

## Trabajo Práctico Grupal - BAIN-091

Prof. Magaly Moraga - Prof. Luis Sánchez.- Prof. Luis Ojeda - Prof. Gerardo Soto

### INSTRUCCIONES GENERALES

- El trabajo es grupal y debe realizarse en equipos de dos integrantes.
- Las respuestas deben entregarse en un informe escrito elaborado con algún editor de texto digital, indicando claramente los nombres de ambos integrantes del grupo.
- El informe debe subirse a la plataforma siveducMD (sección Tareas) en formato pdf. **Plazo máximo de entrega viernes 29 de noviembre de 2024 antes de las 23:59 hrs.**
- No se evaluarán informes entregados fuera de plazo ni aquellos que contengan copias de diferentes fuentes, incluidas herramientas de modelos de lenguaje, como ChatGPT, Bing Chat u otro.

### INSTRUCCIONES DEL TRABAJO

La base de datos para el análisis será enviada a sus correos institucionales. Deberán identificar en ella la variable respuesta (dependiente) y, a partir de las variables independientes, realizar todas las posibles combinaciones de modelos de regresión lineal simple usando RStudio. La base de datos contiene observaciones colectadas en un entorno experimental de las siguientes variables (en paréntesis el nombre de la columna): Dureza del acero [HR] (dure), contenido de carbono [% en peso] (carb), contenido de cromo [% en peso] (crom), contenido de manganeso [% en peso] (mang), contenido de tungsteno [% en peso] (tung), contenido de molibdeno [% en peso] (moli), tamaño de grano [ $\mu\text{m}$ ] (gran), coeficiente de ferrita [%] (ferr), coeficiente de perlita [%] (perl), coeficiente de bainita [%] (bain), tasa de temple [ $^{\circ}\text{C/s}$ ] (temp), tasa de recocido [ $^{\circ}\text{C/s}$ ] (reco), y velocidad de enfriamiento [ $^{\circ}\text{C/s}$ ] (enfr). A usted se le entrega un subconjunto de estas variables, con el objetivo de determinar el mejor modelo predictivo para el fenómeno observado, por lo que es necesario responder las preguntas que se plantean en este documento en el contexto de las variables acá propuestas. A continuación se le solicita responder lo siguiente:

- Elaborar un cuadro resumen que contenga un gráfico de dispersión para cada combinación de variables y determinar su correspondiente coeficiente de correlación lineal de Pearson. Luego, responder para cada combinación: ¿tiene sentido ajustar un modelo de regresión lineal simple?, ¿es necesaria alguna transformación en las variables? (20 Puntos)
- Para los modelos que tienen sentido, en un cuadro resumen, escribir su ecuación de regresión ajustada y explicar la relación en palabras. Graficar dicha recta en el diagrama de dispersión. (20 Puntos)
- Entre los modelos propuestos en b), concluir cuál de ellos es el mejor analizando de manera conjunta: (60 Puntos)
  - el cumplimiento de los supuestos (Usando técnicas gráficas y analíticas)
  - la significancia estadística de los parametros individuales y del modelo global. (Use Pruebas de Hipótesis)
  - el coeficiente de determinación  $R^2$ , RSE ( $\hat{S}_R$ ). Interpretar sus resultados.

Argumentar su decisión.

- Para el modelo seleccionado: (60 Puntos)
  - Interpretar los coeficientes de regresión estimados en el contexto del problema.
  - Obtener un intervalo de confianza del 95 % para el coeficiente  $\beta_1$ . Interpretar el resultado.
  - Ilustrar la utilidad del modelo realizando una predicción para un valor particular de  $X$  que no aparezca en la muestra. Interpretar en palabras tal valor predicho.
  - Obtener un intervalo de predicción para la respuesta media de  $Y$  cuando  $X$  toma un valor particular. Interpretar el intervalo real obtenido.

Nota: incluya un anexo con el script utilizado en R para la solución de cada ítem.