



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.

Razón

- c) ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza?

Muestreo estratificado

- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?

Estadístico, porque estamos calculando la media de la muestra

- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?

La varianza

- 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 hay una interacción directa entre ambas partes, por lo que la probabilidad de recibir respuesta disminuye. El correo puede perderse en spam, ser visto muy tarde o nunca ser visto.

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

Se utiliza para estudiar la variable más importante o relevante para el analista o para el objetivo del estudio, sin tomar en cuenta otras variables que pudiesen influir, puesto que no se busca entender o encontrar relaciones entre ellas. Sólo se tiene interés por el comportamiento o características de una variable por sí misma.

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

La media es el promedio de los datos; la mediana es el valor situado a la mitad de un conjunto de datos al ser ordenados de menor a mayor; la desviación estándar es una medida que permite saber qué tan dispersos están los datos con respecto a la media.

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

Es el proceso por el cual se extrae información útil de un conjunto de datos para así poder predecir resultados o escenarios futuros. Puede ser aplicada en muchas áreas, por ejemplo en el sector financiero para predecir el precio de un activo en base a precios históricos.



UANL

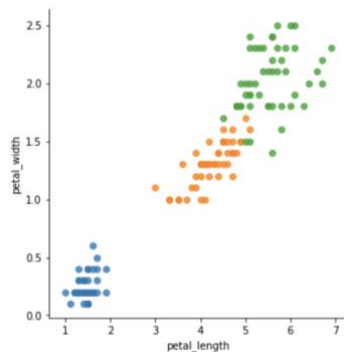
FCFM



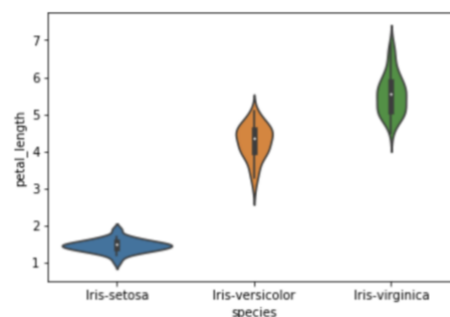
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

- V. **Que lenguajes de programación dominas o en que porcentaje lo manejas:** Estoy un poco familiarizada con los siguientes lenguajes: C (30%), html (35%), css (30%), visual basic (60%)
- VI. **Sabes manejar notebooks digitales como Jupyter, Spyder, Google Collab, etc. (Si la respuesta es otro, anota cuál):** He trabajado con Google Collab un poco.
- 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.



d



c

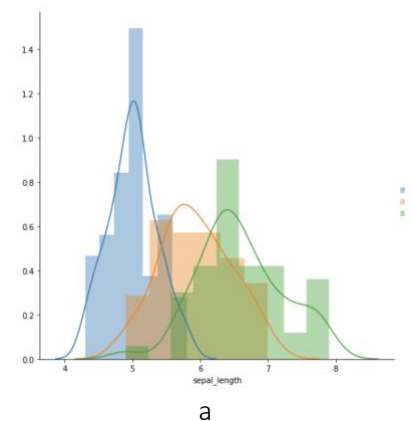


UANL

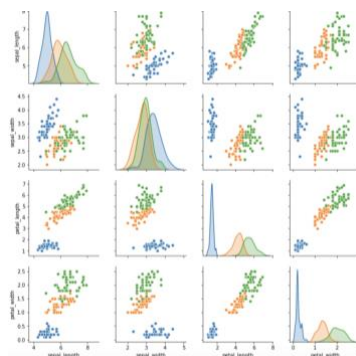
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

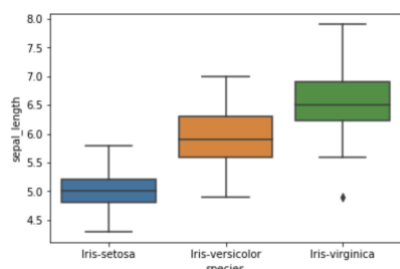
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



a



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
- XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno). No, pero actualmente no aparece la materia
- XII. Que sistema operativo tienes: macOS High Sierra
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en que porcentaje.
 - Github. 10%
 - Jupyter Notebook 0%
 - Google Collab 15%
 - Python 3 5%
 - Terminal de tu computadora. 15%

En general, tengo muy poco conocimiento en dichas herramientas, pero no se me dificulta aprender rápido.