



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 porque tiene un numero finito de valores.
- b) Identifique el nivel de medición (nominal, ordinal, de intervalo, de razón) de los datos muestrales.
  - Intervalo, ya que puede consistir desde 0 hasta 366,000.
- c) ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza?
  - Estratificado, donde cada estado es un estrato.
- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?
  - Parámetro.
- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?
  - Datos concretos, y que sean la cantidad suficiente para que el margen de error no sea mayor a 5%.
- f) ¿Qué es lo que está incorrecto al evaluar la opinión del accionista enviando un cuestionario por correo que los accionistas podrían llenar y regresar por el mismo medio?
  - Si solo consideraras la muestra de 30 accionistas en cada uno de los 50 estados, no puedes hacer un muestreo de opinión/satisfacción de cualquier situación.

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

- Cuando se quiere examinar los efectos de una sola variable en un conjunto.

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

- La media es, también llamado promedio, el valor al cual la mayoría de los datos tienden a él. La mediana es el numero medio del conjunto ordenado. La desviación estándar es cuanto varía los datos ya sea a la derecha o izquierda respecto a la media.

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

- Es el estudio de las muestras con gran cantidad de tamaño, es algo relacionado con la estadística, pero no tan relacionado a dar resultados, sino como analizarlos.



# UANL

# FCFM



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

- V. **Que lenguajes de programación dominas o en qué porcentaje lo manejas:**
- Visual Basic en un 70%
  - C++ en un 25%
  - R en un 25%
  - Python en un 30%-35%
- VI. **Sabes manejar notebooks digitales como Jupyter, Spyder, Google Collab, etc. (Si la respuesta es otro, anota cuál):**
- Solo he utilizado Jupyter, pero vi un poco el interfaz de Spyder y siento que se parece un poco a la de R.
- VII. **¿Trabajas o has trabajado alguna vez en alguna empresa que maneje técnicas de ciencia de datos?**
- Aún no consigo trabajo :c
- VIII. **Has utilizado alguna herramienta de aprendizaje de máquina. Si la respuesta es afirmativa, anota cuál:**
- No he utilizado



# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

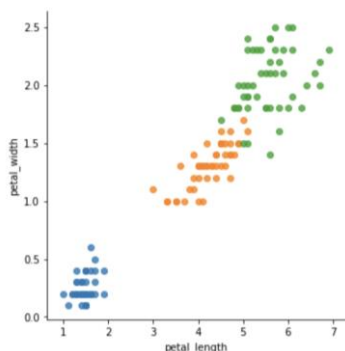
# FCFM

FACULTAD DE CIENCIAS FISICO MATEMÁTICAS

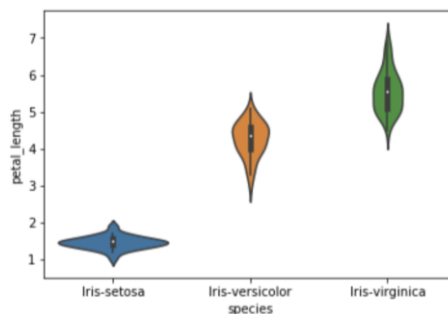


## IX. Relaciona los conceptos con sus gráficas correspondientes:

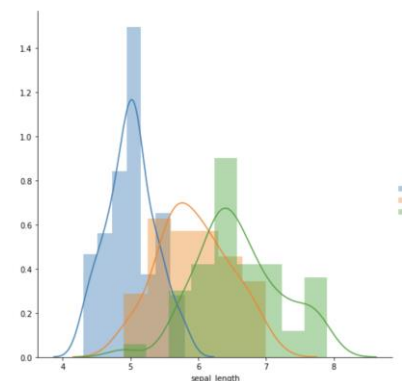
- a) Se utilizan para evaluar visualmente cómo se distribuyen los puntos de datos con respecto a su frecuencia.
- b) Forma estandarizada de mostrar la distribución de datos basada en un resumen de cinco números ("mínimo", primer cuartil (Q1), mediana, tercer cuartil (Q3) y "máximo"). Puede informarle 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.



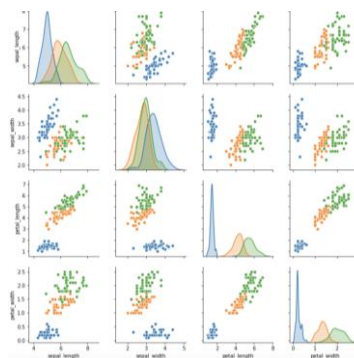
d)



c)



a)



e)

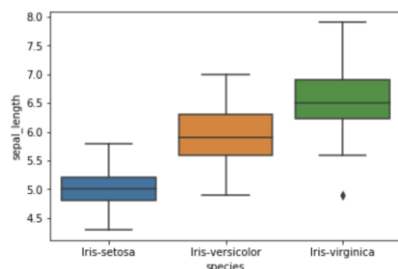


# UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

# FCFM

FACULTAD DE CIENCIAS FISICO MATEMÁTICAS



b)

- X. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos)
- Tengo problemas relacionado con la conexión de internet, es un poco malo.
- XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno)
- No tengo ningún problema
- XII. Que sistema operativo tienes:
- Windows
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en que porcentaje.
- Github. – no tengo ningún conocimiento (0%)
  - Jupyter Notebook – tengo un conocimiento fresco, lo utilice hace semanas para un curso de Python. (35%)
  - Google Collab – no tengo ningún conocimiento (0%)
  - Python 3 – conozco lo básico, ciclos, listas y lo relacionado don data frame (25%)
  - Terminal de tu computadora. – solo lo conozco para poder abrir jupyter notebook (5%)