

Ejemplos de Aplicación - Semana 5

Problema de Dureza Shore

Usted trabaja como Ingeniero de Producto en una autopartista. Junto con el área de Investigación y Desarrollo están desarrollando las especificaciones para una nueva variedad de pastillas de freno. Uno de los atributos más importantes de este producto es la dureza del mismo (medida mediante ensayo de Dureza Shore), la cual, se supone que depende principalmente del porcentaje de azufre y neoprene que posee la ferro-aleación utilizada para fabricar estos discos de freno.

Para esto, se llevó a cabo un experimento realizándole ensayos de Dureza Shore para 9 discos de freno con distintos porcentajes de azufre y neoprene (se utilizaron porcentajes de azufre de 5%, 5.26% y 5.51%; y para cada uno de estos se varió el porcentaje de neoprene al 3%, 4.23% y 5.57%).

Datos: *Ejemplo_Aplicacion_Semana_5.xlsx* Hoja: *Problema1*

Cargamos las librerías

```
# Para importar datos de excel
library(readxl)
# Para graficos más profesionales
library(ggplot2)
# Librerías

library(dplyr)
```

Cargamos los datos del archivo

```
# IMPORTAR DATOS
datos <- read_excel("Ejemplo_Aplicacion_Semana_5.xlsx",
                    sheet = "Problema1")
```

a) Especifique el modelo lineal que utilizará para explicar la Dureza Shore a partir del porcentaje de azufre y de neoprene. Estime el modelo. Indique el la proporción de variabilidad en la dureza que es explicada por el azufre y el neoprene en forma conjunta. Analice la significancia de los coeficientes. JUSTIFIQUE el sentido de los test de hipótesis que utiliza y determine a partir de esto el nivel de significación exacto de cada uno de los test de hipótesis.

```
modelo <- lm(formula = dureza ~ azufre + neoprene, data = datos)
resumen <- summary(modelo)
resumen
```

b) Grafique en 3D

Solo se realiza este gráfico con fines didácticos.

Gráfico del diagrama de dispersión

```
library(rgl) # libreria para graficar en 3 dimensiones

# grafico de dispersion en 3 dimensiones
plot3d(x=datos$azufre, y=datos$neoprene, z=datos$dureza,
       col = "blue",
       xlab = "Porcentaje de azufre", ylab = "Porcentaje de neoprene", zlab = "Dureza shore",
       type = "s")
```

Gráfico del diagrama de dispersión con plano de regresión

```
# Para poder graficar el plano de regresión, guardamos los coeficientes del modelo en 3 variables
b_0 <- resumen$coefficients[1,1]
b_azufre <- resumen$coefficients[2,1]
b_neoprene <- resumen$coefficients[3,1]

# Gráfico de dispersión
plot3d(x=datos$azufre, y=datos$neoprene, z=datos$dureza,
      col = "blue",
      xlab = "Porcentaje de azufre", ylab = "Porcentaje de neoprene", zlab = "Dureza shore",
      type = "s")

# Agregamos el gráfico del plano de regresión
# El software grafica: a*x + b*y + c*z + d = 0
planos3d(a = -b_azufre, b = -b_neoprene, c = 1, d = -b_0, col = "lightgrey")
```

c) Por otra parte, el área de Investigación y Desarrollo le pide que incluya en el informe del experimento como incide en la Dureza Shore las variaciones en las cantidades de azufre y neoprene. Interprete RIGUROSAMENTE.

d) Del área de Investigación y Desarrollo le indican que esta nueva variedad de discos de freno, teóricamente debe llevar un 5% de azufre y un 4,23% de neoprene. A su vez desde el área de Calidad le piden que, a partir de lo hallado experimentalmente especifique entre que valores se encontrará la dureza media para un producto con esta composición con 90% de confianza.

```
# Armamos un nuevo data.frame con los datos a pronosticar
datos_nuevos <- data.frame(azufre = 5,
                           neoprene = 4.23)

# Vemos en pantalla el nuevo dataset
datos_nuevos

predict(object = modelo,
        newdata = datos_nuevos,
        interval="confidence",
        level = 0.90)
```

e) Para la nueva composición, debe especificar una dureza para dar de garantía de forma tal que se sea superada con 99% de seguridad.

```
# Armamos un nuevo data.frame con los datos a pronosticar
datos_nuevos <- data.frame(azufre = 5,
                           neoprene = 4.23)

# Vemos en pantalla el nuevo dataset
datos_nuevos

predict(object = modelo,
        newdata = datos_nuevos,
        interval="prediction",
        level = .98)
```

Problema de la demanda de energía eléctrica

Retomamos el problema 1 de los ejemplos de aplicación de la semana 3.

Una compañía eléctrica estudia la influencia de la temperatura ambiente en el consumo domiciliario de electricidad de una ciudad [MWh/día]. Durante los días de baja temperatura los habitantes de la ciudad usan energía eléctrica para calefacción. Durante los días de alta temperatura los acondicionadores de aire consumen una cantidad significativa de energía. Se toma una muestra, seleccionando al azar varios días.

Datos: *Ejemplo_Aplicacion_Semana_3.xlsx*

```
# IMPORTAR DATOS
datos <- read_excel("Ejemplo_Aplicacion_Semana_3.xlsx")

# Le cambio el nombre a las columnas/variables para que sean mas sencillos
colnames(datos) <- c("temperatura", "consumo")
```

Este es el mejor modelo que habíamos obtenido

```
# Gráfico de dispersion entre la variable explicativa y la de respuesta (mas profesional)
modelo4 <- lm(formula = consumo ~ I((temperatura - 18)^2), data = datos)
resumen4 <- summary(modelo4)
resumen4

# Gráfico de dispersion entre la variable explicativa y la de respuesta (mas profesional)
datos$fit_4 = predict(modelo4)

ggplot(datos) +
  geom_point(aes(x = temperatura, y = consumo)) +
  # Establecemos tema bw (estilo de gráfico mas sobrio)
  theme_bw() +
  # Agregamos en gris ejes cartesianos
  # geom_vline(xintercept = 0, color = "grey")+
  geom_hline(yintercept = 0, color = "grey")+
  # Agregamos el gráfico del modelo lineal
  geom_line(aes(x = temperatura, y = fit_4), colour="blue", size = 1)
#method = "lm", formula = consumo ~ temperatura, color = 'red')
```

El modelo anterior tiene una limitación: requiere que se le indique al modelo el valor del $x_{\text{vertice}} = t_0$ de la parábola. En nuestro caso utilizamos 18°C. Si bien es un valor razonable, también podríamos pedirle al modelo que estime este parámetro.

a) Explique por qué no puede linealizar el modelo tal como está definido para aplicarle un modelo lineal:

$$\text{consumo} = \beta_0 + \beta_1 (\text{temperatura} - t_0)^2$$

b) Una posibilidad es estimar una parábola completa con todos sus términos: $\text{consumo} = \beta_0 + \beta_1 \text{temperatura} + \beta_2 \text{temperatura}^2$.

Estime este modelo y con ambos modelos en el mismo gráfico. ¿Es válido el modelo?

```
# Gráfico de dispersion entre la variable explicativa y la de respuesta (mas profesional)
modelo5 <- lm(formula = consumo ~ temperatura + I(temperatura^2), data = datos)
resumen5 <- summary(modelo5)
resumen5

datos$fit_4 = predict(modelo4)
datos$fit_5 = predict(modelo5)

ggplot(datos) +
  geom_point(aes(x = temperatura, y = consumo)) +
  # Establecemos tema bw (estilo de gráfico mas sobrio)
  theme_bw() +
  # Agregamos en gris ejes cartesianos
  # geom_vline(xintercept = 0, color = "grey")+
  geom_hline(yintercept = 0, color = "grey")+
  # Agregamos el gráfico del modelo lineal
  geom_line(aes(x = temperatura, y = fit_4), colour="blue", size = 1)+
  geom_line(aes(x = temperatura, y = fit_5), colour="green", size = 1)
#method = "lm", formula = consumo ~ temperatura, color = 'red')
```

- c) Explique por que no es sencillo interpretar los valores de los parámetros estimados β_1 y β_2 como lo veníamos haciendo hasta ahora.
- d) Calcule el X_{vertice} de la parábola, compárela con el valor que habíamos asumido gráficamente. ¿Está más o menos en línea con lo que habíamos supuesto? Utilice para determinar un modelo simplificado $\text{consumo} = \beta_0 + \beta_1 (\text{temperatura} - t_0)^2$. Evalúe el modelo.