Análisis de Datos Categóricos

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

Semestre: *2017-II*

Tema: PC 4

Prof. Victor Giancarlo Sal Y Rosas

Pontificia Universidad Católica del Perú
Escuela de Posgrado
Maestría en Estadística

Problema 1: Simulación e interpretación

Se desea hacer un estudio donde se desea corroborar el efecto de una vacuna para prevenir la recaida en abuso de drogas. Variables:

- **X**: (1=recibir vacuna con el medicamento en evaluación, 0=recibir vacuna placebo).
- **Y**: Respuesta (1=recae en drogas, 0=no recae).

Despues de aleatorizar a la persona se le evalua cada mes por seis meses. Se asume el siguiente modelo:

$$log\left[\frac{P(Y_{ij=1} \mid X_{ij})}{P(Y_{ij=0} \mid X_{ij})}\right] = \beta_0 + \beta_1 X_{ij} + b_i \tag{1}$$

donde $b_i \sim N(0, 9)$ y $(\beta_0, \beta_1) = (-1, -0.69)$

a) Se simulará 1000 bases de datos que contengan 200 participantes y cuya data es modelada por

Listing 1: Base de datos: Lista de 1000 bases de 200 participantes evaluada cada 6 meses.

```r
library(lme4)
library(geeM)

# DATOS INICIALES
p=0.5
n=c(1000,200,6)
beta0.media=0
beta0.s=3
beta=c(-1,-0.69)

# DEFINICION DE LA BASE DE DATOS, CLASE: LIST
basededatos=list()

#----SIMULACION-----------------------------------------------------------------
for(i in 1:n[1]){
  X = rep(rbinom(n[2],1,p),n[3])
  basededatos[[i]] = data.frame(X,Y=rbinom(n[3]*n[2], size=1,
                                prob=1/(1+exp(-(beta[1]+beta[2]*X
                                              +rep(rnorm(n[2],beta0.media,
                                                         beta0.s),n[3])))))),
                                participante=rep(1:n[2],n[3]))
}
```

Listing 2: Base de datos k (Output)

```r
# BASE DE DATOS k
k=5
head(basededatos[[k]])
```

```
  X Y participante
1 1 1           1
2 0 0           2
3 1 0           3
4 0 0           4
5 1 1           5
6 1 0           6
```

b) Se estimará β_1 mediante la simulación realizada en item a), usando (i) modelo de regresión logística (ignorando las medidas repetidas), (ii) modelo logístico con efectos mixtos, (iii) El modelo de ecuaciones generalizadas de estimación.

Listing 3: Estimación de β_1

```
beta1.estimado <- function(basededatos){
  beta1.f.L=numeric()
  beta1.f.LM=numeric()
  beta1.f.EG=numeric()
  beta1.f.EG2=numeric()

  for (i in 1:n[1]) {
    #----MODELO LOGISTICO SIMPLE-----------------------------------------------
    beta1.f.L[i] = glm(Y ~ X, data=basededatos[[i]] ,family=binomial)$coef["X"]

    #----MODELO LOGISTICO CON EFECTOS MIXTOS-----------------------------------
    beta1.f.LM[i] = unique(coef(glmer(Y ~ X + (1|participante),
                                      data=basededatos[[i]],family=binomial))
                           $participante$X)

    #----MODELO DE ECUACIONES GENERALIZADAS------------------------------------

    # INDEPENDENCE
    beta1.f.EG[i] = coef(geem(Y ~ X, id = participante,data=basededatos[[i]],
                              family=binomial(link=logit),
                              corstr="independence"))[2]

    # EXCHANGEABLE
    beta1.f.EG2[i] = coef(geem(Y ~ X, id = participante,data=basededatos[[i]],
                               family=binomial(link=logit),
                               corstr="exchangeable"))[2]
        }

    return(cbind(beta1.f.L,beta1.f.LM,beta1.f.EG,beta1.f.EG2))
}

# ESTIMACION DE BETA1 POR CADA BASE
b1est=beta1.estimado(basededatos)
head(b1est,3)
```

	Logístico Simple	Efectos Mixtos	Ecuaciones Generalizadas (Independence)	Ecuaciones Generalizadas (Exchangeable)
basededatos[[1]]	-0.2908022	-0.6058613	-0.2908022	-0.2908022
basededatos[[2]]	-0.3236040	-0.6054216	-0.3236040	-0.3236040
basededatos[[3]]	-0.5419193	-1.0649212	-0.5419193	-0.5419193

c) Compare los promedios de sus estimaciones y discuta los resultados que obtuvo en c)

Listing 4: Promedio de estimaciones β_1

```
# PROMEDIO DE LAS ESTIMACIONES BETA1 OBTENIDAS
prom.beta1.estimado=apply(b1est,2,mean)
prom.beta1.estimado
```

Modelo	Logístico Simple	Efectos Mixtos	Ecuaciones Generalizadas (Independence)	Ecuaciones Generalizadas (Exchangeable)
β_1	-0.325217	-0.7215916	-0.325217	-0.325217

- El β_1 promedio estimado por el modelo GEE, asumiendo independencia (independence) entre observaciones por cada individuo, es el mismo que el obtenido por el modelo GLM; esto es claro debido a que la técnica bajo GEE de este tipo asume nula las correlaciones entre distintos puntos temporales y por tanto puede verse como observaciones de individuos distintos, siendo la misma técnica aplicada por GLM.
- El β_1 promedio estimado por el modelo GEE, asumiendo correlaciones (exchangeable) entre observaciones por cada individuo, parece se las misma que la obtenida por el modelo GEE asumiendo independencia, esto se debe a que las correlaciones estimadas entre distintos puntos temporales son casi nulas, dejando sin efecto el método de este tipo.
- Con respecto al β_1 promedio estimado bajo el modelo GLMER, difere de los obtenidos por los demás modelos, además siendo el de menor sesgo. Esto se debe a que los demás modelos, en particular por GEE (exchangeable), nos colocan en un marco de independencia entre puntos temporales, y dado que el modelo GLMER, acepta tal independencia y asume lo propuesto en (1), por la adición de efectos aleatorios por cada individuo, nos estima un β_1 con menor sesgo.
- Todos los modelos, podrían haber estimado un β_1 promedio semejante con bajo sesgo, si es que la simulación hubiese optado por efectos aleatorios entre individuos de varianza muy pequeña, y de este modo los modelos propuestos por cada técnica resultarían en teoría semejantes.

Problema 2: Aplicación

Los siguientes datos corresponde a un estudio de esquizofrenia y se tienen disponibles cuatro variables:

- id: Identificador de la persona
- y: Indicador de síntomas
- month: meses desde hospitalización
- age: 0 (<20 años) y 1 ($\geq$ 20)
- sex: 0 (hombre) y 1 (mujer)

Listing 5: Base de datos.

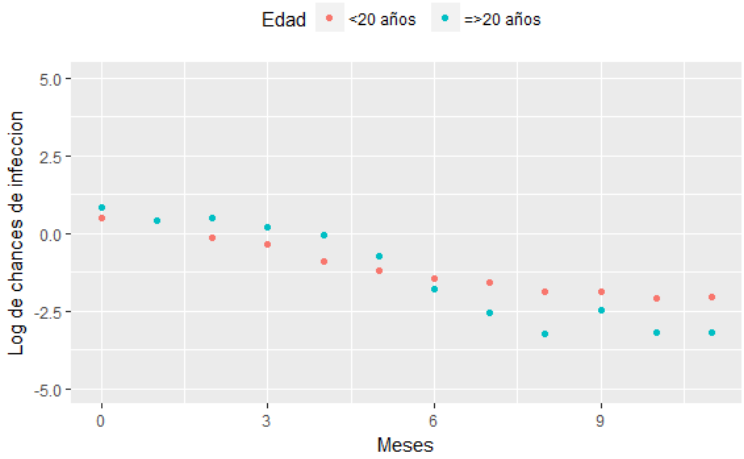
```
libraries("dplyr","ggplot2","foreign","lme4","nlme","bestglm","geeM","vcd","MuMIn",
          "sjPlot")
datos <-read.table("http://faculty.washington.edu/heagerty/Books/AnalysisLongitudinal/
                    madras.data")
datos <- datos[,1:5]
names(datos) <- c("id","y","month","age","sex")
```

Deseamos responder si existe evidencia que la edad o el sexo del paciente están asociados con su evolución y si esta relación varía en el tiempo.

Primero evaluaremos las posibles interacciones que se puedan implementar

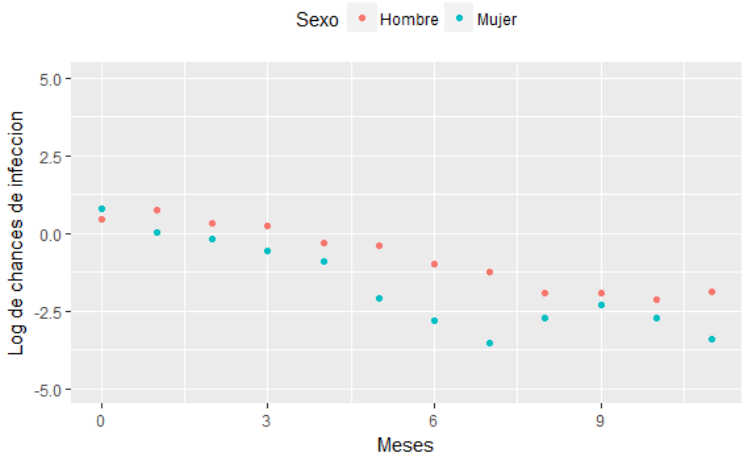
Listing 6: Primer Análisis.

```
summary.datos<-datos %>% group_by(age,month) %>%
               summarize(mean.y=mean(y,na.rm=T), logit.y=log(mean.y/(1-mean.y)))

ggplot(summary.datos ,aes(x=month ,y=logit.y,group=as.factor(age),
                          color=as.factor(age))) + geom_point() +
                          labs(x = "Meses",y="Log de chances de infeccion",size=15) +
                          coord_cartesian(ylim = c(-5, 5)) +
                          theme(legend.position = "top") +
                          scale_colour_discrete(breaks=c("0","1"),
                          labels=c("<20 anos","=>20 anos"), name="Edad")
```



Listing 7: Primer Análisis.

```
summary.datos<-datos %>% group_by(sex,month) %>%
               summarize(mean.y=mean(y,na.rm=T), logit.y=log(mean.y/(1-mean.y)))

ggplot(summary.datos ,aes(x=month ,y=logit.y,group=as.factor(sex),
                          color=as.factor(sex))) + geom_point() +
                          labs(x = "Meses",y="Log de chances de infeccion",size=15) +
                          coord_cartesian(ylim = c(-5, 5)) +
                          theme(legend.position = "top") +
                          scale_colour_discrete(breaks=c("0","1"),
                          labels=c("<20 anos","=>20 anos"), name="Sexo")
```



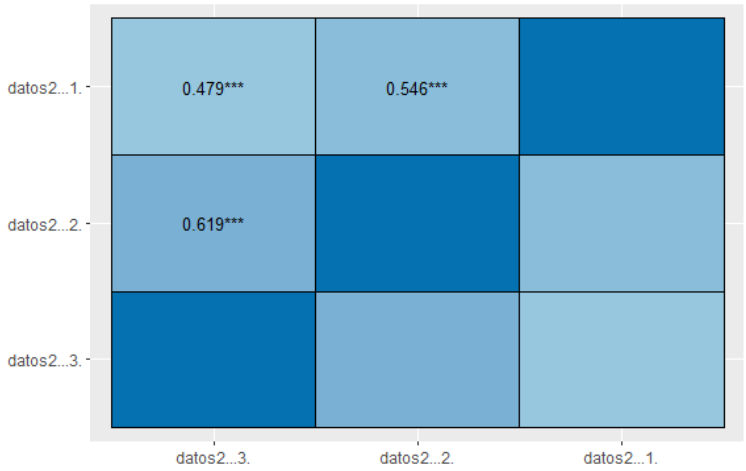
Observamos que ambas gráficas siguen una tendencia lineal casi semejante; sin embargo, se optará por adicionar modelar junto a la interacción respectiva entre edad y meses a causa de que pareciera mejorar considerablemente desde cierto tiempo, el grupo de edades mayores o iguales de 20 respecto al de menores de 20.

a) Use ecuaciones de estimación para responder la pregunta de interés. Justifique su elección de estructura de correlación.

Listing 8: Selección de Estructura de Correlación

```
d0=subset(datos, month==0)[,1:2]
d1=subset(datos, month==1)[,1:2]
d2=subset(datos, month==2)[,1:2]

v=rbind(d0,d1,d2)
v1=rep(1,254)
v=cbind(v,v1)

summary.1<-v %>% group_by(id) %>% summarize(sum.1=sum(v1))

summary.1=data.frame(summary.1)

dat2=subset(datos, id!= 82 & id!=79 & id!= 17)
datos2 = cbind(subset(dat2, month==0)[,2],subset(dat2, month==1)[,2],
subset(dat2, month==2)[,2])

DF1 <- data.frame(datos2[,1],datos2[,2],datos2[,3])
DF1[] <- lapply(DF1,as.integer)
sjp.corr(DF1)
```

```
20 sjt.corr(DF1)
```



Se analizaremos las correlaciones existentes entre observaciones entre las tres primeras medidas temporales, se observa que las correlaciones existen son relativamente fuertes, descartando un posible modelo de ecuaciones generalizadas con el supuesto de una estructura de independencia entre observaciones. Se optará por una estructura intercambiable (exchangeable) debido a que no se observa una relación cuadrática en las correlaciones a medida que avance el tiempo. No se optará por un modelo sin estructura debido a que por la cantidad de medidas observadas por individuo, una estructura de este tipo resulta muy compleja y de mayor costo computacional, además de inadecuada sabiendo que a priori una estructura intercambiable es suficiente, según el análisis.

Listing 9: Modelo 1 GEE (Estructura Intercambiable)

```
1 mod1 = geem(y ~ sex + age*month, id=id, data=datos,family=binomial(link="logit"),
2             corstr="exchangeable" )
3 summary(mod1)
```

```
            Estimates Model SE Robust SE   wald        p
(Intercept)    0.7399  0.22700   0.29750  2.487 1.287e-02
sex           -0.7529  0.29720   0.35660 -2.111 3.474e-02
age            1.0240  0.33260   0.50220  2.039 4.145e-02
month         -0.2985  0.03173   0.05775 -5.170 2.300e-07
age:month     -0.1241  0.05538   0.08966 -1.384 1.664e-01

 Estimated Correlation Parameter:  0.2912
 Correlation Structure:  exchangeable
 Est. Scale Parameter:  1.027

 Number of GEE iterations: 5
 Number of Clusters:  86    Maximum Cluster Size:  12
 Number of observations with nonzero weight:  922
```

Las variables edad y sexo presentan efectos significativos si de evaluar el cociente de odds es requerido, a diferencia del efecto de la interaccion entre edad y meses dado que es significativa.

Optaremos por el modelo siguiente.

Listing 10: Modelo 2 GEE (Estructura Intercambiable)

```
mod2 = geem(y ~ sex + age+month, id=id, data=datos,family=binomial(link="logit"),
corstr="exchangeable" )
summary(mod2)
```

```
             Estimates Model SE Robust SE   wald       p
(Intercept)     0.7657  0.23660    0.2999  2.553 0.01069
sex            -0.7572  0.31510    0.3769 -2.009 0.04454
age             0.7550  0.32950    0.3862  1.955 0.05057
month          -0.3460  0.02849    0.0475 -7.285 0.00000

 Estimated Correlation Parameter:  0.2971
 Correlation Structure:  exchangeable
 Est. Scale Parameter:  1.155

 Number of GEE iterations: 6
 Number of Clusters:  86    Maximum Cluster Size:  12
 Number of observations with nonzero weight:  922
```

- El odds de infección de un hombre es $e^{-0.7572} - 0.47$ veces el odds de infección de una mujer.
- Si la edad aumenta en uno, el odds de infección aumenta en $e^{0.7550} = 2.12$ veces.
- Si transcurre un mes, el odds de infección aumenta en $e^{-0.3460} = 0.71$ veces.

b) Use modelos con efectos aleatorios para responder la pregunta de interés. Justifique su elección de efectos aleatorios.

Debido a que el modelo anterior no nos establece una respuesta adecuada con respecto a la interacción ocurrida entre edad y tiempo, de la cual intuimos si debería existir, se optará por este modelo de efectos mixtos para observar si efectivamente esto se debe a los efectos aleatorios de las personas en estudio.

Listing 11: Modelo GLMR

```
mod3 = glmer ( y ~ sex + age*month + (1| id ), data =datos , family = binomial (
    link ="logit"))
summary(mod3)
```

```
Random effects:
 Groups Name        Variance Std.Dev.
 id     (Intercept) 4.781    2.187
Number of obs: 922, groups:  id, 86

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.19842    0.43973   2.725  0.00642 **
sex         -1.22821    0.54969  -2.234  0.02546 *
age          1.50338    0.67107   2.240  0.02507 *
month       -0.46153    0.04896  -9.426  < 2e-16 ***
age:month   -0.27153    0.09656  -2.812  0.00492 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Correlation of Fixed Effects:
          (Intr) sex    age    month
sex       -0.525
age       -0.388 -0.168
month     -0.424  0.100  0.199
age:month  0.187  0.012 -0.530 -0.414
```

Como se observa, el efecto de la interacción es estadísticamente significativa, asumiendo efectos aleatorios por cada individuo. Podemos decir entonces que dado un individuo aleatoriamente, el odds de infeccion si su edad fuese una unidad mayor en el mes 1, varía en $e^{1.50338-0.27153} = 3.42$. En el mes 6, el odds de infección varía $e^{1.50338-0.27153*6} = 0.88179$.

Pregunta 3 (Aplicación)

La siguiente base de datos describe un estudio cross over para evaluar deficiencia cerebrovascular en los cuales una droga activa A y un placebo B fueron comparados. Treinta y cuatro personas recibieron la droga A (periodo 1) y luego después de un periodo de descanso recibieron la droga B (periodo 2) . Otras 33 personas hicieron el proceso al revés (primero B y luego A).

La respuesta esta definida como 1 si la lectura del electrocardiograma fue normal y 0 en caso contrario.

Listing 12: Base de datos k (Output)

```
library(dplyr)
library(ggplot2)
library(foreign)
library(lme4)
library(nlme)
library(bestglm)
library(geeM)
library(vcd)

datos <- read.table("https://faculty.washington.edu/heagerty
                    /Books/AnalysisLongitudinal/xover1.data")
datos <- datos[,c(1,3,5:6)]
names(datos) <- c("id","y","trt","periodo")

head(datos)
```

```
  id y trt periodo
1  1 1   0       0
2  1 1   1       1
3  2 1   0       0
4  2 1   1       1
5  3 1   0       0
6  3 1   1       1
```

donde

- id: Identificador de la persona
- y: 1 (si la persona esta afectada) y 0 (en caso contrario).
- trt: 1 (si recibio el tratamiento B) y 0 (si recibio el tratamiento A).
- periodo: 1 (si es el segundo periodo) y 0 (si fue el primero).

Deseamos responder si existe evidencia que la droga A es mejor que la droga B.

a) Calcule el cociente de odds asociado con tratamiento y buena respuesta asumiendo independencia entre las observaciones.

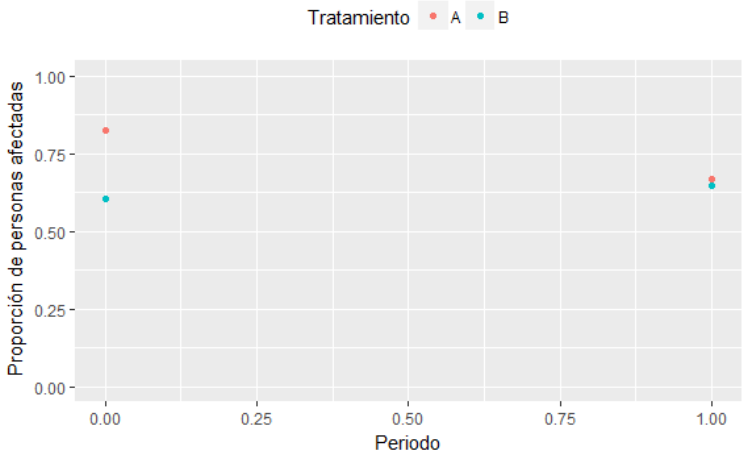
Modelo logístico simple:

Asumiendo independencia entre observaciones de un mismo individuo, el análisis principal tomará todas las medidas como hechas por distintos individuos. Para ello primero visualicemos como varía la proporción de lectura normal en distintos periodos respecto a cada tratamiento.

Listing 13: Primer análisis.

```
summary.datos <- datos  %>%
group_by(trt,periodo)   %>%
summarize(mean.y = mean(y,na.rm=T))

```

```
ggplot(summary.datos, aes(x=periodo,y=mean.y,group=as.factor(trt),
                          color=as.factor(trt))) +
       geom_point() +
       labs(x = "Periodo",y=" Proporcion de personas afectadas",size=18) +
       coord_cartesian(ylim = c(0, 2)) + theme(legend.position = "top") +
       scale_colour_discrete(breaks=c("0","1"),
       labels=c("A","B"),name="Tratamiento")
```



Como podemos observar, parece existir una interacción entre tratamiento y periodo, debido a las distintas pendientes visualizadas en relacion a la proporción de respuesta normal y el periodo. Analicemos estadísticamente la significancia de esta interacción.

Listing 14: Modelo con interacción.

```
model3 = glm(y ~ trt*periodo,data=datos,family=binomial(link="logit"))
```

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.5404     0.4499   3.424 0.000617 ***
trt          -1.1097     0.5739  -1.934 0.053148 .
periodo      -0.8473     0.5820  -1.456 0.145450
trt:periodo   1.0227     0.7710   1.326 0.184714
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Obtenemos p-valores para los efectos del periodo e interacción, estadísticamente no significativos. Es claro que el p-valor para el efecto del tratamiento parece ser significativo pero puede deberse a la influencia obtenida por las demás variables, por lo que optaremos por generar un modelo solo con la variable tratamiento.

Listing 15: Modelo con interacción.

```
model1 = glm(y ~ trt,data=datos,family=binomial(link="logit"))
```

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.0788     0.2808   3.843 0.000122 ***
trt          -0.5600     0.3777  -1.483 0.138120
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

En efecto, obtenemos un p-valor sobre el efecto de tratamiento estadísticamente no significativo. Sin embargo, ahora observemos el ajuste proporcional de ambos modelos tratados anteriormente.

Listing 16: Segundo análisis.

```
datos$pred1 = predict.glm(model1,new.data=datos,type="response")
datos$pred3 = predict.glm(model3,new.data=datos,type="response")

summary.datos2 <- datos  %>%
group_by(trt,periodo)   %>%
summarize(oprob = mean(y,na.rm=T),
prpob1 = mean(pred1,na.rm=T),
prpob3 = mean(pred3,na.rm=T))

summary.datos2
```

	trt <int>	periodo <int>	oprob <dbl>	prpob1 <dbl>	prpob3 <dbl>
1	0	0	0.8235294	0.7462687	0.8235294
2	0	1	0.6666667	0.7462687	0.6666667
3	1	0	0.6060606	0.6268657	0.6060606
4	1	1	0.6470588	0.6268657	0.6470588

donde *oprob* representa las proporciones reales en la muestra, *prpob*1 la predicha por el modelo solo con la variable trt y *prpob*1 la predicha usando todas las variables además de la interacción. Como observamos hay un ajuste perfecto por el modelo que presentaba efectos estadísticamente significativos.
Esto nos lleva a que la interpretación del cociente de odds realizado bajo el modelo con interacción es adecuado matemáticamente en la data tratada debido a las pocas medidas realizadas, sin embargo estadísticamente no es un ajuste óptimo en cuanto a análisis.

Modelo de Ecuaciones generalizadas (Independencia):
En este caso asumimos independencia entre individuos pero ajustando la necesidad de cluster por cada individuo, esto nos lleva a analizar un nuevo modelo

Listing 17: Segundo análisis.

```
mod.GEEM = geem(y ~ trt,id= id, data=datos,family=binomial(link="logit"),
                corstr = "independence")
summary(mod.GEEM)
```

```
             Estimates Model SE Robust SE   wald         p
(Intercept)      1.079   0.2829    0.2808  3.843 0.0001218
trt             -0.560   0.3805    0.2357 -2.376 0.0175000

 Estimated Correlation Parameter:  0
 Correlation Structure:  independence
 Est. Scale Parameter:  1.015

 Number of GEE iterations: 2
 Number of Clusters:  67    Maximum Cluster Size:  2
 Number of observations with nonzero weight:  134
```

Por estudios previos, encontramos este modelo con mejor ajuste y un efecto de tratamiento con alta significancia. Observamos que los efectos determinados son los mismos que los determinados por el efecto logístico simple, sin embargo el supuesto de cluster por individuo en este modelo nos engloba en un resultado estadístico aceptable.

INTERPRETACIÓN:

- El cociente de odds de buena respuesta del tratamiento B sobre el tratamiento A, es $e^{-0.560} = 0.57$. Esto quiere decir, que de hacerse uso del tratamiento B en lugar de A, el odds de buena respuesta disminuye en 43 %.

b) Existe evidencia de un efecto arrastre. Es decir que el hecho que una misma persona reciba ambos tratamientos (uno despues del otro) podria afectar los resultados?

Se llevará a cabo un análisis cross over de los efectos del tratamiento, para ello se implementará una variable a cada individuo que nos indicará si efectivamente el efecto llevado a cabo por el cross over es realmente significativo o no como para tomarlo en cuenta. Sea la variable **secuencia** definida de la siguiente forma:

- Seq: 0 (Si el resultado no proviene de una secuencia de tratamientos), 1 (si proviene de una secuencia de A hacia B) y 2 (si proviene de una secuencia de B hacia A).

El modelo tomaría una forma:

$$\log\left(\frac{\pi}{1-\pi}\right) \quad = \quad \beta_0 + \beta_1 * \text{trt} + \beta_2 * \text{periodo} + \beta_3 * \text{seq} \tag{2}$$

Listing 18: Adición variable secuencia

```
seq=rep(0,dim(datos)[1])

for (i in 1:dim(datos)[1]) {
  if(datos$periodo[i] == 1){
    seq[i] = datos$trt[i] + 1
  }
}

datos=data.frame(datos,seq)
head(datos)
```

```
  id y trt periodo seq
1  1 1   0       0   0
2  1 1   1       1   2
3  2 1   0       0   0
4  2 1   1       1   2
5  3 1   0       0   0
6  3 1   1       1   2
```

Analicemos el modelo, asumiendo independencia entre variables, pero con adición de la variable secuencia, por cuestión de observar si efectivamente se presentaría un efecto arrastre.

Listing 19: Adición variable secuencia

```
mod.GEEM.seq = geem(y ~ trt + seq,id= id, data=datos,
                    family=binomial(link="logit"),corstr = "independence")
summary(mod.GEEM.seq)
```

```
             Estimates Model SE Robust SE    wald        p
(Intercept)   1.11200   0.3069    0.2918  3.8110 0.0001386
trt          -0.52490   0.4014    0.2535 -2.0710 0.0383900
seq          -0.06659   0.2331    0.1695 -0.3929 0.6944000

 Estimated Correlation Parameter:  0
 Correlation Structure:  independence
 Est. Scale Parameter:  1.022
```

Como observamos, el efecto arrastre es estadísticamente no significativo, asumiento independencia entre observaciones. Además podemos acotar que si tomamos en cuenta este efecto,

la influencia sobre el efecto de tratamiento es casi nula pues sigue relativamente siendo el mismo. Podemos concluir que no existe evidencia de un efecto arrastre siempre y cuando se asuma independencia entre observaciones mediante un modelo de ecuaciones generalizadas.

c) Use ecuaciones de estimación para responder la pregunta de interés. Justifique su elección de estructura de correlación.

En este problema nos enfocamos en realizar un modelo asumiento una estructura de correlación adecuada. Además de hacer uso de la variable secuencia con el objetivo de analizar si el efecto arrastre es significativo.

Listing 20: Selección de Estructura de Correlación

```
library(sjPlot)
aux1=numeric()
aux2=numeric()

for (i in 1:67) {
  aux1[i] = datos$y[i*2-1]
  aux2[i] = datos$y[i*2]
}

DF <- data.frame(aux1,aux2)
DF[] <- lapply(DF,as.integer)

sjp.corr(DF)
sjt.corr(DF)
```

	aux1	*aux2*
aux1		0.591***
aux2	0.591***	

Computed correlation used pearson-method with listwise-deletion.

Según los datos logra observarse una correlación alta entre las observaciones medidas en los distintos tiempo, por lo que una estructura de independencia no sería adecuada. Se optará por la estructura intercambiable (exchangeable) a causa de que estima la correlación adecuada. Las demás estructuras son semejantes debido a que la matriz de correlaciones es 2×2.

Los modelos a analizar serán el siguiente

Listing 21: Modelo 1 GEE

```
mod.GEEM2 = geem(y ~ trt,id= id, data=datos,family=binomial(link="logit"),
corstr = "exchangeable")

summary(mod.GEEM2)
```

```
            Estimates Model SE Robust SE   wald        p
(Intercept)     1.079   0.2829    0.2808  3.843 0.0001218
trt            -0.560   0.2346    0.2357 -2.376 0.0175000

 Estimated Correlation Parameter:  0.6234
 Correlation Structure:  exchangeable
 Est. Scale Parameter:  1.015
```

Listing 22: Modelo 2 GEE

```
mod.GEEM2.seq = geem(y ~ trt+seq,id= id, data=datos,family=binomial(link="logit")
    ,
corstr = "exchangeable")

summary(mod.GEEM2.seq)
```

```
            Estimates Model SE Robust SE   wald        p
(Intercept)    1.1650   0.2971    0.2855  4.080 4.505e-05
trt           -0.4813   0.2440    0.2548 -1.889 5.894e-02
seq           -0.1650   0.1480    0.1498 -1.101 2.707e-01

 Estimated Correlation Parameter:  0.6399
 Correlation Structure:  exchangeable
 Est. Scale Parameter:  1.026
```

Como observamos el efecto de arrastre no es significativo cuando se implementa en el modelo por lo que podemos decir que no existe aun evidencia de que tal efecto ocurra.

Con respecto al efecto de tratamiento si es significativo, interpretándose de la siguiente forma: El cociente de odds de buena respuesta del tratamiento B sobre el tratamiento A, es $e^{-0.560} = 0.57$. Esto quiere decir, que de hacerse uso del tratamiento B en lugar de A, el odds de buena respuesta disminuye en 43 %.

d) Use modelos con efectos aleatorios para responder la pregunta de interés. Justifique su elección de efectos aleatorios.

e) ¿Cuál de las dos técnicas estudiadas en el curso, responden mejor la pregunta de interés?. Justifique.