

Marzo 2024

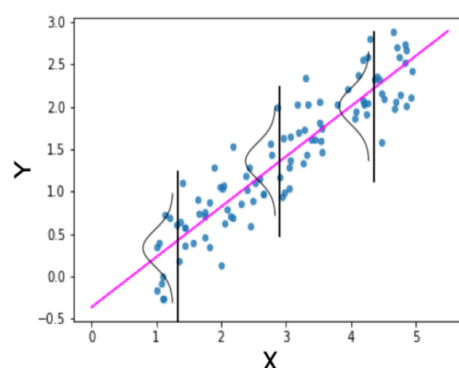
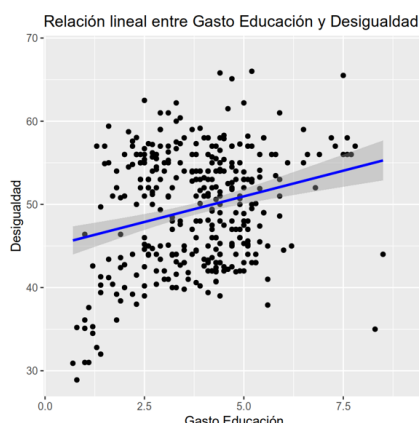
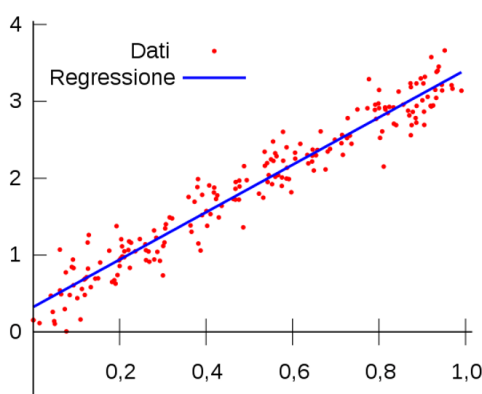
Trabajo final: Modelos lineales aplicados a la investigación



Introducción

Este curso fue diseñado para que los participantes exploraran un conjunto de técnicas estadísticas fundamentales que son ampliamente utilizadas en diversos campos de la

ciencia, la ingeniería, la medicina, las ciencias sociales y más. Los modelos lineales son una herramienta poderosa para comprender y analizar relaciones entre variables, así como para realizar predicciones y tomar decisiones basadas en datos.



Objetivos

Objetivo General:

El curso tiene como finalidad modelar en forma matemática el comportamiento de una variable de respuesta en función de una o más variables independientes (factores), utilizando programas estadísticos. Para esto se han planteado los siguientes objetivos específicos:

- ☐ Comprender los conceptos fundamentales de la regresión lineal.
- ☐ Aprender a aplicar modelos de regresión lineal en contextos profesionales e investigativos.
- ☐ Desarrollar habilidades para interpretar y comunicar los resultados de los modelos de regresión.
- ☐ Aplicar el conocimiento adquirido a través de un proyecto final.

Trabajo final

El trabajo final consiste en el desarrollo de un modelo lineal utilizando una base de datos proporcionada o por una base que se disponga. Los participantes deben seguir un proceso de análisis exploratorio de datos, selección de variables, ajuste de varios modelos lineales y validación de supuestos del último modelo seleccionado. Se espera que por lo menos se hayan probado 2 modelos lineales diferentes. Además, se espera que los participantes describan y justifiquen cada etapa del proceso para llegar modelo final y presenten conclusiones e inferencias basadas en el modelo final. El documento del trabajo final, debe adjuntar una portada donde se indique los nombres y apellidos de las personas que realizaron el trabajo.

Criterios de evaluación (El trabajo se evaluará sobre 70 puntos)

Actividad	¿qué se espera de la actividad?	Puntaje
Análisis Exploratorio de Datos	Se debe incluir un gráfico tipo boxplot de todas las variables de la base. Este gráfico indicará la dispersión que tienen los datos en las variables de la base.	5 puntos
Selección de Variables	<p>Para seleccionar las variables, se debe contar con un análisis de correlación, usar una gráfica para esto.</p> <p>Las variables que se elijan debe estar pensado en el modelo que se desea plantear (ver ejemplos de los modelos que se pueden plantear en descripción de la base)</p>	10 puntos
Desarrollo de Modelos	Se debe partir de un modelo, con las variables adecuadas. Pensar en la relación de la variable dependiente y las predictoras que se desea probar. Realizar una descripción detallada de los diferentes modelos lineales ajustados durante el proceso, incluyendo modelos simples y complejos. Para la selección del modelo usar métodos apropiados de ajuste de modelos, como regresión lineal múltiple, regresión polinómica, regresión ridge, lasso, etc. Para seleccionar el modelo se puede basar en criterios globales (AIC, BIC) o en coeficientes de estadísticas de ajuste (pvalor), puede utilizar métodos como backward, forward o both. Se espera que se tengan al menos 2 modelos.	35 puntos
Validación de Supuestos	<p>Verificación adecuada de los supuestos de la regresión lineal, como linealidad, normalidad de residuos, homocedasticidad, independencia de los residuos, etc.</p> <p>Solo se lo realizará en el modelo final del paso anterior. En caso de que no se cumpla algún supuesto, no es necesaria la transformación de datos.</p>	10 puntos
Conclusiones e Inferencias	Presentación clara de las conclusiones obtenidas a partir del modelo final. Interpretación adecuada de los coeficientes del modelo y sus implicaciones prácticas.	10 puntos

Base de datos

Descripción

Variable	Descripción
X1	Edad del individuo (en años)
X2	Género del individuo (0 = masculino, 1 = femenino)
X3	Nivel de educación (1 = bajo, 2 = medio, 3 = alto)
X4	Ingreso mensual del individuo (en dólares)
X5	Estado civil del individuo (1 = casado, 0 = soltero)
X6	Número de hijos
X7	Horas trabajadas por semana
X8	Índice de masa corporal (IMC)
X9	Consumo de alcohol (1 = sí, 0 = no)
X10	Fumador (1 = sí, 0 = no)

La base de datos propuesta consta de 300 registros de personas que se han levantado diferentes tipos de información. En esta base de datos se pueden proponer los siguientes modelos de ejemplo:

Ejercicios propuestos para diferentes modelos

Modelo de Regresión Simple:

- **Descripción:** Se desea predecir el ingreso mensual del individuo en función de su edad.
- **Ejercicio:** Ajusta un modelo de regresión simple donde la variable dependiente es el ingreso mensual (Y) y la variable independiente es la edad (X1). Luego, interpreta los coeficientes del modelo y realiza una predicción del ingreso para un individuo de 35 años.

Modelo de Regresión Múltiple con Selección de Variables:

- **Descripción:** Se desea predecir el ingreso mensual del individuo en función de varias variables.
- **Ejercicio:** Utilizando el método de backward, se puede ajustar un modelo de regresión múltiple donde la variable dependiente es el ingreso mensual (Y) y las variables independientes son todas las variables

disponibles en la base de datos (X1 a X10). Luego, interpreta los coeficientes del modelo y realiza una predicción del ingreso para un individuo con características específicas.

Modelo de Regresión con Interacción:

- **Descripción:** Se desea predecir el ingreso mensual del individuo en función de varias variables.
- **Ejercicio:** Utilizando las variables Edad del individuo (en años) y Nivel de educación (1 = bajo, 2 = medio, 3 = alto) se puede realizar la interacción que puede tener la edad y el nivel de educación con sus ingresos, es decir evaluar si el efecto de la edad en el ingreso mensual varía según el nivel de educación. Luego, interpreta los coeficientes del modelo y realiza una predicción del ingreso para un individuo con características específicas.

Modelo de Regresión Polinómica:

- **Descripción:** Se desea predecir el ingreso mensual del individuo en función de varias variables.
- **Ejercicio:** Utilizando el método de regresión polinómica ajustar una regresión polinómica de segundo grado para capturar posibles relaciones entre la edad y el ingreso mensual. La variable dependiente es Ingreso mensual del individuo (en dólares) y la predictoras es Edad del individuo (en años), la puede colocar a nivel dos (ejemplo de instrucción en R: `poly(X1, degree = 2)`, o, `I(X1^2)`). Luego, interpreta los coeficientes del modelo y realiza una predicción del ingreso para un individuo con características específicas.

Modelo de Regresión Lasso y Ridge:

- **Descripción:** Se desea ajustar modelos de regresión con regularización Lasso y Ridge para predecir el ingreso mensual del individuo.
- **Ejercicio:** Ajusta un modelo de regresión Lasso y un modelo de regresión Ridge donde la variable dependiente es el ingreso mensual (Y) y las variables independientes son todas las variables disponibles en la base de datos (X1 a X10). Luego, interpreta los coeficientes del modelo y compara los resultados con el modelo de regresión múltiple sin regularización.

Modelo de Regresión Logística Binaria:

- **Descripción:** Se desea predecir si un individuo es fumador o no en función de varias variables.
- **Ejercicio:** Ajusta un modelo de regresión logística binaria donde la variable dependiente es si el individuo es fumador (1 = sí, 0 = no) y las variables independientes son todas las variables disponibles en la base de datos (X1 a X10). Luego, interpreta los coeficientes del modelo y realiza una predicción de la probabilidad de ser fumador para un individuo con características específicas.

Modelo de Regresión Logística Multinomial:

- **Descripción:** Se desea predecir el nivel de educación del individuo en función de varias variables.
- **Ejercicio:** Ajusta un modelo de regresión logística multinomial donde la variable dependiente es el nivel de educación (bajo, medio, alto) y las variables independientes son todas las variables disponibles en la

base de datos (X1 a X10). Luego, interpreta los coeficientes del modelo y realiza una predicción del nivel de educación para un individuo con características específicas.

Estos ejemplos han sido preparados considerando los modelos utilizados. Si desea probar otro modelo, está en total libertad. Los ejemplos son propuestos, si desea utilizarlo está en total libertad; también puede plantear su propio problema a analizar.

Entregables y desarrollo del trabajo

Para los entregables del trabajo serán:

1. El trabajo se puede realizar en grupos (máximo 4 personas) o de forma individual.
2. Documento en formato pdf con los ítems de la tabla **Criterios de evaluación. Por favor añadir** Además, si se ha utilizado R y RStudio para realizar los modelos, por favor adjuntar. Si no se realizó los modelos con el programa R y RStudio, solo adjuntar el documento con el desarrollo de los criterios de evaluación.

Fechas de entrega

De acuerdo con lo acordado en el curso, el trabajo se enviará el día 5 de abril de 2024.

Contacto para información

Juan Carlos Llivisaca

juan.llivisaca@ucuenca.edu.ec ; juancarlosllivisaca@gmail.com ; Whatsapp +593987119669.

Cuenca - Ecuador