

```

---
title: "Trabajo final_datos friabilidad"
author: "Beltran Ruiz Logan Daniel"
date: '2022-06-30'
output: html_document
---

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

## Evaluación de friabilidad (compactado 1 o no compactado 0)

```{r}
####SE USARON LOS ULTIMOS 3 DIGITOS DE LA CEDULA
###1000382874
df <- rnorm_multi(n = 874,
 mu = c(0.5, 300, 30, 35, 0.6),
 sd = c(0.2, 20, 5, 8, 0.15),
 r = c(0.8, 0.7, 0.5, 0.6, 0.8, 0.4, 0.3, 0.4, 0.4, 0.5),
 ## Correlación de pares de variables
 varnames = c('Compact', 'Labranza', 'Arena',
'Arcilla', 'Mecanizado'))
df$Compact <- round(df$Compact)
df
```

## PASO 1: Análisis univariado para CADA VARIABLE

```{r, warning = FALSE, message=FALSE}

univariable_labr <- glm(Compact ~ Labranza, family = binomial, data = df)
summary(univariable_labr)
```

```{r, warning = FALSE, message=FALSE}

univariable_arcll <- glm(Compact ~ Arcilla, family = binomial, data = df)
summary(univariable_arcll)
```

```{r, warning = FALSE, message=FALSE}

univariable_mec <- glm(Compact ~ Mecanizado, family = binomial, data = df)
summary(univariable_mec)
```

```{r, warning = FALSE, message=FALSE}

univariable_labr <- glm(Compact ~ Labranza, family = binomial, data = df)
summary(univariable_labr)
```

## PASO 2: COMPARACION de modelos MULTIVARIABLES

```{r, warning = FALSE, message=FALSE}
modell <- glm(Compact ~ Labranza + Arena + Arcilla + Mecanizado, family =
binomial, data = df)
summary(modell)

```

```
```
```

```
```{r, warning = FALSE, message=FALSE}
Eliminando las variables con el pvalue más alto, que sería el
estadísticamente más insignificante
model2 <- glm(Compact ~ Labranza + Arcilla + Mecanizado, family = binomial,
data = df)
summary(model2)
```
```

```
```{r, warning = FALSE, message=FALSE}
Eliminando las variables con el pvalue más alto, que sería el
estadísticamente más insignificante
model3 <- glm(Compact ~ Labranza + Mecanizado, family = binomial, data =
df)
summary(model3)
```
```

##Se observa que las variables restantes son significantes, por lo que se comparan los cambios en los coeficientes para cada variable restante en el modelo

##Delta beta

```
```{r, warning = FALSE, message=FALSE}
delta.coef <- abs((coef(model3)-coef(model2)[-c(3)]) / coef(model2)[-c(3)])
round(delta.coef, 3)
```
```

```
```{r, warning = FALSE, message=FALSE}
delta.coef <- abs((coef(model3)-coef(model1)[-c(3,4)]) / coef(model1)[-c(3,4)])
round(delta.coef, 3)
```
```

```
```{r, warning = FALSE, message=FALSE}
delta.coef <- abs((coef(model2)-coef(model1)[-c(3)]) / coef(model1)[-c(3)])
round(delta.coef, 3)
```
```

##La función coef() extrae coeficientes estimados de modelo ajustado, por lo que podemos observar que hay un menor impacto en las variables comparadas entre el modelo 2 y el modelo 1; aunque igualmente NO se observa que LOS CAMBIOS SEAN IMPORTANTES ya que los valores de los coeficientes no superan el 20%

##entre el modelo 3 y el modelo 2.

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

```
lrtest(model3, model2)
```
```

##El resultado muestra que los dos modelos no son significativamente diferentes en sus ajustes de datos. En otras palabras, el modelo 3 es tan bueno como el modelo 2

##en el ajuste de datos. Elegimos el modelo 3

```
```{r}
anova(model1, model3, test = 'Chisq')
```
```

```

## PASO 3:      Suposición de linealidad

##USANDO PREDICCIÓN PARA CONSTRUIR LOS GRAFICOS

```{r, warning = FALSE, message=FALSE}
pred <- (model3$fitted.values)
scatter.smooth(df$Labranza, log(pred/(1-pred)), cex = 0.5)
scatter.smooth(df$Arena, log(pred/(1-pred)), cex = 0.5)
scatter.smooth(df$Arcilla, log(pred/(1-pred)), cex = 0.5)
scatter.smooth(df$Mecanizado, log(pred/(1-pred)), cex = 0.5)
```

##PASO 4: Interacciones entre COVARIABLES
```{r, warning = FALSE, message=FALSE}

model.interaction1<-glm(Compact ~ Labranza + Arena + Arcilla + Mecanizado +
Labranza:Arena
 + Labranza:Arcilla + Labranza:Mecanizado +
Arena:Arcilla + Arena:Mecanizado
 + Arcilla:Mecanizado, data=df ,family = binomial)
summary(model.interaction1)

```

```{r, warning = FALSE, message=FALSE}
model.interaction2<-glm(Compact ~ Labranza + Arena + Arcilla + Mecanizado
 + Arcilla:Mecanizado + Arena:Arcilla
 , data=df ,family = binomial)
summary(model.interaction2)

```

```{r}

lrtest(model3,model.interaction2)

```

```{r}
library(dplyr)

data <- df |>
 mutate(lab_c = ifelse(Labranza > mean(Labranza), 'mayor', 'menor'),
 arn_c = ifelse(Arena < mean(Arena), 'bajo','alto'),
 arc_c = ifelse(Arcilla < mean(Arcilla), 'bajo','alto'),
 mec_c = ifelse(Mecanizado < mean(Mecanizado), 'baja', 'alta'))
print('Afección Labranza en la Compactación')
lab_comp <- table(datalab_c, dfCompact); lab_comp
print('Afección Arena en la Compactación')
arn_comp <- table(dataarn_c, dfCompact); arn_comp
print('Afección Arcilla en la Compactación')
arc_comp <- table(dataarc_c, dfCompact); arc_comp
print('Afección Mecanización en la Compactación')
mec_comp <- table(datamec_c, dfCompact); mec_comp
```

```

```

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

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

lrtest(model3, model1)
```

```{r}
##pred son las probabilidades
prob <- ifelse(pred < 0.5, 1, 0)
table(prob, df$Compact)
```

##MODELOS AJUSTADOS POR VARIABLES
```{r}
##LABRANZA
media_l <- mean(df$Labranza)
colores <- ifelse(df$Labranza < media_l, 'blue', 'green')
plot(pred, cex = (df$Labranza * 0.015), pch = 20, col = colores)
abline(h = 0.5, cex = 1.2, col = 'red')
```

```{r}
##ARENA
media_a <- mean(df$Arena)
colores_a <- ifelse(df$Arena < media_a, 'blue', 'green')
plot(pred, cex = (df$Arena * 0.1), pch = 19, col = colores_a)
abline(h = 0.5, cex = 1.2, col = 'red')
```

```{r}
##ARCILLA
media_ar <- mean(df$Arcilla)
colores_b <- ifelse(df$Arcilla < media_ar, 'blue', 'green')
plot(pred, cex = (df$Arcilla * 0.05), pch = 19, col = colores_b)
abline(h = 0.5, cex = 1.2, col = 'red')
```

```{r}
##MECANIZADO
media_m <- mean(df$Mecanizado)
colores_c <- ifelse(df$Mecanizado < media_m, 'blue', 'green')
plot(pred, cex = (df$Mecanizado * 2), pch = 19, col = colores_c)
abline(h = 0.5, cex = 1.2, col = 'red')
```

## Se extraen las predicciones del suelo PARA OBSERVAR SI SE CUMPLE EL
SUPUESTO DE LINEALIDAD

```{r}
model_interaction3=glm(Compact ~ Labranza + Mecanizado
, data=df ,family = binomial)
summary(model_interaction3)

lrtest(model_interaction3,model.interaction2)
```

```

```

```{r, warning = FALSE, message=FALSE}

rta= model_interaction3$fitted.values
prop_ab <- rta*100
cat_lab <- cut(df$Labranza,breaks = 4)
cat_mec <- cut(df$Mecanizado,breaks=4)
data_2 <- data.frame(cat_lab, cat_mec, prop_ab)

tips2 <- data_2 %>%
 group_by(cat_mec, cat_lab) %>%
 summarise(media_prop_co = mean(prop_ab))

Graficando las dos variables
library(ggplot2)
tips2$tip_groups

ggplot(data = tips2) +
 aes(x = cat_lab, y = media_prop_co, color = cat_mec) +
 geom_line(aes(group = cat_mec))
#Se observa que no hay señal de interaccion y hay no hay completa
linealidad
```

## Matriz de confusión valores observados de compactacion con valores
predichos de compactacion

```{r}
library(ResourceSelection)
cut_prob <- ifelse(fitted(model_interaction3) > 0.5, 1, 0)
table(model_interaction3$y, cut_prob)

hoslem.test(model_interaction3$y, fitted(model_interaction3))
```

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

##PASO 5: Evaluación del ajuste del modelo

```{r}
Predprob<-predict(model.interaction2,type="response")
plot(Predprob,jitter(as.numeric(df$Compact),0.5), cex=0.5,
ylab="Compactacion")
abline(v = 0.5, col = 'red')
text(x = 0.8, y = 0.8, 'alta probabilidad de compactacion, \n predicha y
observada')
text(x = 0.2, y = 0.2, 'alta probabilidad de no compactacion, \n predicha y
observada')
```

```