

Curso: Análisis de datos

Práctico 3 Herramientas estadísticas

En este práctico se trabajará con una serie de ejercicios breves. Se deben responder las preguntas que se plantean en cada ejercicio con un breve comentario con la justificación.

Se debe entregar un PDF con una tabla con los resultados de cada punto y un archivo con el código empleado. La tabla debe tener las columnas: Nro Pregunta, Respuesta, Comentario.

1. La universidad ha encuestado a sus estudiantes para averiguar el tiempo semanal medio que dedican a navegar por Internet.
 - a) ¿Cuál es la población?
 - b) ¿Cuál es la muestra?
2. Suponga que es dueño de una tienda de alimentación.
 - a) Proponga un ejemplo de una pregunta que podría responderse utilizando estadística descriptiva.
 - c) Proponga un ejemplo de una pregunta sobre una posible relación entre dos variables que tienen interés para su tienda de alimentación.
 - d) Proponga un ejemplo en la que haya que hacer una predicción.
3. La oficina de relaciones públicas de un equipo de básquetbol profesional quiere información sobre los aficionados que acuden a los partidos después de la temporada. En los partidos que se festejan después de la temporada, se entrega a la entrada un cuestionario a cada aficionado. ¿Es la respuesta a cada una de las siguientes preguntas categórica o numérica? Si es categórica, indique el nivel de medición (nominal u ordinal). Si es numérica, ¿es discreta o continua?
 - a) ¿Tiene usted una entrada de temporada?
 - b) ¿Vive en el estado de Orange?
 - c) ¿Cuánto le costó realmente la entrada para este partido de después de temporada?
 - d) ¿Cuál es su nivel de satisfacción respecto a la experiencia de compra?

4. Se han formulado una serie de preguntas a una muestra de clientes de una tienda de helados. Identifique el tipo de dato (categórico o numérico continuo o discreto) que se pide en cada pregunta.
- a) ¿Cuál es su sabor favorito?
 - b) ¿Cuántas veces al mes toma helado?
 - c) ¿Tiene hijos de menos de 10 años que vivan en casa?
 - d) ¿Ha probado el último sabor de helado?
5. En una muestra aleatoria de 5 semanas se observó que una agencia de cruceros recibía el siguiente número de programas semanales especiales de cruceros al Caribe:
- 20, 73, 75, 80, 82
- a) Calcule la media, la mediana y la moda.
 - b) ¿Qué medida de la tendencia central describe mejor los datos?
6. El director de unos grandes almacenes tiene interés en saber cuántas reclamaciones recibe el departamento de atención al cliente sobre la calidad de los aparatos eléctricos que venden los almacenes. Los registros de un periodo de 5 semanas muestran el siguiente número de reclamaciones semanales:
- 13, 15, 8, 16, 8
- a) Calcule el número medio de reclamaciones semanales.
 - b) Calcule el número mediano de reclamaciones semanales.
 - c) Halle la moda.
7. Los porcentajes de la remuneración total correspondientes al pago de pluses de una muestra de 12 altos ejecutivos son los siguientes:
- 15,8 17,3 28,4 18,2 15,0 24,7 13,1 10,2 29,3 34,7 16,9 25,3
- a) Calcule la mediana muestral.
 - b) Calcule la media muestral.
8. Calcule la varianza y la desviación estándar de los siguientes datos muestrales:
- 6, 8, 7, 10, 3, 5, 9, 8
9. 3.14. Calcule el coeficiente de variación de los siguientes datos muestrales:
- 10, 8, 11, 7, 9

- 10.** Considere la siguiente muestra de cinco valores y las ponderaciones correspondientes:

X_i	w_i
4,6	8
3,2	3
5,4	6
2,6	2
5,2	5

- a) Calcule la media aritmética de los x_i valores sin ponderaciones.
b) Calcule la media ponderada de los x_i valores.

- 11.** Considere la siguiente distribución de frecuencias de una muestra de 40 observaciones:

Clase	Frecuencia
0-4	5
5-9	8
10-14	11
15-19	9
20-24	7

- a) Calcule la media muestral.
b) Calcule la varianza muestral y la desviación estándar muestral.

- 12.** Halle la renta personal media per cápita ponderada de la siguiente muestra aleatoria de siete estados de Estados Unidos.

Estado	Población	Ingreso per cápita
Alabama	4.500.752	26.338
Georgia	8.684.715	29.442
Illinois	12.653.544	33.690
Indiana	6.195.643	28.783
Nueva York	19.190.115	36.574
Pensilvania	12.365.455	31.998
Tennessee	5.841.748	28.455

- 13.** Un fabricante de radios portátiles obtuvo una muestra de 50 radios de la producción de una semana. Las radios se comprobaron minuciosamente y el número de defectos encontrados fue el siguiente:

Número de defectos	Número de radios
0	12
1	15
2	17
3	6

Calcule la desviación estándar.

- 14.** La tabla adjunta muestra la cantidad de tiempo (en horas) dedicada a estudiar para un examen por una muestra aleatoria de 25 estudiantes de una clase numerosa.

Horas de estudio	Número de estudiantes
0 a 4	3
4 a 8	7
8 a 12	8
12 a 16	5
16 a 20	2

- a) Estime la media muestral del tiempo de estudio.
b) Estime la desviación estándar muestral.

- 15.** A continuación, se presenta una muestra aleatoria de siete pares (x, y) de puntos de datos:

$(1, 5)$ $(3, 7)$ $(4, 6)$ $(3, 6)$ $(5, 7)$ $(5, 8)$

b) Calcule el coeficiente de correlación.

a) Calcule la covarianza.

- 16.** Un hospital tiene interés en averiguar la eficacia de un nuevo medicamento para reducir el tiempo necesario para recuperarse totalmente de una operación de rodilla. La recuperación total se mide por medio de una serie de tests de fuerza que comparan la rodilla operada con la rodilla sin operar. El medicamento se administró en dosis diferentes a 18 pacientes durante un periodo de 6 meses. Los datos (x, y) siguientes indican el número de unidades de medicamento, X , y los días necesarios para la recuperación total Y de cada paciente:

(5, 53) (4, 56) (9, 54) (14, 56) (21, 65) (7, 53) (7, 56) (9, 59) (14, 48) (11, 66)
(21,57) (17, 49) (9, 53) (21, 52) (4, 56) (9, 46) (14, 66) (13, 49)

- Calcule la covarianza.
- Calcule el coeficiente de correlación.
- Analice brevemente la relación entre el número de unidades de medicamento y el tiempo de recuperación. ¿Qué dosis deberíamos recomendar basándonos en este análisis inicial?

- 17.** Acme Delivery ofrece tres tarifas distintas de envío de paquetes de menos de 5 libras de Maine a la costa oeste: ordinario, 3 \$; urgente, 5 \$, y superurgente, 10\$. Para comprobar la calidad de estos servicios, un importante minorista de venta por correo envió 15 paquetes de Maine a Tacoma (Washington) en momentos elegidos aleatoriamente. Los paquetes fueron enviados en grupos de tres por los tres servicios al mismo tiempo para reducir las diferencias resultantes del día del envío. Los datos siguientes muestran el coste de envío, X , y el número de días, Y , en pares (x, y) :

(3,7) (3, 6) (3, 5) (5,5) (10,2) (3,9) (5,6) (10,5) (5, 6) (10, 1) (3, 10) (5, 7) (10, 4)
(5,6) (10, 4)

- Describa los datos numéricamente (covarianza y correlación).
- Analice el valor de los servicios de precio más alto desde el punto de vista del envío más rápido.

- 18.** Realizar un análisis de tendencia central, desviación estándar y de relación entre variables utilizando el dataset proporcionado. Explicar los resultados obtenidos.

- 19.** El número de computadores vendidos al día en una tienda viene definido por la siguiente distribución de probabilidad:

X	0	1	2	3	4	5	6
$P(x)$	0,05	0,10	0,20	0,20	0,20	0,15	0,10

- ¿ $P(3 \leq x \leq 6)$?
- ¿ $P(x > 3)$?
- ¿ $P(x \leq 4)$?
- ¿ $P(2 < x \leq 5)$?

- 20.** Considera la siguiente distribución de probabilidad que representa el número de personas que llegan a una estación de tren por hora:

X	0	1	2	3	4	5
P(x)	0,10	0,15	0,20	0,25	0,20	0,10

- ¿Cuál es la probabilidad de que lleguen al menos 2 personas en una hora?
- ¿Cuál es la probabilidad de que llegue más de 3 personas en una hora?
- ¿Cuál es la probabilidad de que llegue como máximo 4 personas en una hora?
- ¿Cuál es la probabilidad de que lleguen entre 1 y 3 personas en una hora?

- 21.** Un concesionario de automóviles calcula la proporción de automóviles nuevos vendidos que se han devuelto varias veces para que se corrijan los defectos durante el periodo de garantía. Los resultados fueron los siguientes:

Número de devoluciones	0	1	2	3	4
Proporción	0,28	0,36	0,23	0,09	0,04

- Graficar la función de probabilidad.
- Calcular y graficar la función de probabilidad acumulada.
- Halle la esperanza del número de devoluciones de un automóvil para que se corrijan los defectos durante el periodo de garantía.
- Halle la varianza del número de devoluciones de un automóvil para que se corrijan los defectos durante el periodo de garantía.

- 22.** Una empresa está especializada en la instalación y el mantenimiento de calefacciones centrales. Antes de que empiece el invierno, las llamadas al servicio de mantenimiento pueden dar como resultado el pedido de una nueva caldera. La tabla adjunta muestra las probabilidades estimadas del número de pedidos de calderas nuevas generados de esta forma en las 2 últimas semanas de septiembre.

Número de pedidos	0	1	2	3	4	5
Probabilidad	0,10	0,14	0,26	0,28	0,15	0,07

- Graficar la función de probabilidad.
- Calcular y graficar la función de probabilidad acumulada.

- c) Halle la probabilidad de que se hagan al menos tres pedidos en este periodo.
- d) Halle la media del número de pedidos de una nueva caldera en este periodo de 2 semanas.
- e) Halle la desviación típica del número de pedidos de una nueva caldera en este periodo de 2 semanas.

- 23.** Un profesor tiene un numeroso grupo de alumnos y ha previsto un examen a las 7 de la tarde en un aula diferente. Estime en la tabla las probabilidades del número de estudiantes que lo llamarán a casa una hora antes del examen preguntándole en qué aula se realizará.

Número de llamadas	0	1	2	3	4	5
Probabilidad	0,10	0,15	0,19	0,26	0,19	0,11

Halle la media y la desviación típica del número de llamadas.

- 24.** Un equipo de reparación es responsable de un tramo de un oleoducto de 2 kilómetros de largo. La distancia (en kilómetros) a la que surge cualquier grieta puede representarse por medio de una variable aleatoria distribuida uniformemente, con una función de densidad de probabilidad:

$$f(x) = 0,5$$

Halle la función de distribución acumulada y la probabilidad de que surja cualquier grieta dada entre 0,5 kilómetros y 1,5 en este tramo del oleoducto.

- 25.** Dentro de la jurisdicción de un equipo de salvamento se encuentran las emergencias que se produzcan en un tramo de un río que tiene 4 kilómetros de largo. La experiencia ha demostrado que la distancia, expresada en kilómetros desde el punto situado más al norte, a la que se produce una emergencia dentro de este tramo puede representarse por medio de una variable aleatoria distribuida uniformemente en el rango 0 a 4 kilómetros. En ese caso, si X representa la distancia (en kilómetros) a la que se produce una emergencia desde el punto situado más al norte de este tramo del río, su función de densidad de probabilidad es:

$$f(x) = \begin{cases} 0,25 & \text{para } 0 < x < 4 \\ 0 & \text{para todos los demás valores de } x \end{cases}$$

- a) Grafique la función de densidad de probabilidad.

- b) Hallar y graficar la función de distribución acumulada.
- c) Hallar la probabilidad de que se produzca una emergencia como máximo a un kilómetro del punto situado más al norte de este tramo del río.
- d) La base del equipo de salvamento se encuentra en el punto medio de este tramo del río. Halle la probabilidad de que se produzca una emergencia a más de 1,5 kilómetros de esta base.