 27 | 61 | 37 | 23 | 46 | 58 | 29 | 36 |

¿Qué se puede comentar al respecto? Responder de manera visual y cuantitativa.

1. Se tiene la siguiente matriz de correlación de 5 variables cuantitativas:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **A** | **B** | **C** | **D** | **E** |
| **A** | 1 | 0.58 | -0.34 | 0.22 | 0.06 |
| **B** | 0.58 | 1 | 0.10 | -0.12 | -0.65 |
| **C** | -0.34 | 0.10 | 1 | 0.79 | -0.02 |
| **D** | 0.22 | -0.12 | 0.79 | 1 | 0.18 |
| **E** | 0.06 | -0.65 | -0.02 | 0.18 | 1 |

1. ¿Entre qué variables existe una mayor asociación?
2. ¿Entre qué variables existe una menor asociación?
3. ¿Cuáles variables presentan la mayor asociación positiva?
4. ¿Cuáles variables presentan la mayor asociación negativa?
5. Una pizzería local desea conocer un poco más sobre el comportamiento de sus pedidos. Para esto, selecciona una muestra aleatoria de 9 pedidos realizados la última semana.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Orden** | App | Web | Web | Web | App | App | App | Web | App |
| **Distancia de traslado (km)** | 7.5 | 13.3 | 4.7 | 14.6 | 8.4 | 12.6 | 6.2 | 10.5 | 9.8 |
| **Tiempo de entrega (min)** | 25 | 32 | 26 | 38 | 34 | 41 | 36 | 37 | 31 |
| **Satisfacción de entrega** | 5 | 3 | 2 | 2 | 2 | 3 | 1 | 4 | 4 |

1. Identificar el tipo de cada una de las características recopiladas.
2. ¿Qué tipo de distribución presenta la distancia de traslado del pedido? *Considerar calcular el coeficiente de asimetría.*
3. ¿Cuál debe ser el tiempo de entrega máximo del pedido para pertenecer al 10% de los pedidos que se entregan más rápido?
4. ¿La distancia de traslado presenta datos más homogéneos que el tiempo de entrega?
5. Cuantificar e interpretar la asociación entre el tiempo de entrega del pedido y la distancia de traslado.
6. ¿El nivel de satisfacción de la entrega del pedido está relacionado a la forma de orden del pedido? Sustentar su respuesta.

Los mismos datos se encuentran en el archivo “Pizzeria.xlsx” o “Pizzeria.csv”.

1. Los datos presentados a continuación corresponden a los montos y el tipo de pago realizados por una muestra de 100 clientes de una cadena de supermercados. Los directivos de esta empresa pidieron tomar la muestra para conocer el comportamiento de los **pagos de los clientes**.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cheque** | | | | | | | | | |
| 5.00 | 17.87 | 18.77 | 21.11 | 22.67 | 25.96 | 30.60 | 31.07 | 31.74 | 34.67 |
| 35.38 | 36.09 | 36.48 | 37.20 | 37.60 | 37.94 | 38.58 | 39.55 | 40.51 | 41.10 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Efectivo** | | | | | | | | | |
| 1.09 | 1.27 | 1.85 | 2.44 | 2.87 | 2.96 | 3.31 | 3.65 | 4.34 | 4.75 |
| 5.08 | 5.15 | 5.91 | 5.98 | 6.93 | 7.02 | 7.17 | 7.22 | 7.40 | 7.41 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tarjeta** | | | | | | | | | |
| 14.44 | 19.78 | 22.59 | 25.57 | 26.57 | 26.91 | 27.66 | 27.89 | 33.76 | 44.53 |
| 46.13 | 46.24 | 48.11 | 50.30 | 52.35 | 52.63 | 53.32 | 54.19 | 55.21 | 57.55 |

* 1. Para las tres formas de pago halle e interprete la media, la mediana y la moda
  2. ¿Cuál de las tres formas de pago presenta mayor homogeneidad?
  3. Para las tres formas de pago halle e interprete el coeficiente de asimetría

1. Novecientos cincuenta escolares se clasificaron de acuerdo a sus hábitos alimenticios y a su coeficiente intelectual:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Estado de nutrición** | Coeficiente intelectual | | | |
| < 80 | 80 - 90 | 90 - 99 | ≥100 |
| Nutrición buena | 245 | 228 | 177 | 219 |
| Nutrición pobre | 31 | 27 | 13 | 10 |

Indique si existe asociación entre las variables y la fuerza de asociación

1. Un investigador está interesado en verificar si el rendimiento en una planta de un grupo de personas está asociado con la antigüedad en la planta. En el estudio se eligieron al azar 10 trabajadores y se registraron los siguientes datos:

|  |  |
| --- | --- |
| **Rendimiento** | **Antigüedad** |
| Muy bajo | 10 |
| Muy alto | 15 |
| Bajo | 18 |
| Medio | 22 |
| Bajo | 18 |
| Muy bajo | 12 |
| Medio | 15 |
| Medio | 14 |
| Alto | 25 |
| Medio | 22 |

Calcule e interprete la medida de asociación más adecuada entre el rendimiento y la antigüedad en la planta.

1. Al gerente de comercialización de una cadena grane de supermercados le gustaría determinar si hay correlación entre el espacio (en pies) en estantes sobre las ventas (en cientos de soles) de comida para mascotas. Se selecciona una muestra aleatoria de 12 supermercados de igual tamaño y los resultados se presentan a continuación:

|  |  |  |
| --- | --- | --- |
| **Tienda** | **Espacio** | **Ventas** |
| 1 | 5 | 1.6 |
| 2 | 5 | 2.2 |
| 3 | 5 | 1.4 |
| 4 | 10 | 1.9 |
| 5 | 10 | 2.4 |
| 6 | 10 | 2.6 |
| 7 | 15 | 2.3 |
| 8 | 15 | 2.7 |
| 9 | 15 | 2.8 |
| 10 | 20 | 2.6 |
| 11 | 20 | 2.9 |
| 12 | 20 | 3.1 |

Halle e interprete el coeficiente de correlación para las variables en estudio