



¡Felicidades!

Aprobaste el curso. Ya puedes acceder a tu diploma digital.

9.6

Calificación

24 / 25

Aciertos

1. ¿Qué es la **imputación de valores faltantes**?

Estimar los valores ausentes con base en los valores válidos de otras variables y/o casos de muestra o modelos.



2. Como profesional en data, podrás crear múltiples modelos para entender el comportamiento de tus datos. ¿Qué problemas puedes enfrentar al crear modelos con datos con valores faltantes?

Todas las opciones son correctas.



3. Al realizar cualquier tipo de imputación de datos deberás realizar una **exploración y limpieza de datos**. ¿Cuál de las siguientes acciones **NO** es un paso válido en este proceso?

Eliminar todos los valores faltantes en el conjunto de datos sin previamente revisarlos.



4. ¿Es posible eliminar los valores faltantes sin una revisión previa y que esto **NO** introduzca sesgos?

No, es necesario revisar los valores faltantes para saber si es posible eliminarlos.



5. ¿Cuál de las siguientes afirmaciones **NO** es una implicación de los datos **Missing Completely At Random (MCAR)**?

La imputación es necesaria.



6. ¿Cuál de las siguientes afirmaciones **SÍ** es una implicación de los datos **Missing At Random (MAR)**?

Todas las opciones son correctas.



7. ¿Cuál de las siguientes afirmaciones **SÍ** es una implicación de los datos **Missing Not At Random (MNAR)**?

Todas las opciones son correctas.



8. Después de examinar tus datos con valores faltantes, tus colegas te sugieren mejorar tu diseño experimental o realizar un análisis de sensibilidad. ¿Qué tipo de mecanismo de valores faltantes creen que actúa sobre tus datos?

MNAR



9. Una vez ajustamos y transformamos nuestros datos con codificador, podemos hacer uso de su atributo `_____` para obtener las categorías únicas de nuestro conjunto de datos.

`encoder.categories_`



10. ¿Cuál es el papel de la función `sklearn.compose.make_column_transformer()` ?

Transformar cada columna con un transformador asociado a esta y guardar la información necesaria para realizar el procesamiento inverso.



11. ¿Qué es la **codificación one-hot encoding**?

Convertirá cada valor categórico en una nueva columna categórica y le asignará un valor binario de 1 o 0 a esas columnas.



12. ¿Cuál es una ventaja de realizar one-hot encoding utilizando sklearn en lugar de Pandas?

Realizar one-hot encoding puede llegar a ser más robusto debido a que guarda la información necesaria de las categorías involucradas e incluye una forma de realizar transformaciones inversas.



13. ¿Cuándo utilizarías una **codificación one-hot** en lugar de una codificación ordinal?

Cuando mis variables categóricas no tienen un orden natural.



14. ¿Por cuál de las siguientes razones podrías utilizar la **imputación por media, mediana y moda**?

Todas las opciones son correctas.



15. ¿Cuál de las siguientes afirmaciones **NO** es una desventaja de la **imputación por llenado hacia atrás y hacia adelante**?

Existen trucos para evitar romper las relaciones entre variables.

Repasar

16. Al realizar una **imputación por llenado hacia atrás y hacia adelante**, existen trucos que pueden ayudarte a enfrentar el problema de que las **relaciones multivariantes** puedan ser **distorsionadas**. ¿Cuál de las siguientes opciones **NO** es uno de estos trucos?

Realizar una imputación con dominios creados completamente al azar.



17. ¿Cuál de las siguientes afirmaciones es una desventaja de la **imputación por interpolación**?

Puede introducir valores fuera de rango.



18. Recibes datos de series temporales, ¿cuál método de Pandas podría ayudarte a probar distintos modelos de interpolación?

`df.interpolate()`



19. El algoritmo de **imputación por K-vecinos más cercanos (KNN)** es un método de imputación con base en modelos. Esto es:

Falso, es basada en árboles.



20. Necesitas cuantificar la **distancia únicamente entre variables numéricas**. ¿Cuál de las siguientes métricas de distancia podría ser más útil?

Euclidiana



21. Necesitas cuantificar la **distancia únicamente entre variables categóricas ordinales (tipo factor)**. ¿Cuál de las siguientes métricas de distancia podría ser más útil?

Manhattan



22. Necesitas cuantificar la **distancia entre variables de distintos tipos (numéricas, categóricas)**. ¿Cuál de las siguientes métricas de distancia podría ser más útil?

Gower



23. Dependiendo del software utilizado, algunos algoritmos pueden o no ordenar las variables de acuerdo a su valor de faltantes. ¿Cómo podría afectar esto a los resultados de la imputación?

La imputación podría estar sesgándose debido a utilizar estimados provenientes de variables con alta cantidad de valores faltantes en lugar de variables con baja o nula cantidad de valores faltantes.



24. ¿Cuál de las siguientes afirmaciones es una desventaja de la **imputación con base en modelos**?

Todas las opciones son correctas.



25. ¿Cuál de las siguientes afirmaciones es una desventaja de **Imputaciones Múltiples por Ecuaciones Encadenadas (MICE)**?

Para funcionar bien necesitas pensar en el modelo de imputación y el modelo de análisis.



[Ver menos](#)

[Ir a Inicio](#)