Ejemplos de Aplicación - Semana 7

Problema de los dos catalizadores

Una empresa química desea modelas la influencia de un catalizador en el tiempo de reacción para obtener un determinado producto. Para ello recolecta los registros de las últimas 100 reacciones realizadas conteniendo información de los tiempos de reaccion requeridos, expresados en horas, la candidad de componente D expresada en Kg por lote y el tipo de cataliador utilizado. Esta reacción requiere si o sí un catalizador.

Datos: *Ejemplo_Aplicacion_Semana_7.xlsx* Hoja: *Problema1*

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_7.xlsx",
                    sheet = "Problema1")
```

a) Dejando de lado (por ahora) el catalizador, grafique un diagrama de dispersión para evaluar la relación entre la cantidad de componente D y el tiempo de reacción. Estime el modelo lineal. Valide el modelo. Genere un gráfico para evaluar el supuesto de homocedasticidad. ¿Se cumple el supuesto? el interprete el valor de la pendiente en téminos del problema. Agregue al gráfico de dispersión la recta de regresión estimada.

```r
ggplot(datos)+
  geom_point(aes(kg_D, tiempo))+
  theme_bw()

modelo1 <- lm(tiempo ~ kg_D, data = datos)

summary(modelo1)

plot(modelo1, 1)
```

```r
datos$pred1 = predict(modelo1)

ggplot(datos)+
  geom_point(aes(kg_D, tiempo))+
  # Agregamos una línea con los valores predichos para mostrar la recta de regresión
  geom_line(aes(kg_D, pred1), color = 'blue')+
  theme_bw()
```

b) Estime ahora un nuevo modelo incluyendo ambas variables. Valide el modelo. Genere un gráfico para evaluar el supuesto de homocedasticidad. ¿Se cumple el supuesto? Escriba las ecuaciones de regresión para cada uno de los catalizadores. Interprete el valor ambas pendientes en téminos del problema.

$$tiempo = \beta_0 + \beta_1 KgD + \beta_2 Catalizador$$

```r
modelo2 <- lm(tiempo ~ kg_D + cat, data = datos)

summary(modelo2)

plot(modelo2, 1)
```

```
datos$pred2 = predict(modelo2)

ggplot(datos)+
  geom_point(aes(kg_D, tiempo, color = cat))+
  # Agregamos una línea con los valores predichos para mostrar la recta de regresión
  # los parametros group color son para separar las rectas en 2
  geom_line(aes(kg_D, pred2, group = cat, color = cat), size = 1)+
  # La siguiente linea establece la paleta de colores que se va a utilizar
  scale_color_brewer(palette="Set1")+
  theme_bw()
```

c) Estime ahora un otro modelo que incluya la interacción enter el catalizador y los kg de componente D (se llama interacción a los productos entre variables). Valide el modelo. Genere un gráfico para evaluar el supuesto de homocedasticidad. ¿Se cumple el supuesto? Escriba las ecuaciones de regresión para cada uno de los catalizadores. Interprete el valor ambas pendientes en téminos del problema.

$$tiempo = \beta_0 + \beta_1 \ KgD + \beta_2 \ KgD \ Catalizador$$

```
# En las formulas de R, las interacciones se incluyen con ":"
modelo3 <- lm(tiempo ~ kg_D + kg_D:cat, data = datos)

summary(modelo3)

plot(modelo3, 1)
```

```
datos$pred3 = predict(modelo3)

ggplot(datos)+
  geom_point(aes(kg_D, tiempo, color = cat))+
  # Agregamos una línea con los valores predichos para mostrar la recta de regresión
  # los parametros group color son para separar las rectas en 2
  geom_line(aes(kg_D, pred3, group = cat, color = cat), size = 1)+
  # La siguiente linea establece la paleta de colores que se va a utilizar
  scale_color_brewer(palette="Set1")+
  theme_bw()
```

d) Estime ahora un último modelo que permita modelar un cambio de pendiente y de ordenada al origen de manera simultánea. Valide el modelo. Genere un gráfico para evaluar el supuesto de homocedasticidad. ¿Se cumple el supuesto? Escriba las ecuaciones de regresión para cada uno de los catalizadores. Interprete el valor ambas pendientes en téminos del problema.

$$tiempo = \beta_0 + \beta_1 \ Catalizador + \beta_1 \ KgD + \beta_2 \ KgD \ Catalizador$$

```
# En las formulas de R, las interacciones se incluyen con ":"
modelo4 <- lm(tiempo ~ kg_D + cat + kg_D:cat, data = datos)

summary(modelo4)

plot(modelo4, 1)
```

```
datos$pred4 = predict(modelo4)

ggplot(datos)+
  geom_point(aes(kg_D, tiempo, color = cat))+
  # Agregamos una línea con los valores predichos para mostrar la recta de regresión
  # los parametros group color son para separar las rectas en 2
  geom_line(aes(kg_D, pred4, group = cat, color = cat), size = 1)+
  # La siguiente linea establece la paleta de colores que se va a utilizar
  scale_color_brewer(palette="Set1")+
  theme_bw()
```

Problema proyecciond demanda con variables indicadoras

Una empresa cervecera que opera en el país, desea realizar un modelo para pronóstico de la demanda global en litros de cerveza para los próximos trimestres. Para ello utiliza la infromación de ventas totales (en litros) para los últimos años con una granularidad trimestral.

Datos: *Ejemplo_Aplicacion_Semana_7.xlsx* Hoja: *Problema2*

```
# IMPORTAR DATOS
datos <- read_excel("Ejemplo_Aplicacion_Semana_7.xlsx",
                    sheet = "Problema2")
```

a) Grafique la serie de ventas y observe que particularidades posee

```
ggplot(datos)+
  geom_line(aes(periodo, ventas))+
  ylim(0, NA)+
  theme_bw()
```

b) Estime un modelo que permita determinar la tendencia. Valide el modelo el interprete la pendiente en términos del problema. Grafique el diagrama de dispersion junto al modelo de regresión.

$$ventas = \beta_0 + \beta_1 \; periodo$$

```
modelo1 <- lm(formula = ventas ~ periodo, data = datos)
summary(modelo1)
datos$pred1 <- predict(modelo1)
ggplot(datos)+
  geom_line(aes(periodo, ventas))+
  geom_line(aes(periodo, pred1), size = 1, color = 'blue')+
  ylim(0, NA)+
  theme_bw()
```

c) Estime un nuevo modelo que permita determinar la tendencia y tenga en cuenta la estacionalidad utilizando variables dummies. Valide el modelo el interprete todas los coeficientes del modelo en términos del problema. ¿Tiene sentido en este caso la interpretación de β_0? Grafique el diagrama de dispersion junto al modelo de regresión.

```
modelo2 <- lm(formula = ventas ~ periodo + trimestre, data = datos)
summary(modelo2)
datos$pred2 <- predict(modelo2)
ggplot(datos)+
  geom_line(aes(periodo, ventas), size = 1)+
  geom_line(aes(periodo, pred2), size = 1, color = 'blue')+
  ylim(0, NA)+
  theme_bw()
```

d) Utilizando el modelo, proyecte las ventas del próximo año. Grafique el intervalo de predicción obtenido.

```
datapred <- data.frame(periodo = 55:58,
                       trimestre = c("Q3", "Q4", "Q1", "Q2"))
pred <- predict(modelo2, newdata = datapred, interval = "prediction")
datos_plot_pred <- cbind(datapred,pred)
ggplot(datos)+
  geom_line(aes(periodo, ventas), size = 1)+
  geom_line(aes(periodo, pred2), size = 1, color = 'blue')+
  geom_line(data = datos_plot_pred, aes(periodo, fit), size = 1, color = 'red')+
  geom_line(data = datos_plot_pred, aes(periodo, upr), color = 'red')+
  geom_line(data = datos_plot_pred, aes(periodo, lwr), color = 'red')+
  ylim(0, NA)+
  theme_bw()
```

e) Para pensar... ¿Estoy extrapolando? Si lo estuviera, ¿Son válidos los resultados? ¿Que hubiera pasado con los pronosticos del modelo durante la pandemia? ¿Pierde validez el modelo?.