

Tarea #2

Regresión Lineal Multivariada

En la presente tarea seguiremos trabajando con los datos utilizados para la primera Tarea. Esta vez, analizaremos el salario de las personas. El objetivo es identificar aquellas variables que explican la diferencia de salario entre los empleados y cuantificar su efecto. Recuerde que para realizar este análisis contamos con datos de 1470 trabajadores de una misma empresa en el dataset *employee_attrition.csv*. La variable dependiente que analizaremos es *MonthlyIncome*.

PARTE 1: EXPLORACIÓN DE DATOS Y VARIABLES

1. Investigue acerca de cuáles variables influyen en explicar el salario de una persona [1.5 ptos.]
2. Analice las variables con las que cuenta según lo investigado en la parte anterior. Evalúe la necesidad de normalizar/estandarizar los datos o transformar las variables (por ejemplo a dummy o aplicar logaritmo). Justifique. [1.5 pto.]
3. Realice un gráfico de dispersión para algunas variables que considere relevantes en este problema. Realice graficos para ver la relacion entre las variables independientes con la dependiente [1.5 ptos.]
4. Analice la relevancia que tiene cada variable con la variable dependiente y entre ellas. [1.5 ptos.]

PARTE 2: MODELAMIENTO

Plantee 3 modelos de regresión lineal para la variable dependiente de interés. Recuerde que puede utilizar transformaciones de las variables (cuadráticas, logarítmicas) o agregar interacciones entre ellas. Además puede eliminar observaciones que ensucien su análisis. La elección de las variables utilizadas y su forma de ser incorporada al modelo debe sustentarse en los análisis realizados en la Parte 1 de esta Tarea.

Luego, analice los resultados obtenidos: compare R^2 , test de hipótesis correspondientes y concluya con respecto a los efectos (significativos o no) estimados. Luego, responda las siguientes dos preguntas:

1. ¿A qué se debe los resultados que obtiene? [1.0 ptos.]
2. ¿Identifica algún problema de endogeneidad en su análisis? ¿Cómo podría solucionarse? [0.5 ptos.]

La distribución de puntaje en esta parte es de 1.5 puntos por cada modelo planteado y su análisis (0.5 ptos. por el modelo y su justificación, 1.0 pto por el análisis de resultados).

Reglas de la Tarea:

- **Fecha de entrega:** Viernes 09 de julio de 2021 hasta las 23:59. No se aceptan atrasos.
- **Entregables:** Debe entregar un reporte con sus resultados y los códigos utilizados. El código debe correr sin errores. Además deberá entregar completado el archivo *Evaluacion_Tarea2.xlsx* agregándole su nombre al nombre del archivo. Ejemplo: *Evaluacion_Tarea2_ConstanzaContreras.xlsx*.
- **Lenguajes permitidos:** R y Python.
- **Formato de entrega:** Puede entregar un archivo PDF con el reporte y un archivo .R o .py con los códigos, o puede entregar un archivo .ipynb, .rmd (con el .html respectivo) que contenga reporte y códigos.
- Esta Tarea puede realizarla de manera individual o en parejas.
- **Auxiliares Encargados:** Rubén Ortega, Camila Pulgar y Florencia Signorini.
- Favor realizar las preguntas sobre la Tarea en el foro de U-cursos para que todo el curso posea la misma información.
- **Nota Tarea:** La nota de la presente Tarea será ponderada de acuerdo a la pauta de evaluación que entregue usted y su pareja. Se evaluarán 3 dimensiones en cada pregunta: diseño de la tarea, análisis de resultados/conclusiones y código. Para más detalles revisar el archivo *Evaluacion_Tarea2.xlsx*.