XS-2130 Modelos de Regresión Aplicados II Sem 2020 Grupo 2

Práctica para clase.

1. Preguntas de cálculos y teóricas a partir de procedimientos en cómputo

Se le da el archivo vacaslecherasA.Rdata en formato de R. Se refiere a datos recolectados por un agrónomo que está interesado en los nutrientes que mejoran la producción de leche en vacas. Según la literatura, él encontró que hay tres nutrientes que ayudan a maximizar la producción lechera: nut1, nut2 y nut3. Todos los nutrientes están en gramos por centímetro cúbico (g/cm3). Diseña un experimento en el que a 22 vacas les da distintas combinaciones de los nutrientes en un diseño experimental complejo. La variable dependiente es litros/día en cada vaca durante 10 días. Con base en el archivo de datos vacaslecherasA.Rdata, en donde las variables predictoras (conocidas en experimentos como factores tratamiento) se llaman nut1, nut2 y nut3, y la producción lechera se llama litrosxdia, conteste las siguientes preguntas:

1. Con gráficos explore la linealidad entre la variable dependiente y cada una de las independientes. Escriba en menos de 5 líneas con cuáles variables se ve más clara la asociación lineal. Justifique (4 ptos.)
2. Estime un modelo de regresión en el que los litros de leche por día estén en función de los nutrientes, y a partir de este modelo, conteste las siguientes preguntas (Si estima el modelo incorrectamente, pero las preguntas las contesta correctamente con base en el modelo malo, se le asignará parte de los puntos, pero no todos):
   1. Independientemente de la probabilidad asociada, interprete el coeficiente de la variable nut1 y el intercepto (4 ptos.).
   2. Calcule los coeficientes de regresión estandarizados usando el procedimiento de las matrices de correlación y diga cuál variable está más fuertemente asociada con los litros diarios de leche (4 ptos.)
   3. Contraste la hipótesis nula de que los coeficientes estandarizados para nut2 y nut3 son iguales, con un =0.05. Escriba las hipótesis nula y alternativa y conteste apropiadamente al resultado de la prueba de hipótesis (5 ptos.)
   4. En términos prácticos relacionados con el aumento de la producción lechera, qué significa el resultado de la pregunta anterior? (2 ptos.)
   5. Interprete el coeficiente de determinación múltiple de la “salida” del lm (2 ptos.)
   6. Calcule matriz de variancia-covariancia de los betas y explique, con base en los resultados del procedimiento, por qué las covariancias entre i y j son 0 (exceptuando 0) (3 ptos.)
3. Se le da abajo el programa que escribió el profesor Ricardo Alvarado que contiene la función para simular datos de un modelo de regresión. Dicha función se usó posteriormente para representar gráficamente el error estándar. Simule 1000 veces el generar el par de variables (X y Y) con los valores de los parámetros dados en el programa (a=2; b=3; sig=5; n=10). Usando esa función, estime el valor ajustado o predicho cuando X=5.5 (la media teórica de la variable X) para cada una de las 1000 simulaciones y asígnelo a un vector. Haga lo mismo para el valor ajustado cuando X=10 (el valor teórico máximo de la variable X). Haga un histograma de cada uno de los dos vectores y compárelos. Diga qué diferencias hay en la variabilidad de los histogramas y a qué se debe dicha variabilidad (4 ptos.).

sim.beta=function(a,b,sig,n,min=1,max=10){

x=runif(n,min,max)

y=a+b\*x+rnorm(n,0,sig)

return(list("x"=x,"y"=y))

}

a=2; b=3; sig=5; n=10

s1=sim.beta(a,b,sig,n)

s1

summary(lm(s1$y~s1$x))

anova(lm(s1$y~s1$x))