Ejemplos de Aplicación - Semana 3

Problema de la demanda de energía eléctrica

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*

Cargamos las librerías

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

library(dplyr)
```

Cargamos los datos del archivo

```
# 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")
```

a) Graficar las variables registradas. Plantear la relación causal. Analice la curvatura. Mas allá de la estadística, ¿le encuentra un sentido práctico a la forma de la curvatura que observa?

Generamos el gráfico

```
# Gráfico de dispersion entre la variable explicativa y la de respuesta (mas profesional)
ggplot(datos, aes(x = temperatura, y = consumo)) +
  geom_point() +
  # fijamos los límites inferiores de los ejes en 0.
  # NA = Not avaiable. Dejamos que R decida la mejor opción
  xlim(0, NA)+
  ylim(0, NA)+
  # 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")
```

b) Estimar y evaluar el modelo $Y = \beta_0 + \beta_1 * X + \epsilon$. Grafique las observaciones junto con el modelo propuesto. Es válido el modelo desde el punto de vista estadístico. ¿Que opina de este modelo?

```
# Gráfico de dispersion entre la variable explicativa y la de respuesta (mas profesional)
modelo1 <- lm(formula = consumo ~ temperatura, data = datos)
resumen1 <- summary(modelo1)
resumen1
```

```
# Gráfico de dispersion entre la variable explicativa y la de respuesta (mas profesional)
ggplot(datos, aes(x = temperatura, y = consumo)) +
  geom_point() +
  # fijamos los límites inferiores de los ejes en 0.
  # NA = Not avaiable. Dejamos que R decida la mejor opción
  xlim(0, NA)+
  ylim(0, NA)+
  # 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")+
```

```r
  # Agregamos el gráfico del modelo lineal
  geom_smooth(se = FALSE, method = lm)
```

c) Estimar y evaluar el modelo $Y = \beta_0 + \beta_1 * X^2 + \epsilon$.

¿Qué opina de este modelo? ¿Se trata de un modelo lineal?

```r
# Gráfico de dispersion entre la variable explicativa y la de respuesta (mas profesional)
modelo2 <- lm(formula = consumo ~ I(temperatura^2), data = datos)
resumen2 <- summary(modelo2)
resumen2
```

```r
# Gráfico de dispersion entre la variable explicativa y la de respuesta (mas profesional)
datos$fit_2 = predict(modelo2)

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_2), colour="blue", size = 1)
#method = "lm", formula = consumo ~ temperatura, color = 'red')
```

d) Estimar y evaluar el modelo $Y = \beta_0 + \beta_1 * X^3 + \epsilon$.

¿Qué opina de este modelo? ¿Se trata de un modelo lineal?

```r
# Gráfico de dispersion entre la variable explicativa y la de respuesta (mas profesional)
modelo3 <- lm(formula = consumo ~ I(temperatura^3), data = datos)
resumen3 <- summary(modelo3)
resumen3
```

```r
# Gráfico de dispersion entre la variable explicativa y la de respuesta (mas profesional)
datos$fit_3 = predict(modelo3)

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_3), colour="blue", size = 1)
#method = "lm", formula = consumo ~ temperatura, color = 'red')
```

e) Se plantea el siguiente modelo: $Y = a * (X - 18)^2 + b$, siendo a y b pendiente y ordenada al origen, respectivamente. ¿Es este un modelo lineal en los parámetros? Justifique. Grafique, estime y evalúe este modelo. ¿Qué diferencias tiene este modelo con respecto a los anteriores?

```
# 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')
```

f) Con el último modelo, calcular el consumo diario esperado en cada estación del año, si las temperaturas medias en la ciudad son: 37°C en verano, 10°C en otoño, 0°C en invierno, y 25°C en primavera. Construir intervalos al 90% de confianza. Grafique el intervalo de confianza y de predicción en el gráfico anterior.

```
# Observe la diferencia entre estos 2 comandos
predict(modelo4)
predicciones <- predict(modelo4, interval = "confidence", level = .95) # interval = 'predicti
# El primero de ellos devuelve un vector con los pronósticos y pico del modelo
# El segundo comando devuelve un data.frame con fit = y pico, y lwr - upr son los límites in
head(predicciones)

# Hasta aquí pronosticamos puntos que están en dataset original. Para estimar observacione so
data_predict <- data.frame(temperatura = -10:40)
head(data_predict)
predicciones <- predict(modelo4, newdata = data_predict, interval = "confidence", level = .95
# cbind pega columnas de dataframes y genera un nuevo data.frame
data_predict <- cbind(data_predict, predicciones)
head(data_predict)
```

g) Un cuestionamiento que se le puede hacer al modelo anterior es haber fijado el valor $X = 18$ para centar la parábola, por ello se propone un nuevo modelo $Y = a * (X - c)^2 + b$, siendo a, b y c parámetros a estimar del modelo. ¿Es este un modelo lineal en los parámetros? Justifique. Indique por que razón este modelo no es estomable así como se presenta.