Estadísticas Contemporánea

Alumno:
Huertas Quispe, Anthony Enrique ***Cod:*** *20173728*

Semestre: *2018-II*

Tema: PC 2

Prof. Luis Valdivieso

Pontificia Universidad Católica del Perú
Escuela de Posgrado
Maestría en Estadística

Problema 1.

Definimos ciertas condiciones, entre ellas el hecho de que $\phi = 1$ (parámetro de precisión)

Listing 1: Simulación de 1 base de datos y Respuestas (Método de 2 partes).

```
library("MASS")
library("poLCA")
library(ggplot2)
library(betareg)
library(pryr)
library(nnet)
library(foreign)

set.seed(1993)
n=1000 #numero de sujetos
phi=10 #precision
mu.NM <- c(0,2)
Sigma.NM <- matrix(c(1,0.2,0.2,1),2,2)
p1 = c(0.3,0.5,0.2) #probabilidades de la variable x3

theta = c(1,-0.9,0.35,1.5)
omega = c(0.7,theta) #ln(delta0/P(0<Y<1) = Xt*omega
eta = c(1,theta)     #ln(delta1/P(0<Y<1) = Xt*eta
beta = c(1.35,theta) #logit(mu) = Xt*beta

# Parametros beta observados: beta.observado = (b0,c1,c2,c3,c4)
beta.observado <- c(1.35,theta.observado[1:4])

#Parametros observados: theta.observado = (c1,c2,c3,c4,w0,n0,b0)
theta.observado <- c(theta,0.7,1,1.35)
```

a)

Se tiene un modelo de regresión de dos partes con las siguientes ecuaciones

$$ln\left(\frac{P(Y=0)}{P(0<Y<1)}\right) = 0.7 + x_1 - 0.9 * x_2 + 0.35 * x3_a + 1.5 * x3_b \tag{1}$$

$$ln\left(\frac{P(Y=1)}{P(0<Y<1)}\right) = 1 + x_1 - 0.9 * x_2 + 0.35 * x3_a + 1.5 * x3_b \tag{2}$$

$$logit(\mu) = 1.35 + x_1 - 0.9 * x_2 + 0.35 * x3_a + 1.5 * x3_b \tag{3}$$

Listing 2: Simulación de 1 base de datos y Respuestas (Método de 2 partes).

```
x.1_2 <- mvrnorm(n,mu.NM,Sigma.NM) #Variables x1,x2 ~ N2(mu.NM,Sigma.NM)

x3 = rmulti(matrix(c(p1[3],p1[1],p1[2]),
                   nrow = n,ncol = length(p1),
                   byrow = TRUE)) # Variable x3 ~ multinomial(..); Referencia sobre 0<
                      Y<1
x3 = factor(x3)
dummies = model.matrix(~x3) #x3 ---> dummies
```

```r
x3a <- dummies[,2]
x3b <- dummies[,3]

X = cbind(x.1_2[,1],x.1_2[,2],x3a,x3b) # Covariables de 1000 sujetos

# Metodo 2 partes
Y=numeric(n)

for (i in 1:n) {
        delta_0 = exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(omega))/
              (1+exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(omega))
                    + exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(eta)))
        delta_1 = exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(eta))/
              (1+exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(omega))
               + exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(eta)))

        prob = c(delta_0,delta_1,1-delta_0-delta_1)
        y = rmulti(prob)
        if(y==1){
                Y[i]=0
        }else{
                if(y==2){
                        Y[i]=1
                }else{
                        mu = 1/(1+exp(-t(as.matrix(c(1,X[i,])))%*%as.matrix(beta)))
                        Y[i]=rbeta(1,phi*mu,phi*(1-mu))
                }
        }
}

dataset <- data.frame(X,Y) # X(covariables), Y(respuesta)
```

Listing 3: Histograma normalizado.

```r
n = dim(dataset)[1]

Dat0 = dataset[dataset$Y==0,]
Dat1 = dataset[dataset$Y==1,]
Datbeta = dataset[dataset$Y>0 & dataset$Y<1,]

p_0 = (dim(Dat0)[1])/n
p_1 = (dim(Dat1)[1])/n
p_beta = 1-p_0-p_1

h1 = ggplot(Datbeta,aes(Datbeta$Y)) +
geom_histogram(breaks=seq(0,1,by=1/15),
aes(y=..density..*0.395),col="red",fill="yellow") +
labs(title="Beta inflacionada 0 y 1",
     x="Respuesta",
     y="Proporcion") +
geom_segment(aes(x=0, y=0, xend=0, yend=p_0),size=2)
    +
geom_segment(aes(x=1, y=0, xend=1, yend=p_1),size=2)

h1
```

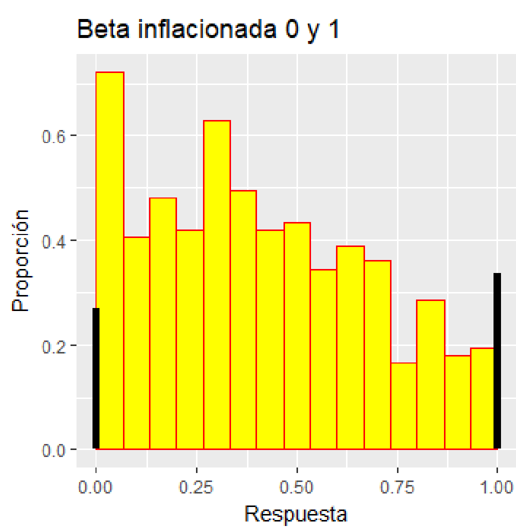

Figura 1: Histograma normalizado.

b)

Se tiene un modelo de regresión a la media con las siguientes ecuaciones

$$ln(\alpha_0) = 0.7 + x_1 - 0.9 * x_2 + 0.35 * x3_a + 1.5 * x3_b \tag{4}$$

$$ln(\alpha_1) = 1 + x_1 - 0.9 * x_2 + 0.35 * x3_a + 1.5 * x3_b \tag{5}$$

$$logit(\gamma) = 1.35 + x_1 - 0.9 * x_2 + 0.35 * x3_a + 1.5 * x3_b \tag{6}$$

Listing 4: Simulación de 1 base de datos y Respuestas (Método de regresion a la media).

```
# Se tomaran las mismas covariables X determinadas en a)

set.seed(1993)
Y=numeric(n)

for (i in 1:n) {
        alpha_0 = 1/(1+exp(-t(as.matrix(c(1,X[i,])))%*%as.matrix(omega)))
        alpha_1 = 1/(1+exp(-t(as.matrix(c(1,X[i,])))%*%as.matrix(eta)))
        gamma = 1/(1+exp(-t(as.matrix(c(1,X[i,])))%*%as.matrix(beta)))

        prob = c(alpha_0*(1-gamma),alpha_1*gamma,1-alpha_0*(1-gamma)-alpha_1*gamma)
        y = rmulti(prob)
        if(y==1){
                Y[i]=0
        }else{
                if(y==2){
                        Y[i]=1
                }else{
                        mu = gamma*(1-alpha_1)/(1-alpha_0*(1-gamma)-alpha_1*gamma)
                        Y[i]=rbeta(1,phi*mu,phi*(1-mu))
                }
        }
}

dataset2 <- data.frame(X,Y)
```

Listing 5: Histograma normalizado.

```
n = dim(dataset2)[1]

Dat0 = dataset2[dataset2$Y==0,]
Dat1 = dataset2[dataset2$Y==1,]
Datbeta = dataset2[dataset2$Y>0 & dataset2$Y< 1,]

p_0 = (dim(Dat0)[1])/n
p_1 = (dim(Dat1)[1])/n
p_beta = (dim(Datbeta)[1])/n


h2 = ggplot(Datbeta,aes(Datbeta$Y)) +
```

```
        geom_histogram(breaks=seq(0,1,by=1/15),
        aes(y=..density..*0.538),col="red",fill="yellow") +
        labs(title="Beta inflacionada 0 y 1",
        x="Respuesta",y="Proporcion")+
        geom_segment(aes(x=0, y=0, xend=0, yend=p_0),size=2)+
        geom_segment(aes(x=1, y=0, xend=1, yend=p_1),size=2)

h2
```

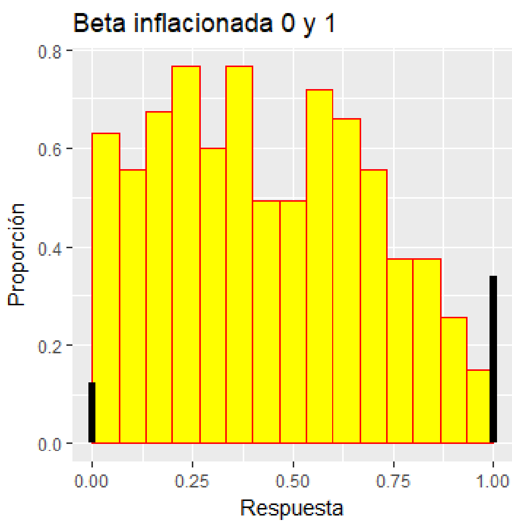

Figura 2: Histograma normalizado.

c)

Si un sujeto tuviese valores

$$x_{new} = (x1 = 0.8, x2 = 1.75, x3 = 2) \quad \equiv \quad x_{new} = (x1 = 0.8, x2 = 1.75, x3_a = 0, x3_b = 1)$$

Se calculará la respuesta esperada en ambos casos, sabiendo que

$$\begin{aligned} E[Y] &= P(Y=0)*(0) + P(Y=1)*(1) + (1 - P(Y=0) - P(Y=1))E[Y \mid 0 < Y < 1] \quad (7) \\ &= \delta_1 + (1 - \delta_0 - \delta_1)\mu \end{aligned}$$

Listing 6: Respuesta Esperada.

```
X_new=c(0.8,1.75,0,1)

####### Respuesta media (modelo dos partes)
delta_0 = exp(t(as.matrix(c(1,X_new)))%*%as.matrix(omega))/
          (1+exp(t(as.matrix(c(1,X_new)))%*%as.matrix(omega))
          +exp(t(as.matrix(c(1,X_new)))%*%as.matrix(eta)))

delta_1 = exp(t(as.matrix(c(1,X_new)))%*%as.matrix(eta))/
          (1+exp(t(as.matrix(c(1,X_new)))%*%as.matrix(omega))
          +exp(t(as.matrix(c(1,X_new)))%*%as.matrix(eta)))

mu = 1/(1+exp(-t(as.matrix(c(1,X_new)))%*%as.matrix(beta)))
```

```
Y_new_modelo1 = delta_1+(1-delta_0-delta_1)*mu
Y_new_modelo1

####### Respuesta media (modelo regresion a la media)

Y_new_modelo2 = 1/(1+exp(-t(as.matrix(c(1,X_new)))%*%as.matrix(beta)))
Y_new_modelo2
```

	Modelo de Regresión 2 partes	Modelo de Regresión a la media
$E(Y \mid X_{new})$	0.6035972	0.8884495

Si x2 incrementase en 0.5

$$x_{new.2} = (x1 = 0.8, x2 = 1.75 + 0.5, x3 = 2) \quad \equiv \quad x_{new.2} = (x1 = 0.8, x2 = 2.25, x3_a = 0, x3_b = 1)$$

Listing 7: Respuesta Esperada (Incremento de 0.5 en x2).

```
# x2 aumentado en 0.5:
X_new[2]=X_new[2]+0.5

####### Respuesta media (modelo dos partes)
delta_0 = exp(t(as.matrix(c(1,X_new)))%*%as.matrix(omega))/
          (1+exp(t(as.matrix(c(1,X_new)))%*%as.matrix(omega))
          +exp(t(as.matrix(c(1,X_new)))%*%as.matrix(eta)))

delta_1 = exp(t(as.matrix(c(1,X_new)))%*%as.matrix(eta))/
          (1+exp(t(as.matrix(c(1,X_new)))%*%as.matrix(omega))
          +exp(t(as.matrix(c(1,X_new)))%*%as.matrix(eta)))

mu = 1/(1+exp(-t(as.matrix(c(1,X_new)))%*%as.matrix(beta)))

Y_new_modelo1_incremento = delta_1+(1-delta_0-delta_1)*mu
Y_new_modelo1_incremento

####### Respuesta media (modelo regresion a la media)

Y_new_modelo2_incremento = 1/(1+exp(-t(as.matrix(c(1,X_new)))%*%as.matrix(beta)))
Y_new_modelo2_incremento

## En cuanto se ve afectada la respuesta media

Y_dif_modelo1 = Y_new_modelo1_incremento-Y_new_modelo1; Y_dif_modelo1
Y_dif_modelo2 = Y_new_modelo2_incremento-Y_new_modelo2; Y_dif_modelo2
```

	Modelo de Regresión 2 partes	Modelo de Regresión a la media
$E(Y \mid X_{new.2})$	0.6105484	0.8354835
$E(Y \mid X_{new.2}) - E(Y \mid X_{new})$	0.006951235	-0.05296592

Interpretaciones:

- En el modelo de dos partes, la respuesta media aumenta en 0.007
- En el modelo de regresión a la media, la respuesta media disminuye en 0.053

d)

Se crearán 500 bases de datos de 1000 sujetos, siguiendo el proceso en a)

Listing 8: 500 bases de datos de 1000 sujetos.

```
base_de_datos=list()

for (j in 1:500){
        x.1_2 <- mvrnorm(n,mu.NM,Sigma.NM) #Variables x1,x2 ~ N2(mu.NM,Sigma.NM)

        x3 = rmulti(matrix(c(p1[3],p1[1],p1[2]),
                nrow = n,ncol = length(p1),
                byrow = TRUE)) # Variable x3 ~ multinomial(..); Referencia sobre 0<Y<1
        x3 = factor(x3)
        dummies = model.matrix(~x3)
        x3a <- dummies[,2]
        x3b <- dummies[,3]

        X = cbind(x.1_2[,1],x.1_2[,2],x3a,x3b) # Covariables de 1000 sujetos

        # Metodo 2 partes
        Y=numeric(n)

        for (i in 1:n) {
                delta_0 = exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(omega))/
                          (1+exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(omega))
                          +exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(eta)))
                delta_1 = exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(eta))/
                  (1+exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(omega))
                  +exp(t(as.matrix(c(1,X[i,])))%*%as.matrix(eta)))

                prob = c(delta_0,delta_1,1-delta_0-delta_1)
                y = rmulti(prob)
                if(y==1){
                        Y[i]=0
                }else{
                        if(y==2){
                                Y[i]=1
                        }else{
                                mu = 1/(1+exp(-t(as.matrix(c(1,X[i,])))%*%as.matrix(
                                    beta)))
                                Y[i]=rbeta(1,phi*mu,phi*(1-mu))
                        }
                }
}

base_de_datos[[j]] <- data.frame(X,Y)
}
```

Se plantea un modelo de la siguiente forma

$$ln\left(\frac{P(Y=0)}{P(0<Y<1)}\right) = \omega_0 + \theta_1 * x_1 + \theta_2 * x_2 + \theta_3 * x3_a + \theta_4 * x3_b \tag{8}$$

$$ln\left(\frac{P(Y=1)}{P(0<Y<1)}\right) = \eta_0 + \theta_1 * x_1 + \theta_2 * x_2 + \theta_3 * x3_a + \theta_4 * x3_b \tag{9}$$

$$logit(\mu) = \beta_0 + \theta_1 * x_1 + \theta_2 * x_2 + \theta_3 * x3_a + \theta_4 * x3_b \tag{10}$$

Teniéndose esto, por el método de 2 partes estimamos en la primera parte $\theta_1, \theta_2, \theta_3, \theta_4$, y en la segunda β_0. Se reportará la media de las estimaciones de las 500 bases además del ECM (Error Cuadrático Medio).

Listing 9: Estimación (Media-ECM).

```
g <- function(data,thetax){
        as.matrix(cbind(1,data))%*%as.matrix(thetax)
}

#Parametros iniciales
theta0 = rep(1,6)
beta0 = c(1)

theta.estimado = matrix(rep(0,7*500),nrow = 500,ncol = 7)

# Estimaciones
for (j in 1:500){
        data=base_de_datos[[j]]
        log.multi= function(thetax){
                theta.0 = c(thetax[5],thetax[1:4])
                theta.1 = c(thetax[6],thetax[1:4])
                -sum(sum(g(subset(data,Y==0)[,-c(5)],theta.0)),
             sum(g(subset(data,Y==1)[,-c(5)],theta.1)),
             sum(-log(1+exp(g(data[,-c(5)],theta.0))
                      + exp(g(data[,-c(5)],theta.1)))))
}

# Estimando theta1,theta2,theta3,theta4,omega0,eta0
theta.estimado[j,1:6] = optim(theta0,
                              log.multi,
                              method = "L-BFGS-B")$par

log.multi2 = function(betax){
        theta.mu = c(betax,theta.estimado[j,1:4])
        databeta = subset(data, Y>0 & Y<1)
        Xb = databeta[,-c(5)]
        Yb = databeta[,5]
        mu = 1/(1+exp(-(as.matrix(cbind(1,Xb)))%*%as.matrix(theta.mu)))
    -sum(log(dbeta(Yb,phi*mu,phi*(1-mu))))
}

# Estimando beta0
theta.estimado[j,7] = optim(beta0,
                            log.multi2,
                            method = "L-BFGS-B")$par
}


theta.mean_Modelo1 <- apply(theta.estimado,2,mean)
theta.var_Modelo1 <- apply(theta.estimado,2,var)
theta.MSE_Modelo1 <- theta.var_Modelo1+(theta.observado-theta.mean_Modelo1)^2

estimaciones1 <- rbind(theta.observado,theta.mean_Modelo1,theta.MSE_Modelo1)
colnames(estimaciones1)<-c("a1","a2","a3","a4","w0","n0","b0")
```

```
rownames(estimaciones1)<-c("Observados","Media estimada","ECM")
estimaciones1
```

	θ_1	θ_2	θ_3	θ_4	ω_0	η_0	β_0
θ	1	-0.9	0.35	1.5	0.7	1	1.35
$E(\hat{\theta})$	1.011233111	-0.912005857	0.34623519	1.51475242	0.72253083	1.02148510	1.38537905
$ECM(\hat{\theta})$	0.009043747	0.008138641	0.04367832	0.04746863	0.06541448	0.06251102	0.06472707

Como se observa las estimaciones son muy precisas, el ECM es bajo en comparación con la escala de estimación.

Obtengamos los parámetros de regresión mediante el método Papke and Wooldridge, y comparemos

Listing 10: Método Papke and Wooldridge.

```
estimacion.Papke <- matrix(rep(0,5*500),nrow = 500,ncol = 5)

beta.ini=rep(1,5)

for(j in 1:500){
        data=base_de_datos[[j]]
        log.multi3 <- function(betax){
                pred = as.matrix(cbind(1,data[,-c(5)]))%*%betax
                P = 1/(1+exp(-pred))
                -sum(data[,c(5)]*log(P) +(1-data[,c(5)])*log(1-P))
}

estimacion.Papke[j,] = optim(beta.ini,
                             log.multi3,
                             method = "L-BFGS-B")$par
}

theta.mean_Modelo.Papke <- apply(estimacion.Papke,2,mean)
theta.var_Modelo.Papke <- apply(estimacion.Papke,2,var)
theta.MSE_Modelo.Papke <- theta.var_Modelo.Papke+(beta.observado-theta.mean_Modelo.
    Papke)^2

estimaciones.Papke <- rbind(beta.observado,theta.mean_Modelo.Papke,theta.MSE_Modelo.
    Papke)
colnames(estimaciones.Papke)<-c("b0","b1","b2","b3","b4")
rownames(estimaciones.Papke)<-c("Observados","Media estimada","ECM")
estimaciones.Papke
```

	β_0	β_1	β_2	β_3	β_4
β	1.35	1	-0.9	0.35	1.5
$E(\hat{\beta})$	0.3518575	0.3497718	-0.3202265	0.17579164	0.5638970
$ECM(\hat{\beta})$	1.0213362	0.4263945	0.3396583	0.04903656	0.8943651

No pueden compararse debido a que estamos tratando con una beta inflacionada en 0 y 1, en altos porcentajes; por lo que en caso se desea usar dicho método sería conveniente hacerlo para estimar la media condicional (en lugar de la verosmilitud para una regresión beta) en el algoritmo de regresion a 2 partes.

e)

El método de la transformación a una beta de Smithson and Verkuilen, realiza el mecanismo siguiente:

$$Y^* \quad = \quad \frac{(n-1)*Y+0.5}{n} \tag{11}$$

Pudiendo transformar las respuestas $Y \in [0,1]$ en $Y^* \in (0,1)$, y con ello implementar un método de regresión beta para la estimación correspondiente de μ.

Listing 11: Método de la transformación a una beta de Smithson and Verkuilen.

```
theta.estimado.Smithson <- matrix(rep(0,5*500),nrow = 500,ncol = 5)

for (j in 1:500) {
        base_de_datos2 = base_de_datos[[j]]
        base_de_datos2$Y = ((n-1)*base_de_datos2$Y+0.5)/n
        theta.estimado.Smithson[j,] = betareg(Y~ V1+V2+x3a+x3b,data = base_de_datos2)$
            coefficients$mean
}

theta.mean_Modelo.Smithson <- apply(theta.estimado.Smithson,2,mean)
theta.var_Modelo.Smithson <- apply(theta.estimado.Smithson,2,var)
theta.MSE_Modelo1.Smithson <- theta.var_Modelo.Smithson+(beta.observado-theta.mean_
    Modelo.Smithson)^2

estimaciones.Smithson <- rbind(beta.observado,theta.mean_Modelo.Smithson,theta.MSE_
    Modelo1.Smithson)

colnames(estimaciones.Smithson)<-c("b0","b1","b2","b3","b4")
rownames(estimaciones.Smithson)<-c("Observados","Media estimada","ECM")
estimaciones.Smithson
```

	β_0	β_1	β_2	β_3	β_4
β	1.35	1	-0.9	0.35	1.5
$E(\hat{\beta})$	0.2058915	0.1240508	-0.1155636	0.06183388	0.1986495
$ECM(\hat{\beta})$	1.3196754	0.7686464	0.6167446	0.09095841	1.7010330

Como se observa, estas estimaciones no son buenas pues generan un sesgo muy grande; sin embargo, el efecto si puedo determinarse correctamente (efectos negativos y positivos). Cabe mencionar que estas estimaciones al haberse determinado con betareg, modelan la precisión. Lo ideal sería comparar este tipo de regresión manteniendo fijo $\phi = 1$.

Problema 2.

Listing 12: Lectura de datos.

```
library(readxl)
library(dplyr)
library(table)
install.packages(table)
setwd("C:/Users/saban/Documents/PUCP/2018 - 2/ESTADISTICA CONTEMPORANEA/RegresionInflacionada/Datos")
df = read_excel("SW.xlsx", sheet = 1)
df <- data.frame(df)
```

Se usan como parámetros iniciales para los modelos de regresión como sigue:

Listing 13: Parámetros iniciales.

```
#Parametros iniciales
theta0 = rep(1,10)
```

Para el modelo de **regresión de 2 partes** estimamos ω, η, para luego determinar β y ϕ.

Listing 14: Modelo de 2 partes.

```
#------------REGRESION 2 PARTES
theta.estimado = rep(0,10)

g <- function(data,thetax){
        as.matrix(cbind(1,data))%*%as.matrix(thetax)
}

log.multi = function(thetax){
       -sum(sum(g(subset(df,Barthel.index==0)[,-c(3)],thetax[1:3])),
                sum(g(subset(df,Barthel.index==1)[,-c(3)],thetax[4:6])),
                sum(-log(1+exp(g(df[,-c(3)],thetax[1:3]))
                 + exp(g(df[,-c(3)],thetax[4:6])))))
}

# Estimacion de omega,eta
theta.estimado[1:6] = optim(theta0[1:6],log.multi,method = "L-BFGS-B")$par

databeta = subset(df, Barthel.index>0 & Barthel.index<1)
log.multi2 = function(betax){
        Xb = databeta[,-c(3)]
        Yb = databeta[,3]
        mu = 1/(1+exp(-(as.matrix(cbind(1,Xb)))%*%as.matrix(betax[1:3])))
       -sum(log(dbeta(Yb,betax[4]*mu,betax[4]*(1-mu))))
}

# Estimacion de beta,phi
theta.estimado[7:10] = optim(theta[7:10],log.multi2,method = "L-BFGS-B")$par

theta.estimado <- matrix(theta.estimado,ncol=10)
colnames(theta.estimado) <- c("w0","w1","w2","n0","n1","n2","b0","b1","b2","phi")
rownames(theta.estimado) <- c("Estimacion")
theta.estimado
```

	ω_0	ω_1	ω_2	η_0	η_1	η_2
Estimación	-2.804793	0.4247591	0.1824472	1.401442	-0.213898	0.7279867

	β_0	β_1	β_2	ϕ
Estimación	2.236848	-0.264585	0.2295762	2.498583

Por tanto, tenemos que

$$\frac{P(Y=0)}{P(0<Y<1)} = \exp(-2.804793 + 0.4247591 * Edad + 0.1824472 * rt.PA1) \tag{12}$$

Interpretaciones:

- Si la Edad aumenta en 1, las chances de que la respuesta sea $Y = 0$ en lugar de $0 < Y < 1$, aumenta a $exp(0.4247591) = 1.52$ veces.
- Si se aplica el tratamiento especial, las chances de que la respuesta sea $Y = 0$ en lugar de $0 < Y < 1$, aumenta a $exp(0.1824472) = 1.2$ veces.

$$\frac{P(Y=1)}{P(0<Y<1)} = \exp(1.401442 - 0.213898 * Edad + 0.7279867 * rt.PA1) \tag{13}$$

Interpretaciones:

- Si la Edad aumenta en 1, las chances de que la respuesta sea $Y = 1$ en lugar de $0 < Y < 1$, disminuye a $exp(-0.213898) = 0.80$ veces.
- Si se aplica el tratamiento especial, las chances de que la respuesta sea $Y = 1$ en lugar de $0 < Y < 1$, aumenta a $exp(0.7279867) = 2.07$ veces.

$$\mu = \frac{1}{1 + \exp(-(2.236848 - 0.264585 * Edad + 0.2295762 * rt.PA1))} \tag{14}$$

Interpretaciones:

- Si la Edad aumenta, la media condicional μ de la variable respuesta disminuye.
- Si se aplica el tratamiento especial, la media condicional μ de la variable respuesta aumenta.
- El parámetro de precisión es de 2.498583

Para el modelo de **regresión a la media** estimamos ω, η, β y ϕ, en un solo paso final

Listing 15: Modelo Regresión a la media.

```
#-----------REGRESION A LA MEDIA
theta0 = rep(1,10)

theta.estimado2 = rep(0,10)

g.logit <- function(theta,data){
        1/(1+exp(-as.matrix(cbind(1,data[,-c(3)]))%*%theta))
```

```r
}

log.multi3 <- function(thetax){
        alpha_0.b = g.logit(thetax[1:3],subset(df, Barthel.index>0 & Barthel.index<1))
        alpha_1.b = g.logit(thetax[4:6],subset(df, Barthel.index>0 & Barthel.index<1))
        gamma.b = g.logit(thetax[7:9],subset(df, Barthel.index>0 & Barthel.index<1))

        alpha_0.y0 = g.logit(thetax[1:3],subset(df, Barthel.index==0))
        alpha_1.y0 = g.logit(thetax[4:6],subset(df, Barthel.index==0))
        gamma.b.y0 = g.logit(thetax[7:9],subset(df, Barthel.index==0))

        alpha_0.y1 = g.logit(thetax[1:3],subset(df, Barthel.index==1))
        alpha_1.y1 = g.logit(thetax[4:6],subset(df, Barthel.index==1))
        gamma.b.y1 = g.logit(thetax[7:9],subset(df, Barthel.index==1))

        c = 1 - alpha_0.b*(1-gamma.b)-alpha_1.b*gamma.b
        mu = gamma.b*(1-alpha_1.b)/c
        aux.0 = sum(log(alpha_0.y0*(1-gamma.b.y0)))
        auX.1 = sum(log(alpha_0.y1*gamma.b.y1))
        aux.beta = sum(log(c*dbeta(subset(df, Barthel.index>0 & Barthel.index<1)[,3],
                                thetax[10]*mu,thetax[10]*(1-mu))))
        -(aux.0+auX.1+aux.beta)
}

# Estimacion de omega,eta,beta,phi
theta.estimado2 = optim(theta0,log.multi3,method = "L-BFGS-B")$par

theta.estimado2 <- matrix(theta.estimado2,ncol=10)
colnames(theta.estimado2) <- c("w0","w1","w2","n0","n1","n2","b0","b1","b2","phi")
rownames(theta.estimado2) <- c("Estimacion")
theta.estimado2
```

	ω_0	ω_1	ω_2	η_0	η_1	η_2
Estimación	0.8549056	0.03334929	0.3573334	0.5287433	-1.978516	0.8506672

	β_0	β_1	β_2	ϕ
Estimación	2.531572	-0.4338705	0.4034744	2.626877

Por tanto, tenemos que

$$logit(\alpha_0) = 0.8549056 + 0.03334929 * Edad + 0.3573334 * rt.PA1 \tag{15}$$

$$logit(\alpha_1) = 0.5287433 - 1.978516 * Edad + 0.8506672 * rt.PA1 \tag{16}$$

$$\gamma = \frac{1}{1 + \exp(-(2.531572 - 0.4338705 * Edad + 0.4034744 * rt.PA1))} \tag{17}$$

Interpretaciones:

- Si la Edad aumenta, la respuesta media γ disminuye.
- Si se aplica el tratamiento especial, la respuesta media γ de la variable respuesta aumenta.