**Universidad de Costa Rica**

**Programa de Posgrado en Estadística**

**SP-1652 Modelos Lineales Generalizados**

**Trabajo Final: Avance II**

**Miguel Coto García**

**Natalia Díaz Ramírez**

1. **Cuadros y gráficos “vacíos”**
2. Potencia de la prueba para sus coeficientes

|  |
| --- |
| **Gráfico 1. Potencia de la prueba para los coeficientes variando el tamaño de la muestra y con mil simulaciones** |
|  |

1. Análisis de residuos de deviancia con simulaciones (linealidad, normalidad) y distintos tamaños de muestra.

|  |  |  |  |
| --- | --- | --- | --- |
| **Gráfico 2. Residuos de deviancia vs Predichos por cantidad de predictores y variando el tamaño de muestra** | | | |
|  | **Predictores** | | |
| **Tamaño muestra** | **P=** | **P=** | **P=** |
| **n=50** |  |  |  |
| **n=100** |  |  |  |
| **n=1000** |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Gráfico 3. Gráfico de normalidad QQ-Plot por cantidad de predictores y variando el tamaño de muestra** | | | |
|  | **Predictores** | | |
| **Tamaño muestra** | **P=** | **P=** | **P=** |
| **n=50** |  |  |  |
| **n=100** |  |  |  |
| **n=1000** |  |  |  |

1. Cuadros con la utilización del modelo con algún ejemplo práctico (los datos pueden ser simulados)

|  |
| --- |
| **Gráfico 4. Cantidad de variables seleccionadas según el lambda** |
|  |

|  |
| --- |
| **Gráfico 5. Cantidad de variables seleccionadas con la regresión lasso y con stepwise para una regresión logística** |
|  |

|  |
| --- |
| **Gráfico 6. Error cuadrático medio para la regresión lasso y la regresión logística** |
|  |

1. **Modelo**

Se adjunta la estimación de los coeficientes del modelo con mle y mle2. Se utilizaron datos recolectados por el Banco Interamericano de Desarrollo (BID) que contiene variables sobre 9558 personas de 2973 hogares para Costa Rica. La variable objetivo se refiere a la categorización usando el PMT que clasifica a los hogares en “No vulnerables”, “Vulnerables”, “Pobres” y “En extrema pobreza”, en este caso se categorizó como 1 pobres o en extrema pobreza y como 0 las categorías restantes.

|  |  |
| --- | --- |
| **mle** | **mle2** |
|  |  |

***Código en R:***

LL <- function(b0,b1,b2,b3,b4,b5,b6,b7,b8,b9,b10) {

n<-nrow(hogar)

x<-model.matrix(pobre~.,hogar)

betas = c(b0,b1,b2,b3,b4,b5,b6,b7,b8,b9,b10)

xb = x%\*%betas

xb\_g = 1/(1+exp(-xb))

vero1 = suppressWarnings(dbinom(hogar$pobre,

size = 1,

prob = xb\_g,

log=FALSE))

parte1<- -sum(log(vero1))

parte2<-sum(abs(betas))

vero= parte1+(n\*lambda\*parte2)

}

# Usando validacion cruzada para determinar el mejor Lambda

x<-model.matrix(pobre~.,hogar)

sal.cv<-cv.glmnet(x[,-1],hogar$pobre,alpha=1, family = binomial)

#El lambda de 1 error estandar

lambda<-sal.cv$lambda.1se

## Valores iniciales a partir de la salida de glmnet

mod1 <- mle(LL, start = list( b0= -1,

b1= 0.38,

b2= -0.32,

b3= -0.32,

b4= 0.65,

b5= 0.0000000001,

b6= -0.005,

b7= 0.0000000001,

b8= -0.35,

b9= 0.05,

b10= 0.0000000001

))

summary(mod1)

mod2 <- mle2(LL, start = list(

b0= -1,

b1= 0.38,

b2= -0.32,

b3= -0.32,

b4= 0.65,

b5= 0.0000000001,

b6= -0.005,

b7= 0.0000000001,

b8= -0.35,

b9= 0.05,

b10= 0.0000000001

))

summary(mod2)