



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

## 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**
- 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?  
**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 número de acciones que cada accionista tiene.**
- 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?  
**Debido a que es una encuesta de opinión, va a haber discrepancia en los datos por el pensamiento individual que cada persona tiene.**

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

**El análisis uni-variante nos ayuda a analizar una sola variable, por ende, podemos ver los cambios independientes de dicha variable, sin tener a otras impactando en los resultados de esta.**

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

**Media: Es un valor que se obtiene al sacar el promedio de una muestra.**

**Mediana: Conjunto es un valor que se encuentra a la mitad de los otros valores, es decir, que al ordenar los números de menor a mayor, éste se encuentra justamente en medio entre los que están por arriba.**

**Desviación Estándar: Indica que tan dispersos están los datos con respecto a la media, mientras mayor sea, habrá mayor dispersión de los datos.**



# UANL

# FCFM



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

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

La minería de datos es cuando se agrupan grandes cantidades de información para poder ser trabajada, analizada y estudiada. Esto con el fin de lograr resultados como tendencias, predicciones, comportamientos en las personas entre otras cosas. Lo podemos ver día a día en redes sociales como Instagram o Facebook que lo utilizan para recomendarnos productos que nos gustan o necesitamos.

V. Que lenguajes de programación dominas o en que porcentaje lo manejas:

R – 30%

C – 50%

SQL – 10%

Visual Basic Excel – 70%

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?

No

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.

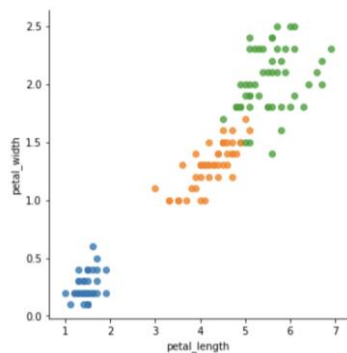


# UANL

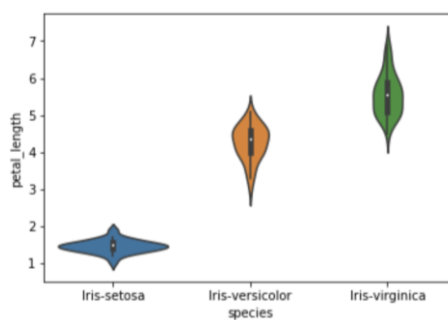
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM

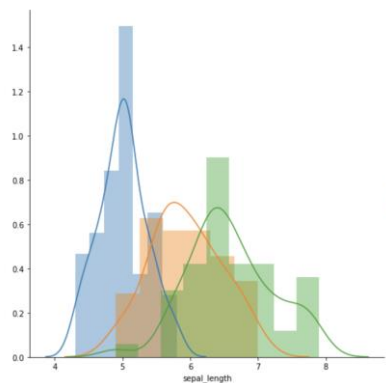
FACULTAD DE CIENCIAS FISICO MATEMÁTICAS



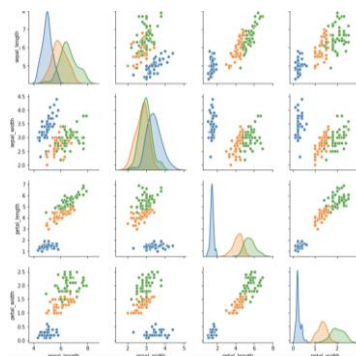
a)



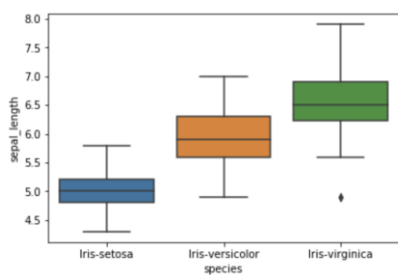
c)



d)



e)



b)

X. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos):

No hay ningún problema de este tipo

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



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

- XII. No hay ningún problema de este tipo
- XIII. Que sistema operativo tienes:  
Windows 10
- XIV. De las siguientes herramientas menciona si tienes conocimiento de alguna y en que porcentaje.
- Github. 0%
  - Jupyter Notebook 0%
  - Google Collab 0%
  - Python 5%
  - Terminal de tu computadora. 5%