

Ejercitación regresión múltiple

1

Problema 1: Considere el modelo lineal con intercepto y sin variable independiente:

$$Y_i = \mu + e_i, e_i \sim (0, \sigma^2) \quad i = 1 \dots n,$$

- a. Representelo matricialmente
- b. Construya las ecuaciones normales en este caso
- c. Obtenga el estimador de cuadrados mínimos

2

Problema 2:

Construya un modelo lineal que represente el comportamiento de las muestras de tamaño n y m de dos poblaciones normales (Y^A, Y^B), con medias respectivas μ_A, μ_B ; independientes y con varianzas iguales.

Represente su modelo matricialmente.

Obtenga los estimadores de cuadrados mínimos de las medias y el estimador de la varianza común

3

Una situación aplicada

Considere un experimento aleatorio en el que 30 personas son divididas aleatoriamente en dos grupos uno de 16 personas y otro de 14. A los miembros del primer grupo, se les administra una droga llamada "B" y, a los del segundo, otra droga llamada "G".

Se desea saber cómo las drogas influyen en el tiempo de coagulación.

Planteando el problema con el modelo anterior, la solución del problema conduce a la comparación de medias de dos poblaciones.

¿Se podría ver la modelación de otra forma que me permita hacer otro tipo de formulación del problema más cercana a lo que ya vimos en el análisis de regresión?

4

Una situación aplicada

Considere un experimento aleatorio en el que 30 personas son divididas aleatoriamente en dos grupos uno de 16 personas y otro de 14. A los miembros del primer grupo, se les administra una droga llamada “B” y, a los del segundo, otra droga llamada “G”.

Se desea saber cómo las drogas influyen en el tiempo (en minutos) de coagulación.

$$tiempo_i = \beta_0 + \beta_1 * droga_i + e_i$$

$$droga = \begin{cases} 1, & \text{si recibió droga B} \\ 0, & \text{si recibió droga G} \end{cases}$$

¿En este caso qué significan los parámetros?

¿Cómo serían los estimadores de cuadrados mínimos en este caso?

5

Ajuste el modelo de regresión y responda a la pregunta

- (b) A un nivel de significancia del 5% los datos de la muestra evidencian que el tiempo de coagulación es mayor para quienes recibieron la droga G que para quienes recibieron la droga B? (valor 0.3)

$$tiempo_i = \beta_0 + \beta_1 * droga_i + e_i$$

$$droga = \begin{cases} 1, & \text{si recibió droga B} \\ 0, & \text{si recibió droga G} \end{cases}$$

6

Datos: supervisor

Problema de psicología industrial

Estudio con empleados administrativos de una empresa

Interés del estudio: satisfacción de los empleados con el supervisor

Y: variable con la satisfacción general del empleado con el trabajo del supervisor

Análisis: Valorar variables que influyen o se relacionan con esa satisfacción

Variables de relaciones interpersonales

X1: Atención a quejas del empleado

X2: Permitir privilegios especiales

X5: Críticas excesivas por resultados deficientes

Variables relacionadas con el tipo de empleo

X3: Oportunidad de aprender cosas nuevas

X4: Mejoras salariales basadas en el rendimiento

Percepción del empleado sobre su situación en la compañía

X6: Tasa de avance hacia mejores empleos

- (1) Ajuste la recta de cuadrados mínimos, analice los errores de estimación y construya los correspondientes intervalos de confianza
- (2) Suponiendo homocedasticidad, estime la varianza en el modelo
- (3) Valore el efecto conjunto de las variables en el modelo y el efecto individual de cada variable
- (4) Valore el efecto individual de X1 y el efecto conjunto de las variables relacionadas con el tipo de empleo, ¿qué piensa de los resultados de su análisis?. Para responder esta pregunta ajuste los modelos de regresión correspondientes
- (5) ¿Qué piensa del efecto de las variables y del modelo de regresión que construyó?

Error
0.01

7

Datos: fumar

Problema de salud

Estudio de factores que influyen en el hábito de fumar

Variables

Edad: Mediana de la edad

ES: Porcentaje de personas de más de 25 años con educación superior

ingreso: ingreso per cápita

color: % personas raza negra

mujeres: % mujeres

precio: precio promedio (centavos / dólar) de una cajetilla

ventas: per cápita de cajetillas vendidas

Error
0.01

8

Datos: exámenes

Problema de rendimiento académico

Estudio de resultados docentes en un curso de estadística

Interés del estudio: ver como dos parciales influyen en el resultado final de la asignatura

Variables

F: nota final

P1: nota parcial 1

P2: nota parcial 2

(1) Ajuste los tres modelos de regresión siguientes

(a) Con las dos variables independientes

(b) Con cada una por separado

En cada caso ajuste la recta de cuadrados mínimos, analice los errores de estimación y construya los correspondientes intervalos de confianza. Valore en cada caso el efecto de las variables y la calidad del ajuste.

(2) Si tuviera que seleccionar alguno de los modelos de regresión para modelar la influencia de factores relevantes sobre el examen final, ¿cuál utilizaría?

(3) Si tuviera que predecir el puntaje final de un estudiante con valores $P1=78$ y $P2=85$, ¿cuál sería su predicción en este caso?

Error
0.01