

SEGUNDO LABORATORIO EVALUADO

Un portal inmobiliario está interesado en generar nuevo contenido para los usuarios de su plataforma. Por ello está interesado en identificar variables que puedan explicar el precio de arriendo de las propiedades.

Las variables que disponibles para este análisis son las siguientes:

Variable	descripción
m2	metros cuadrados construidos
Precio	precio del arriendo en dólares
distancia	distancia al centro urbano, en metros
categoría	Zona de la ciudad (A o B)

A usted y a su equipo se les solicita que proponga un modelo que permita explicar el precio de arriendo en función de una de las variables disponibles. El objetivo es proporcionar a los usuarios del portal inmobiliario la información necesaria para que puedan tomar decisiones bien fundamentadas acerca de las propiedades disponibles.

Para crear este modelo considere lo siguiente:

1. Realice un análisis exploratorio de datos que permita visualizar las variables y ver si estas están relacionadas. Incluya gráficos y medidas estadísticas. Interprete.
2. Seleccione la variable que usted considera que explica de mejor forma el precio de arriendo y proponga un modelo que represente dicha relación. Identifique quien es la variable explicativa y la variable a explicar (variable dependiente e independiente)
3. Usando Rstudio, ajuste un modelo de regresión lineal. Escriba el modelo ajustado e interprete los coeficientes en el contexto del problema.
4. Grafique el modelo ajustado en conjunto con los datos.
5. Investigue y describa brevemente que representa el coeficiente de determinación lineal. Reporte este valor e interprételo en el contexto del problema.
6. Investigue y describa a que corresponden los gráficos cuantil-cuantil (qqplot) y realice dichos gráficos para las variables de su modelo. ¿Sus variables siguen una distribución normal? Justifique con el gráfico propuesto.
7. Investigue y describa brevemente a que corresponden los test de bondad de ajuste, en particular los test de normalidad. Plantee las hipótesis respectivas y determine si sus variables siguen una distribución normal. ¿Qué confianza utilizó para su prueba?

Para la sección de noticias del portal, se requiere saber si el precio promedio de arriendo ha aumentado y si este es diferente según zona de la ciudad, para este análisis considere:

8. Si los datos que usted maneja corresponden a una muestra aleatoria de la variable precio y si el precio promedio de arriendo del año anterior corresponde a \$970. ¿Puede afirmar dicha creencia? Use una confianza del 95%. Use Rstudio para realizar los cálculos necesarios.
9. Se está interesado en saber si el promedio del precio de arriendo es distinto según zona de la ciudad. Usando R realice una prueba de hipótesis que permita responder a esa pregunta. Use una significancia del 5%.

INSTRUCCIONES

- El laboratorio se desarrolla en grupos de 3 estudiantes.
- En caso de realizar test de hipótesis, recuerde plantear las hipótesis, el estadístico de prueba y valor-p. Concluya con dicha información en el contexto del problema.
- Debe entregar un informe en formato PDF y un script con el código respectivamente comentado.
- El nombre de archivo (informe y script) debe contener los apellidos de los integrantes del grupo separados por guión. Ejemplo:
Apellido1_Apellido2_Apellido3.pdf
- Fecha de entrega martes 24 de octubre a las 23:59 hrs vía buzón de tareas. No se aceptarán entregas por otra vía.

ESTRUCTURA DEL INFORME

El informe de laboratorio debe contar con la siguiente estructura:

- Introducción: en esta parte se debe explicitar las variables registradas y la importancia de su análisis. También puede incluir resultados generales, por ejemplo, el número total de clientes, promedios, entre otras medidas.
- Desarrollo: Responda a las preguntas a través de un hilo conductor coherente. Incluya tablas de frecuencias y gráficos con sus respectivos comentarios e interpretaciones.
- Conclusión: redacte una breve conclusión a nivel general.