



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

**Nombre:** Delfino Azael Ávila Rodríguez

**Matricula:** 1861240

**TIPO DE EVALUACIÓN:** EXAMEN DIAGNÓSTICO.

**MATERIA/UNIDAD DE APRENDIZAJE:** Minería de datos.

**I. Lee el caso de ejemplo y responde las preguntas correspondientes:**

La Coca Cola Company tiene 366,000 accionistas y efectúa una encuesta por medio de la selección aleatoria de 30 accionistas de cada uno de los 50 estados de Estados Unidos. Se registra el número de acciones de cada accionista de la muestra.

- a) ¿Los valores obtenidos son discretos o continuos?  
Discretos, por que son contables (son todos los números enteros positivos), ya que no medias acciones.
- b) Identifique el nivel de medición (nominal, ordinal, de intervalo, de razón) de los datos muestrales.  
-De razón
- c) ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza?  
-Muestreo aleatorio estratificado
- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?  
Estadístico, ya que no hemos calculado la media de toda la población.
- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?  
La media, ya que me interesaría saber cual es la media de acciones que tienen los accionistas de la empresa.
- f) ¿Qué es lo que está incorrecto al evaluar la opinión del accionista enviado un cuestionario por correo que los accionistas podrían llenar y regresar por el mismo medio?  
-Que no tendríamos la certeza de que el accionista sea el que verdaderamente contesto la encuesta.

**II. Describe con tus palabras para que se utiliza el análisis uni-variante de datos:**

Es el análisis de datos basado en una sola variable.

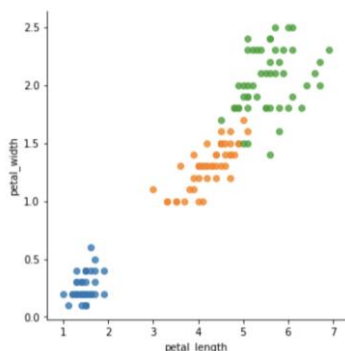
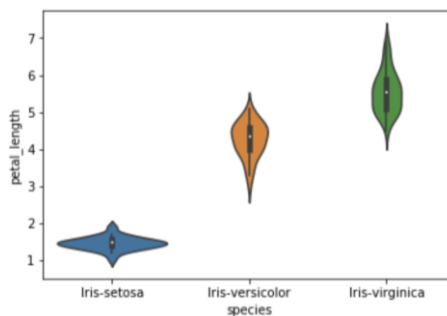
**III. Describe qué es la media, la mediana y la desviación estándar:**

La media y mediana son medidas de tendencia central, y la primera nos muestra el valor promedio de los datos, y la segunda nos muestra el dato que se encuentra justamente en la mitad de la muestra (tiene el 50% de los datos arriba y el 50% de los datos abajo). La desviación estándar es una medida de dispersión y nos muestra que tan dispersos están los datos en relación a la media.

**IV. Con tus palabras explica qué es minería de datos y en donde puede ser aplicada:**

Es el proceso de analizar una gran cantidad de datos para hallar anomalías o patrones, y puede ser aplicada en todos los campos que generen datos.

- V. **Que lenguajes de programación dominas o en que porcentaje lo manejas:**  
R – Avanzado  
Python – Principiante  
Visual Basic – Intermedio
- VI. **Sabes manejar notebooks digitales como Jupyter, Spyder, Google Collab, etc. (Si la respuesta es otro, anota cuál):**  
Si, solo Jupyter.
- VII. **¿Trabajas o has trabajado alguna vez en alguna empresa que maneje técnicas de ciencia de datos?**  
Si, me encuentro realizando practicas en una empresa que se dedica al análisis de datos.
- VIII. **Has utilizado alguna herramienta de aprendizaje de máquina. Si la respuesta es afirmativa, anota cuál:**  
No, nunca he utilizado una herramienta de aprendizaje de maquina.
- IX. **Relaciona los conceptos con sus gráficas correspondientes:**
- a) Se utilizan para evaluar visualmente cómo se distribuyen los puntos de datos con respecto a su frecuencia.
  - b) Forma estandarizada de mostrar la distribución de datos basada en un resumen de cinco números ("mínimo", primer cuartil (Q1), mediana, tercer cuartil (Q3) y "máximo"). Puede informarle si sus datos son simétricos, qué tan estrechamente están agrupados y si sus datos están sesgados.
  - c) Método para visualizar la distribución de datos numéricos de diferentes variables. Es similar al diagrama de caja, pero con un diagrama rotado en cada lado, que brinda más información sobre la estimación de densidad en el eje y.
  - d) Visualización de datos bidimensionales que utiliza puntos para representar los valores obtenidos para dos variables diferentes: una trazada a lo largo del eje "xy" otra trazada a lo largo del eje y.
  - e) Produce una matriz de relaciones entre cada variable en sus datos para un examen instantáneo de nuestros datos. También puede ser un gran punto de partida para determinar los tipos de análisis de regresión a utilizar.

**d****c**

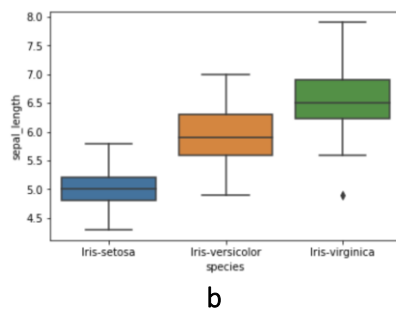
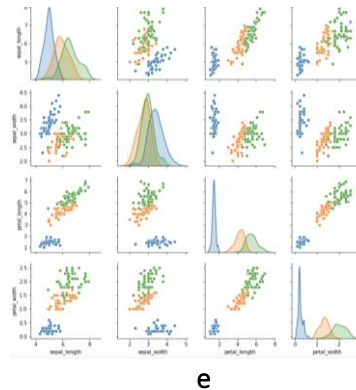
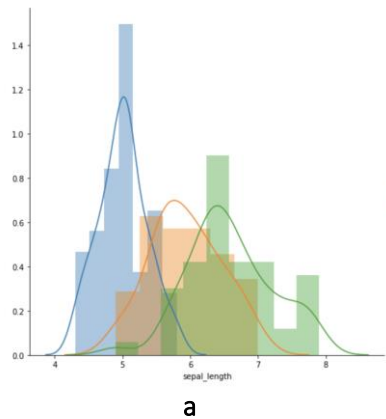


UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



- X. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos):  
**No, no tengo problemas para compartir pantalla o hablar en micrófono por Teams.**
- XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno)  
**Si, hasta el momento no me aparece el grupo de Minería de Datos en Nexus.**
- XII. Qué sistema operativo tienes:  
**Windows**
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en que porcentaje.
- Github.  
**No tengo conocimiento.**
  - Jupyter Notebook  
**Si tengo conocimiento, un 30%.**
  - Google Collab  
**No tengo conocimiento.**
  - Python 3



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

**No tengo conocimiento.**

- Terminal de tu computadora.

**No tengo conocimiento.**