



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

## TIPO DE EVALUACIÓN: EXAMEN DIAGNÓSTICO.

MATERIA/UNIDAD DE APRENDIZAJE: Minería de datos.

Nombre del alumno: Marlene Michel Sepulveda Bermudez      Matricula: 1862646

### 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
- b) Identifique el nivel de medición (nominal, ordinal, de intervalo, de razón) de los datos muestrales.  
El nivel de medición es de razón.
- c) ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza?  
El tipo de muestreo es estratificado
- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?  
Estadístico
- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?  
El numero promedio de las acciones de los accionistas.
- 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?  
La mayoría de las veces al evaluar un cuestionario a través de un correo es que a veces se pierde con otros correos no deseados o el accionista no lo contesta.

### II. Describe con tus palabras para que se utiliza el análisis uni-variante de datos: es el análisis de datos con una sola variante

Es el análisis de datos mas sencillo, ya que los datos analizados solo contienen una variable y su principal objetivo del análisis es describir los datos y encontrar los patrones que existen entre los mismos. Un ejemplo de una uni-variante es la edad.

### III. Describe qué es la media, la mediana y la desviación estándar: la media es el promedio de los datos, la mediana es el número que esta justo en medio de los datos, la desviación estándar es para cuantifica variaciones de datos.



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

La media es el promedio de una distribución de datos, la mediana es el dato central de una distribución de datos ordenados de manera que se pueden dividir en 2 partes iguales y la desviación estándar es la raíz cuadrada de la varianza y indica que tan dispersos están el conjunto de datos con respecto a la media.

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

La minería de datos es el proceso de hallar patrones, anomalías y correlaciones en grandes conjuntos de datos para predecir resultados que ayuden en el análisis de esos mismos datos.

**V. Que lenguajes de programación dominas o en que porcentaje lo manejas:** programación c un 80%, programación en r 50% y visual basic un 95%.

**VI. Sabes manejar notebooks digitales como Jupyter, Spyder, Google Collab, etc. (Si la respuesta es otro, anota cuál):** No

**VII. ¿Trabajas o has trabajado alguna vez en alguna empresa que maneje técnicas de ciencia de datos?**

Ahorita estoy en practicas en un banco donde he visto algunas técnicas donde se manejan los datos de los clientes.

**VIII. Has utilizado alguna herramienta de aprendizaje de máquina. Si la respuesta es afirmativa, anota cuál:** No

**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.

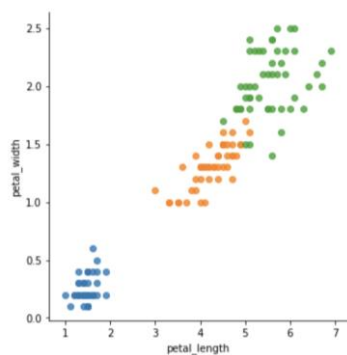


# UANL

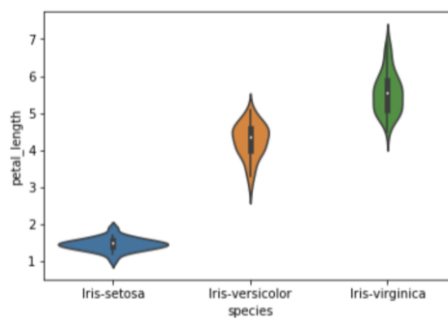
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM

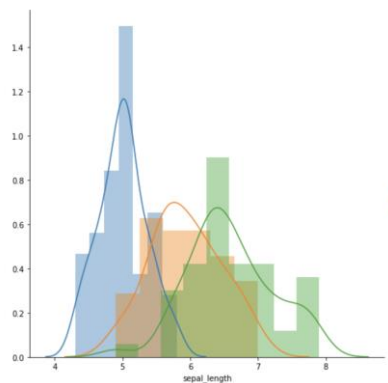
FACULTAD DE CIENCIAS FISICO MATEMÁTICAS



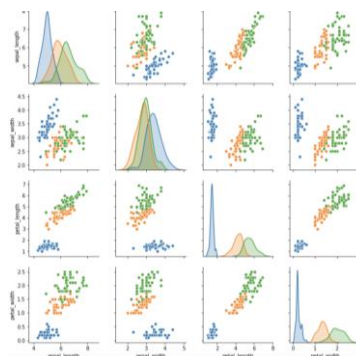
d) Grafico de dispersión



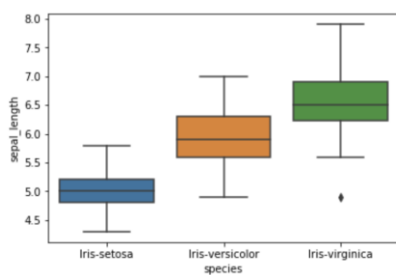
c) Diagrama de caja rotado



a) Histograma



e) Matrices de gráficos de dispersión e histograma



b) Diagrama de caja

- X. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos): En algunas ocasiones si prendo mi cámara o comparto pantalla mi laptop se vuelve muy lenta o se traba algunas veces y no me deja trabajar, pero mas bien es el internet.



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno)

Si la materia no me aparece

- XII. Que sistema operativo tienes: Sistema operativo de 64 bits, procesador basado en x64
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en que porcentaje.
- Github. 5%
  - Jupyter Notebook 5%
  - Google Collab 5%
  - Python 20%
  - Terminal de tu computadora. 15%