

```

---
title: "Articulo 1"
author: "Beltran Ruiz Logan Daniel"
date: '2022-06-23'
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

##CREACION DE LAS VARIABLES

```{r}
library(faux)

dfa <- rnorm_multi(n = 1000,
 mu = c(67, 30, 30, 320),
 sd = c(2, 3, 5, 10),
 varnames = c('edad', 'dap', 'rto', 'clolA'),
 r = c(0.4, 0.6, 0.5, 0.6, 0.7, 0.8))

dfa$hibrido <- round(runif(n = 1000, min = 0,max = 1.2))

w <- 0.5 * dfa$clolA - 0.01 * dfa$dap - 0.6 * dfa$rto - 0.02 * dfa$edad
SI ABORTO ES IGUAL A 1 Y NO ABORTO ES IGUAL A 0
Si = 1
No = 0
dfa$abortos <- ifelse(w > 140, Si, No)
dfa

```

## Análisis univariado para CADA VARIABLE

```{r}

univariable_edad <- glm(abortos ~ edad, family = binomial, data = dfa)
summary(univariable_edad)
```

```{r}
univariable_dap <- glm(abortos ~ dap, family = binomial, data = dfa)
summary(univariable_dap)
```

```{r}
univariable_clolA <- glm(abortos ~ clolA, family = binomial, data = dfa)
summary(univariable_clolA)
```

```{r}
univariable_h <- glm(abortos ~ hibrido, family = binomial, data = dfa)
summary(univariable_h)
```

```{r}
univariable_rto <- glm(abortos ~ rto, family = binomial, data = dfa)
summary(univariable_rto)

```

```
```
```

```
```{r, warning = FALSE, message=FALSE}
modell1 <- glm(abortos ~ edad + dap + hibrido + rto + clolA, family =
binomial, data = dfa)
summary(modell1)
```
```

```
```{r, warning = FALSE, message=FALSE}
model2 <- glm(abortos ~ edad + dap + rto + clolA, family = binomial, data =
dfa) # Eliminando las variablse con el pvalue más alto, que seria el
estadisticamente mas insignificante
summary(model2)
```
```

```
##Se comparan los cambios en los coeficientes para cada variable restante
en el modelo2
##Delta beta
```

```
```{r, warning = FALSE, message=FALSE}
delta.coef <- abs((coef(model2)-coef(modell1)[-c(4)])/coef(modell1)[-c(4)])
round(delta.coef, 3)
```
```

```
##LOS CAMBIOS NO SON IMPORTANTES YA QUE NO SUPERAN EL 20%, LA VARIABLES NO
DEBIO SER ELIMINADAS
```

```
```{r, warning = FALSE, message=FALSE}
model3 <- glm(abortos ~ edad + rto + clolA, family = binomial, data = dfa)
Eliminando la variable con el pvalue más alto
summary(model3)
```
```

```
```{r, warning = FALSE, message=FALSE}
delta.coef <- abs((coef(model3)-coef(model2)[-c(3)])/coef(model2)[-c(3)])
round(delta.coef, 3) # Este es el modelo que escogemos
```
```

```
##LOS CAMBIOS SON IMPORTANTES YA QUE SUPERAN EL 20%, LA VARIABLE NO DEBIO
SER ELIMINADA
```

```
```{r, warning = FALSE, message=FALSE}
model_final <- glm(abortos ~ edad + dap + rto + clolA, family = binomial,
data = dfa) # Eliminando la variable con el pvalue más alto
summary(model_final)
```
```

```
```{r, warning = FALSE, message=FALSE}
##COMPARACION DE LOS MODELOS FINALES
library(lmtest)
```

```
lrtest(modell1, model_final)
```
```

```
```{r}
anova(modell1, model_final, test = 'Chisq')
```
```

```
##USANDO PREDICCION PARA CONSTRUIR LOS GRAFICOS
```

```

```{r}
pred <- (model_final$fitted.values)
scatter.smooth(dfa$edad, pred, cex = 0.5)
scatter.smooth(dfa$dap, pred, cex = 0.5)
scatter.smooth(dfa$clolA, pred, cex = 0.5)
scatter.smooth(dfa$rto, pred, cex = 0.5)
```

```{r}
##SUPONIENDO LA PROBABILIDAD DE ABORTO
par(mfrow = c(2,3))
z <- 0.22 * dfa$edad - 0.12 * dfa$clolA + dfa$dap -8 #Variable artificial
pr <- 1/(1+exp(-z)) # Probabilidad de aborto

scatter.smooth(dfa$edad, log(pr/(1-pr)), cex = 0.5)
scatter.smooth(dfa$dap, log(pr/(1-pr)), cex = 0.5)
scatter.smooth(dfa$clolA, log(pr/(1-pr)), cex = 0.5)
scatter.smooth(dfa$rto, log(pr/(1-pr)), cex = 0.5)
```

```{r}
summary(model_final)
hist(model_final$fitted.values)

```

```{r}
edades <- cut(dfa$edad, breaks = 5)
tabla <- table(edades, dfa$abortos)
prop.table(tabla)
```

```{r}
length(dfa$edad)
```

```{r}
probabilidades <- model_final$fitted.values
prob <- ifelse(probabilidades < 0.5, 1, 0)
table(prob, dfa$abortos)
```

```{r}
media_e <- mean(dfa$edad)
colores <- ifelse(dfa$edad < media_e, 'blue', 'green')
plot(model_final$fitted.values, cex = (dfa$edad * 0.015), pch = 20, col =
colores)
abline(h = 0.5, cex = 1.2, col = 'red')
```

```{r}
media_d <- mean(dfa$dap)
colores_d <- ifelse(dfa$dap < media_d, 'blue', 'green')
plot(model_final$fitted.values, cex = (dfa$dap * 0.015), pch = 19, col =
colores_d)
abline(h = 0.5, cex = 1.2, col = 'red')
```

```{r}
media_c <- mean(dfa$clolA)
colores_c <- ifelse(dfa$clolA < media_c, 'blue', 'green')

```

```

plot(model_final$fitted.values, cex = (dfa$clolA * 0.005), pch = 19, col =
colores_c)
abline(h = 0.5, cex = 1.2, col = 'red')
```

## Estudiando las interacciones del modelo

```{r}
model_inter <- glm(abortos ~ edad + dap + clolA + edad:dap, family =
binomial, data = dfa)
summary(model_inter)
```

```{r}
library(dplyr)

data <- dfa |>
 mutate(edad_c = ifelse(edad > mean(edad), 'mayor', 'menor'),
 dap_c = ifelse(dap < mean(dap), 'angosto', 'ancho'),
 cloA_c = ifelse(clolA < mean(clolA), 'baja', 'alta'))
print('edad abortos')
edad_abortos <- table(data$edad_c, data$abortos); edad_abortos
print('dap abortos')
dap_abortos <- table(data$dap_c, data$abortos); dap_abortos
print('cloA abortos')
cloa_abortos <- table(data$cloA_c, data$abortos); cloa_abortos
```

## probabilidades cercanas a 1 es más probabilidad de supervivencia ##

```{r, warning = FALSE, message=FALSE}
mean(dfa$cloA)
library(lmtest)

lrtest(model_final, model1)
```

```{r}
anova(model_final, model2, test = 'Chisq')
```

```{r}
model_inter2=glm(abortos ~edad + dap + clolA , family = binomial, data =
dfa)
summary(model_inter2)
lrtest(model_inter,model_inter2)
Se toma el modelo mas sencillo
```

## SE EXTRAEN LAS PREDICCIONES DE LOS ABORTOS
```{r, warning = FALSE, message=FALSE}

rta= model_inter2$fitted.values
prop_ab <- rta*100
cat_edad <- cut(dfa$edad,breaks = 4)
cat_clolA <- cut(dfa$clolA,breaks=4)

```

```

data_2 <- data.frame(cat_edad, cat_clolA, prop_ab)

tips2 <- data_2 %>%
 group_by(cat_clolA, cat_edad) %>%
 summarise(media_prop_abortos = mean(prop_ab))

Graficando las dos variables
library(ggplot2)
tips2$tip_groups

ggplot(data = tips2) +
 aes(x = cat_edad, y = media_prop_abortos, color = cat_clolA) +
 geom_line(aes(group = cat_clolA))
#Se observa que no hay señal de interaccion
```

## Matriz de confusión valores observados de abortos con valores predichos
de aborto

```{r}
library(ResourceSelection)
cut_prob <- ifelse(fitted(model_inter) > 0.5, 1, 0)
table(model_inter$y, cut_prob)

hoslem.test(model_inter$y, fitted(model_inter))
```

## Como el p_value es mayor a 0.05 no hay diferencias significativas entre
valores observados y predichos ##

```{r}
Predprob<-predict(model_inter,type="response")
plot(Predprob,jitter(as.numeric(dfa$abortos),0.5), cex=0.5, ylab="Abortos")
abline(v = 0.5, col = 'red')
text(x = 0.8, y = 0.8, 'alta probabilidad de abortos, \n predicha y
observada')
text(x = 0.2, y = 0.2, 'alta probabilidad de no abortos, \n predicha y
observada')
```

##SE CONCLUYE LA REVISION DEL ARTICULO

```