

TAREA 1

Análisis exploratorio manual de datos químicos

Introducción al análisis de datos para otras carreras

Escuela de Informática y Computación

Información de la tarea

Curso: EIY403 - Introducción al análisis de datos para otras carreras

Valor: 5 % de la nota final

Modalidad: Individual

Fecha de entrega: Clase 4 - Martes 20 de agosto, 6:00 PM

Formato: Manuscrito a puño y letra, entrega física al inicio de clase

1. Introducción al caso

Usted trabaja como asistente de laboratorio en una empresa química industrial que produce soluciones especializadas para tratamiento de aguas. Durante el mes de julio 2025, se realizaron mediciones de **pH** en 50 muestras de una nueva formulación de solución buffer para control de acidez en procesos industriales.

Contexto químico importante:

Las soluciones buffer son cruciales en la industria química para mantener el pH estable durante procesos químicos. Un pH fuera del rango óptimo (6.8 - 7.2) puede causar:

- Corrosión de equipos metálicos
- Precipitación de compuestos indeseados
- Reducción en la eficiencia del proceso
- Problemas de calidad del producto final

2. Dataset: mediciones de pH

Las siguientes 50 mediciones de pH fueron tomadas bajo condiciones controladas de laboratorio:

Muestra	pH	Muestra	pH	Muestra	pH	Muestra	pH	Muestra	pH
1	7.02	11	6.89	21	7.15	31	6.92	41	7.08
2	6.95	12	7.11	22	6.87	32	7.03	42	6.88
3	7.18	13	6.84	23	7.06	33	6.99	43	7.12
4	6.91	14	7.05	24	6.93	34	7.17	44	6.86
5	7.09	15	6.97	25	7.01	35	6.85	45	7.04
6	6.83	16	7.13	26	6.96	36	7.14	46	6.98
7	7.16	17	6.90	27	7.07	37	6.94	47	7.10
8	6.87	18	7.04	28	6.82	38	7.19	48	6.89

9	7.05	19	6.98	29	7.00	39	6.91	49	7.06
10	6.94	20	6.85	30	6.88	40	7.15	50	6.95

3. Ejercicios a resolver

3.1. Parte I: organización y clasificación de datos (15 puntos)

Ejercicio 1.1: Clasificación del tipo de datos (5 puntos)

Clasifique la variable pH según:

- a) Tipo de dato (cuantitativo/cualitativo, discreto/continuo)
- b) Escala de medición (nominal, ordinal, intervalo, razón)
- c) Justifique su respuesta explicando qué operaciones matemáticas son válidas

Ejercicio 1.2: Organización de datos (10 puntos)

- a) Ordene los 50 valores de pH de menor a mayor
- b) Identifique el valor mínimo y máximo
- c) Calcule el rango de los datos
- d) Cree una tabla de frecuencias agrupando los datos en 6 intervalos de clase

3.2. Parte II: medidas de tendencia central (25 puntos)

Ejercicio 2.1: Media aritmética (8 puntos)

- a) Calcule la media aritmética usando la fórmula: $\bar{x} = \frac{\sum x}{n}$
- b) Muestre todos los pasos del cálculo
- c) Interprete el resultado en el contexto químico

Ejercicio 2.2: Mediana (8 puntos)

- a) Calcule la mediana de los datos ordenados
- b) Explique el procedimiento para $n = 50$ (número par)
- c) Compare el resultado con la media. ¿Qué indica esta comparación sobre la distribución?

Ejercicio 2.3: Moda (9 puntos)

- a) Identifique el(los) valor(es) que aparece(n) con mayor frecuencia
- b) Clasifique la distribución (unimodal, bimodal, multimodal o amodal)
- c) Explique cuál medida de tendencia central es más apropiada para este dataset y por qué

3.3. Parte III: medidas de tendencia no central (20 puntos)**Ejercicio 3.1: Cuartiles (15 puntos)**

- a) Calcule Q1 (percentil 25)
- b) Calcule Q3 (percentil 75)
- c) Calcule el rango intercuartílico ($IQR = Q3 - Q1$)
- d) Interprete qué significa cada cuartil en términos de control de calidad

Ejercicio 3.2: Detección de outliers (5 puntos)

Use el método IQR para detectar outliers:

- a) Calcule los límites: Inferior = $Q1 - 1.5 \times IQR$, Superior = $Q3 + 1.5 \times IQR$
- b) Identifique si hay outliers en los datos
- c) Si encuentra outliers, discuta posibles causas químicas (errores de medición, contaminación, etc.)

3.4. Parte IV: medidas de dispersión (25 puntos)**Ejercicio 4.1: Varianza y desviación estándar (15 puntos)**

- a) Calcule la varianza muestral usando: $s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$
- b) Calcule la desviación estándar: $s = \sqrt{s^2}$
- c) Muestre al menos 5 cálculos individuales de $(x - \bar{x})^2$
- d) Interprete la desviación estándar: ¿es alta o baja la variabilidad?

Ejercicio 4.2: Coeficiente de variación (10 puntos)

- Calcule el coeficiente de variación: $CV = \frac{s}{\bar{x}} \times 100 \%$
- Clasifique la variabilidad (baja: $\leq 15 \%$, moderada: $15-30 \%$, alta: $\geq 30 \%$)
- Evalúe si esta variabilidad es aceptable para un proceso de control de calidad

3.5. Parte V: análisis integral y toma de decisiones (15 puntos)**Ejercicio 5.1: Control de calidad químico (15 puntos)**

Basándose en sus cálculos, responda:

- ¿Qué porcentaje de las muestras está dentro del rango óptimo (6.8 - 7.2)?
- Usando la regla empírica (68-95-99.7), ¿qué rango de pH esperaría para el 95 % de las muestras?
- ¿Recomendaría aprobar este lote de solución buffer? Justifique con evidencia estadística
- Proponga 3 acciones específicas para mejorar el control de calidad del proceso

4. Criterios de evaluación

Criterio	Puntos	Descripción
Precisión en cálculos	40	Fórmulas correctas, operaciones precisas, resultados exactos
Procedimientos mostrados	25	Pasos claramente detallados, trabajo organizado
Interpretaciones químicas	20	Contexto industrial aplicado, conexiones relevantes
Análisis crítico	10	Evaluación de resultados, recomendaciones fundamentadas
Presentación	5	Legibilidad, organización, formato profesional
Total	100	

5. Instrucciones importantes

Requisitos obligatorios:

- Trabajo individual - No se permite colaboración
- Mostrar todos los cálculos paso a paso
- Usar al menos 4 decimales en cálculos intermedios
- Redondear resultados finales a 3 decimales
- Incluir interpretaciones químicas en cada sección
- Escribir con letra legible o a máquina
- Entregar en formato físico antes de la Clase 3

Fórmulas de referencia:

$$\text{Media: } \bar{x} = \frac{\sum x}{n} \quad (1)$$

$$\text{Varianza muestral: } s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \quad (2)$$

$$\text{Desviación estándar: } s = \sqrt{s^2} \quad (3)$$

$$\text{Coeficiente de variación: } CV = \frac{s}{\bar{x}} \times 100 \% \quad (4)$$

$$\text{Posición cuartil: } Q_k = \frac{k(n + 1)}{4} \quad (5)$$

Esta tarea integra conceptos fundamentales del análisis exploratorio de datos con aplicaciones reales en química industrial, preparando a los estudiantes para el uso de herramientas computacionales en las siguientes clases.