



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?
Los valores obtenidos son discretos.
- b) Identifique el nivel de medición (nominal, ordinal, de intervalo, de razón) de los datos muestrales.
El nivel de medición de los datos muestrales es de razón dado lo que representaría el cero en estos casos.
- c) ¿Qué tipo de muestreo (aleatorio, sistemático, de conveniencia, estratificado, por racimos) se utiliza?
El tipo de muestreo que se utiliza es el muestreo estratificado debido a cómo se formula el problema.
- d) Si se calcula el número promedio (la media) de acciones. ¿El resultado es un estadístico o un parámetro?
La media de acciones en este caso es un estadístico ya que el valor de dicha media proviene de una muestra y no de la población completa.
- e) Si usted fuera ejecutivo en jefe de la Coca Cola Company, ¿qué característica del conjunto de datos consideraría que es extremadamente importante?
Lo primero que revisaría es que los datos fueran verdaderos, ya que no se puede hacer un estudio con datos falsos o que no tengan cierto nivel de confianza. Después revisaría la fecha en que se recopilaron los datos, para ver si me funcionarían o no, de acuerdo con las necesidades de mi estudio. Finalmente, revisaría si realmente son significativos y me ayudan a predecir comportamientos en la población o no.
- 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?
No se sabría con franqueza si el accionista contestó el cuestionario el mismo o alguien más lo hizo. Podría haber considerablemente perdida de información o falta de respuesta por parte de los accionistas al no ser un trato directo. Se puede perder el correo o simplemente tener la dirección electrónica incorrecta, lo que causaría problemas porque no se dirigiría necesariamente a un accionista.



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

- II. **Describe con tus palabras para que se utiliza el análisis uni-variante de datos:**
El análisis uni-variante de datos se utiliza cuando solo se tiene una variable que revisar. Usualmente se hace con el fin de estudiar a fondo los estadísticos principales de la variable (media, mediana, moda, desviación, etc.). Este análisis no se preocupa del por qué de los datos como cuando se mezclan más variables y se tratan de relacionar. Con este análisis se busca encontrar patrones o distribuciones a las cuales se les asigne esta variable para predecir a futuro comportamientos.
- III. **Describe qué es la media, la mediana y la desviación estándar:**
La media se podría traducir de alguna manera como el valor promedio de los datos; Es un valor que nos ayuda a determinar en dónde se encuentra teóricamente la mayor concentración de datos refiriéndonos a sus valores. La mediana es el dato que ocupa el lugar central una vez que los datos están ordenados, nos ayuda a visualizar de mejor manera los cuartiles y a conocer el centro de los datos, que puede ser utilizado para referirnos a que el 50% de los datos son mayores o menores a cierto valor, entre otras cosas. Finalmente, la desviación estándar es un valor que nos dice cuánto se desvían los valores de la media o promedio mencionado anteriormente, nos puede indicar la concentración de datos cerca o lejos de la media, a mayor desviación podremos afirmar que los datos son más dispersos y a menor desviación diremos que los datos son más concentrados.
- IV. **Con tus palabras explica qué es minería de datos y en donde puede ser aplicada:**
La minería de datos es la extracción de información valiosa proveniente de bases de datos que en ocasiones son muy grandes. La minería de datos se encarga de recopilar la información, hacer resúmenes, encontrar patrones, métricas, indicadores y encontrar toda información que pueda utilizarse. Puede ser aplicada en empresas que se interesan en conocer a sus empleados y tienen bases de datos con esta información; puede aplicarse para conocer información histórica de gastos, ventas, ingresos de una compañía y ayudarlos a saber en qué mejorar; entre otras cosas.
- V. **Qué lenguajes de programación dominas o en qué porcentaje lo manejas:**
Lenguaje de programación C (70%), Visual Basic (75%), Rstudio (40%) y Python(20%).
- VI. **Sabes manejar notebooks digitales como Jupyter, Spyder, Google Collab, etc. (Si la respuesta es otro, anota cuál):**
Únicamente he trabajado con Jupyter, me sentí cómodo cuando lo utilicé.
- VII. **¿Trabajas o has trabajado alguna vez en alguna empresa que maneje técnicas de ciencia de datos?**
En la empresa en la que estoy trabajando actualmente se utiliza la ciencia de datos para poder resumir cantidades de información que son extremadamente grandes, también para visualizar los históricos de compra y venta, así como de pagos y cobros que se han realizado. Se cuenta con una aplicación que es capaz de realizar todo esto para entregar reportes eficientes.
- VIII. **Has utilizado alguna herramienta de aprendizaje de máquina. Si la respuesta es afirmativa, anota cuál:**
No he utilizado herramientas de machine learning, o al menos conscientemente no, tal vez de manera intuitiva sin saber qué eran.



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

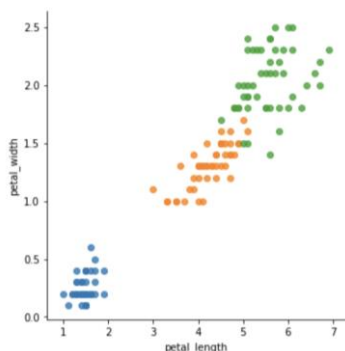
FCFM

FACULTAD DE CIENCIAS FISICO MATEMÁTICAS

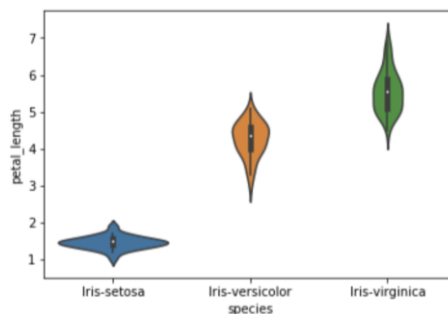


IX. Relaciona los conceptos con sus gráficas correspondientes:

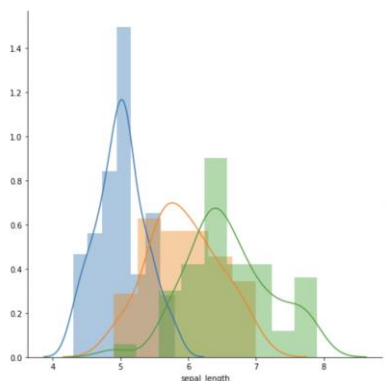
- a) Se utilizan para evaluar visualmente cómo se distribuyen los puntos de datos con respecto a su frecuencia.
- b) Forma estandarizada de mostrar la distribución de datos basada en un resumen de cinco números ("mínimo", primer cuartil (Q1), mediana, tercer cuartil (Q3) y "máximo"). Puede informarle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



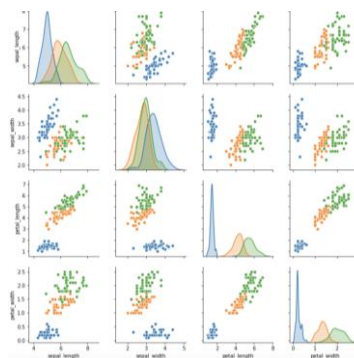
D



C



A



E

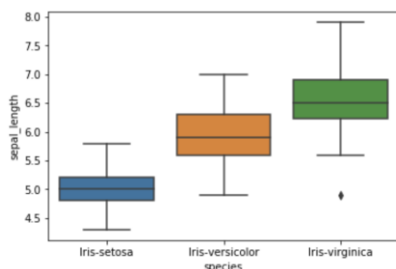


UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



B

- X. Tienes problemas de compartir pantalla, o hablar en micrófono en el Teams: (Menciona cuales son los problemas en caso de tenerlos)
No tengo ningún problema para realizar estas actividades.
- XI. Tienes problemas de abrir o utilizar Nexus: (Menciona cuales si tienes alguno)
No tengo problema al utilizar Nexus. De momento no ha cargado la asignatura todavía pero espero se solucione pronto.
- XII. Qué sistema operativo tienes:
Windows 10
- XIII. De las siguientes herramientas menciona si tienes conocimiento de alguna y en qué porcentaje.
- Github.
 - Jupyter Notebook
 - Google Collab
 - Python 3
 - Terminal de tu computadora.

No he trabajado con Github ni con Google Collab anteriormente. Jupyter notebook lo trabajé durante un curso, por lo que sé lo básico solamente, así como también de Python.