

**MICROECONOMETRÍA****Estudiantes: David Gerardo Mora Salazar, Manfred Ramírez Alfaro****Tarea 3:**

Profesora: Ivonne Zúñiga, PhD

Fecha límite de entrega: 30 de junio de  
2021, 6pm

In [1]:

```
install.packages("devtools")
install.packages("pscl")
install.packages("MASS")
install.packages("margins")
install.packages("haven")
install.packages("mfx")
install.packages("dplyr")
install.packages("VGAM")
install.packages("MixAll")
install.packages("ggeffects")
install.packages("emmeans")
install.packages("HSAUR")
install.packages("survival")
install.packages("mclogit")
install.packages("mlogit")
```

```
There is a binary version available but the source version is later:
      binary source needs_compilation
devtools  2.4.1  2.4.2                FALSE
```

```
installing the source package 'devtools'
```

```
package 'pscl' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
      C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
package 'MASS' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
      C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
package 'margins' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
      C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
package 'haven' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
```

```
C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
package 'mfx' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
```

```
There is a binary version available but the source version is later:
  binary source needs_compilation
dplyr 1.0.6 1.0.7 TRUE
```

```
Binaries will be installed
package 'dplyr' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
package 'VGAM' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
package 'MixAll' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
package 'ggeffects' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
```

```
There is a binary version available but the source version is later:
  binary source needs_compilation
emmeans 1.6.0 1.6.1 FALSE
```

```
installing the source package 'emmeans'
```

```
package 'HSAUR' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
package 'survival' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
package 'mclogit' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
package 'mlogit' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
C:\Users\David Mora Salazar\AppData\Local\Temp\RtmpmIYaMW\downloaded_packages
```

In [16]:

```
library(pscl)
library(MASS)
library(margins)
library(haven)
library(mfx)
library(dplyr)
library(VGAM)
library(MixAll)
library(ggeffects)
library(emmeans)
library(HSAUR)
library(margins)
```

```
library(survival)
library(mclglogit)
library(margins)
library(mlogit)
```

### Primera parte:

In [6]:

```
names(plasma)
dim(plasma)
(plasma)
```

1. 'fibrinogen'

2. 'globulin'

3. 'ESR'

1. 32

2. 3

	fibrinogen	globulin	ESR
<b>1</b>	2.52	38	ESR < 20
<b>2</b>	2.56	31	ESR < 20
<b>3</b>	2.19	33	ESR < 20
<b>4</b>	2.18	31	ESR < 20
<b>5</b>	3.41	37	ESR < 20
<b>6</b>	2.46	36	ESR < 20
<b>7</b>	3.22	38	ESR < 20
<b>8</b>	2.21	37	ESR < 20
<b>9</b>	3.15	39	ESR < 20
<b>10</b>	2.60	41	ESR < 20
<b>11</b>	2.29	36	ESR < 20
<b>12</b>	2.35	29	ESR < 20
<b>16</b>	3.15	36	ESR < 20
<b>18</b>	2.68	34	ESR < 20
<b>19</b>	2.60	38	ESR < 20
<b>20</b>	2.23	37	ESR < 20
<b>21</b>	2.88	30	ESR < 20
<b>22</b>	2.65	46	ESR < 20
<b>24</b>	2.28	36	ESR < 20
<b>25</b>	2.67	39	ESR < 20
<b>26</b>	2.29	31	ESR < 20

	fibrinogen	globulin	ESR
27	2.15	31	ESR < 20
28	2.54	28	ESR < 20
30	3.34	30	ESR < 20
31	2.99	36	ESR < 20
32	3.32	35	ESR < 20
13	5.06	37	ESR > 20
14	3.34	32	ESR > 20
15	2.38	37	ESR > 20
17	3.53	46	ESR > 20
23	2.09	44	ESR > 20
29	3.93	32	ESR > 20

1. Utilice una regresión Logit para conocer la asociación entre ESR y las proteínas plasmáticas. ¿Cuáles son sus resultados? ¿Qué se puede decir de la relación entre las variables? ¿Cuáles coeficientes son significativos? ¿Cómo se interpretan correctamente estos coeficientes?

In [7]:

```
#PRIMERA PARTE
#1. Regresión Logit

logit_plasma <- glm(ESR ~fibrinogen + globulin,
                    family=binomial(link='logit'),data=plasma)

logit_plasma
summary(logit_plasma)
```

Call: glm(formula = ESR ~ fibrinogen + globulin, family = binomial(link = "logit"), data = plasma)

Coefficients:

(Intercept)	fibrinogen	globulin
-12.7921	1.9104	0.1558

Degrees of Freedom: 31 Total (i.e. Null); 29 Residual

Null Deviance: 30.88

Residual Deviance: 22.97 AIC: 28.97

Call:

```
glm(formula = ESR ~ fibrinogen + globulin, family = binomial(link = "logit"),
    data = plasma)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.9683	-0.6122	-0.3458	-0.2116	2.2636

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-12.7921	5.7963	-2.207	0.0273 *
fibrinogen	1.9104	0.9710	1.967	0.0491 *
globulin	0.1558	0.1195	1.303	0.1925

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 30.885 on 31 degrees of freedom  
Residual deviance: 22.971 on 29 degrees of freedom  
AIC: 28.971

Number of Fisher Scoring iterations: 5

Los resultados muestran una relación positiva de odd ratio del ESR ante cambios en las proteínas.

Vemos que en este resultado solo el fibrinogen se considera significativa a 5% de significancia.

La interpretación de esta regresión solo se basa en la dirección del efecto, pero no en su magnitud.

1. Interprete el efecto marginal promedio de las variables.

In [8]:

```
#2. Efecto marginal promedio
EMprom_plasma<-margins(logit_plasma)
summary(EMprom_plasma)
```

	factor	AME	SE	z	p	lower	upper
	fibrinogen	0.20986577	0.08930123	2.350088	0.01876896	0.034838579	0.38489296
	globulin	0.01711319	0.01222729	1.399590	0.16163606	-0.006851849	0.04107823

Solo se interpreta la variable con el coeficiente significativo. En este caso se mide si el nivel del ESR es mayor a 20. Este resultado muestra que ante un aumento de una unidad de fibrinogen, la probabilidad que el nivel de ESR sea mayor a 20 aumenta en promedio 20.99 puntos porcentuales.

1. Interprete el efecto marginal en la media de las variables.

In [9]:

```
#3. Efecto marginal en las medias
medias<-colMeans(plasma[, c("globulin", "fibrinogen")])
(medias)
EM_medias_logit<-margins(logit_plasma, at=list(fibrinogen=medias["fibrinogen"],
globulin=medias["globulin"]))

summary(EM_medias_logit)
```

**globulin** 35.65625  
**fibrinogen** 2.78875

	factor	fibrinogen	globulin	AME	SE	z	p	lower	upper
	fibrinogen	2.78875	35.65625	0.21460871	0.11317806	1.896204	0.05793304	-0.007216204	0.43643363
	globulin	2.78875	35.65625	0.01749995	0.01289568	1.357039	0.17476872	-0.007775124	0.04277502

Los resultados muestran que un aumento en una unidad de fibrinogen aumenta la probabilidad de ESR mayor a 20 en 21.46 puntos porcentuales para un individuo promedio.

1. Calcule la tabla de clasificación. Con respecto dicha tabla, estime la exactitud o  $R^2$ , la sensibilidad y la especificidad.

In [10]:

```
#4. tabla de clasificación
plogit <- predict(logit_plasma, type = "response")

table(plasma$ESR, plogit>0.5)

#con la tabla obtenemos que la exactitud es 0.875
(26+2)/32
#sensibilidad = 0.867
26/30
#La especificidad =1
2/2
```

	FALSE	TRUE
ESR < 20	26	0
ESR > 20	4	2

0.875

0.8666666666666667

1

Con la tabla obtenemos que la exactitud es 0.875. La sensibilidad = 0.867. La especificidad =1

1. Ahora, estime la asociación entre ESR, fibrinogen y globulin con una regresión Probit. Interprete los coeficientes.

In [11]:

```
#5. Regresión probit
plasma_probit <- glm(ESR ~ globulin + fibrinogen,
                     family=binomial(link='probit'), data=plasma)

summary(plasma_probit)
```

Call:

```
glm(formula = ESR ~ globulin + fibrinogen, family = binomial(link = "probit"),
    data = plasma)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.0170	-0.6438	-0.3443	-0.1666	2.2474

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-7.53544	3.06374	-2.460	0.0139 *
globulin	0.09700	0.06667	1.455	0.1457
fibrinogen	1.06791	0.50231	2.126	0.0335 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 30.885 on 31 degrees of freedom  
 Residual deviance: 22.947 on 29 degrees of freedom  
 AIC: 28.947

Number of Fisher Scoring iterations: 6

Los resultados son similares a los de la regresión logit, se continua observando al fibrinogen como

variable significativa para la variable dependiente, Igualmente se observa las direcciones positivas de los efectos, pero no la magnitud de sus coeficientes.

1. Interprete el efecto marginal promedio de las variables.

In [12]:

```
#6. EM Promedio
EMprom_probit_plasma<-margins(plasma_probit)
summary(EMprom_probit_plasma)
```

factor	AME	SE	z	p	lower	upper
fibrinogen	0.21168779	0.08319481	2.544483	0.01094396	0.048628962	0.37474661
globulin	0.01922726	0.01229311	1.564067	0.11780181	-0.004866806	0.04332132

Si se tiene que un aumento en una unidad de fibrinogen aumenta en promedio la probabilidad de ESR>20 en 21.17 puntos porcentuales.

1. Interprete el efecto marginal en la media de las variables.

In [13]:

```
#7. EM en La media
EM_medias_probit_plasma<-margins(plasma_probit, at=list(globulin=medias["globulin"],
                                                         fibrinogen=medias["fibrinogen"])
summary(EM_medias_probit_plasma)
```

factor	globulin	fibrinogen	AME	SE	z	p	lower	upper
fibrinogen	35.65625	2.78875	0.23296349	0.11621446	2.004600	0.04500584	0.005187331	0.46073964
globulin	35.65625	2.78875	0.02115969	0.01413834	1.496617	0.13449281	-0.006550952	0.04887034

Se muestra que un aumento de una unidad de fibrinogen hace que la probabilidad de obtener ESR>20 aumente en 23.3 puntos porcentuales para el individuo promedio.

1. ¿Existe alguna asociación entre la probabilidad de una lectura de ESR superior a 20 mm / h y los niveles de las dos proteínas plasmáticas? Concluya de acuerdo con lo que encontró en los incisos anteriores. Profundice en su respuesta.

Según los resultados obtenidos hay evidencia para decir que existe relación positiva entre la probabilidad de lectura de un ESR>20 y el nivel de Fibrinogen. Sin embargo, no se encuentra evidencia de que el coeficiente para globulin sea diferente de cero, por lo que no hay relación con una lectura de ESR>20. Sin embargo en la literatura existe indicaciones que si hay relación de las proteínas y el aumento de ESR por lo que se podría cuestionar las pocas observaciones en esta base de datos (29 obs) y también que aquellas que muestras ESR>20 con pocas (6/29 obs).

**Tercera parte:** La base Cereales (que se adjunta) contiene características sobre un set de opciones de cereales, dichas características representan lo que el individuo toma en cuenta a la hora de elegir cual comprar. Cabe destacar que las variables presentan información subjetiva de la opinión del individuo, no calzan necesariamente con los datos reales.

1. ¿Qué características debe presentar una base de datos, la variable dependiente en particular, para que sea adecuado utilizar un modelo logit o probit condicional? ¿Cómo se diferencia de un logit o probit multinomial?

La base de datos deben de contener variables que muestren la característica de elección del individuo para realizar un logit o probit condicional. Esto para obtener la utilidad de escogencia de la variable dependiente dado a sus características, omitiendo la relevancia de alternativas irrelevantes. El modelo probit condicional a diferencia de un modelo multinomial, es estimado usando un único vector de parámetros beta. Estos betas, a diferencia de un modelo logit si se interpretan.

1. Corra una regresión condicional de la elección del consumidor. Interprete al menos 3 coeficientes. No utilice la variable cereal, individuos ni identif en la regresión.

```
In [17]: cereales <- read.csv("C:/Users/David Mora Salazar/Documents/ECONOMÍA UNIVERSIDAD DE COS
#2. Regresión Logit condicional
cereales["cheerios"]<- ifelse(cereales$cereal=="Cheerios",c(1),c(0))
cereales["Froot_Loops"]<- ifelse(cereales$cereal=="Froot Loops",c(1),c(0))
cereales["Rice_Krispies"]<- ifelse(cereales$cereal=="Rice Krispies",c(1),c(0))
cereales["SpecialK"]<- ifelse(cereales$cereal=="Special K",c(1),c(0))

cereal_clogit<-clogit(eleccion~grasa+azucar+vitaminas+proteina+calorias+
                      tazas+fibra+carbo+
                      strata(individuos),data = cereales)

cereal_clogit
```

Call:

```
clogit(eleccion ~ grasa + azucar + vitaminas + proteina + calorias +
      tazas + fibra + carbo + strata(individuos), data = cereales)
```

	coef	exp(coef)	se(coef)	z	p
grasa	-0.289392	0.748719	0.253935	-1.140	0.254
azucar	-0.059446	0.942286	0.063754	-0.932	0.351
vitaminas	0.074082	1.076895	0.075918	0.976	0.329
proteina	-0.240432	0.786288	0.207107	-1.161	0.246
calorias	0.011711	1.011780	0.008942	1.310	0.190
tazas	-0.016085	0.984044	0.720141	-0.022	0.982
fibra	-0.699210	0.496978	0.507247	-1.378	0.168
carbo	0.038051	1.038784	0.089793	0.424	0.672

Likelihood ratio test=10.37 on 8 df, p=0.24  
n= 48, number of events= 12

Los resultados para las características escogidas no se muestran significativas para explicar la elección del individuo sobre que cereal consumir. De todas formas, si interpretamos los coeficiente diríamos que el nivel de vitaminas tiene efecto positivo sobre la elección de un cereal con una utilidad marginal de 0.07 mientras que la grasa y el azúcar tienen el efecto contrario, con utilidad marginal negativa en magnitud de -0.3 y -0.06 respectivamente.

1. Interprete las probabilidades predichas para los 3 primeros individuos.

```
In [18]: #3. Probabilidades predichas

prob_clogit<-predict(cereal_clogit,type="expected")
prob_clogit[1:12]
```



1	0.380245937299341
2	0.204475061848844
3	0.348584793063657
4	0.0666942077881579
5	0.546855004323538
6	0.0653804807700171
7	0.289940155337097
8	0.0978243595693481
9	0.0512075707520117
10	0.566008159578124
11	0.156916640172488
12	0.225867629497377

En nuestra base de datos tenemos un total de 12 individuos que deciden que cereal tomar. Vemos que el primer individuo tiene un 0.38 de probabilidad de seleccionar cheerios, un 0.20 de elegir Froot Loops, 0.35 de Rice Krispies y un 0.066 de Special K. Para el segundo individuo, escoge Cheerios con 0.54 de probabilidad, toma Froot Loops con 0.065, Rice Krispies con 0.29 y con probabilidad de 0.098 elije Special K. Para un tercer individuo, este elije Cheerios con 0.051 de probabilidad, Froot Loops con 0.56, Rice Krispies con 0.15 y con 0.22 de probailidad selecciona Special K.

1. Estime los efectos marginales promedio, interprete 3 de ellos.

In [20]:

```
#4. Efectos marginales promedios
cereales_mcllogit<-mcllogit(cbind(eleccion,individuos)~grasa+azucar+vitaminas+proteina+ca
                             tazas+fibra+carbo,data=cereales)

#removing specialk

mpc_mcllogit<-margins(cereales_mcllogit)
summary(mpc_mcllogit)
```

```
Iteration 1 - deviance = 23.15069 - criterion = 0.2802134
Iteration 2 - deviance = 22.90686 - criterion = 0.01059797
Iteration 3 - deviance = 22.90137 - criterion = 0.0002385062
Iteration 4 - deviance = 22.90137 - criterion = 2.192388e-07
Iteration 5 - deviance = 22.90137 - criterion = 2.078986e-13
converged
```

factor	AME	SE	z	p	lower	upper
azucar	-0.007430775	0.007969231	-0.93243311	0.3511127	-0.0230501800	0.008188631
calorias	0.001463894	0.001117756	1.30967147	0.1903070	-0.0007268687	0.003654656
carbo	0.004756407	0.011224163	0.42376501	0.6717372	-0.0172425472	0.026755362
fibra	-0.087401270	0.063405897	-1.37844071	0.1680673	-0.2116745441	0.036872005
grasa	-0.036173949	0.031741905	-1.13962755	0.2544415	-0.0983869396	0.026039041
proteina	-0.030053941	0.025888325	-1.16090713	0.2456797	-0.0807941260	0.020686244
tazas	-0.002010565	0.090017680	-0.02233522	0.9821806	-0.1784419749	0.174420845

factor	AME	SE	z	p	lower	upper
vitaminas	0.009260205	0.009489754	0.97581078	0.3291582	-0.0093393721	0.027859781

Se continua obteniendo resultados no significativos, por lo que ninguna característica cambia la probabilidad de escoger entre un cereal u otro. Si así lo fuese, se diría que el azúcar y la grasa en promedio disminuyen la probabilidad de elegir un cereal en 0.74 y 3.6 puntos porcentuales respectivamente, mientras que las vitaminas tendrían un efecto contrario de 0.93 pp.

1. Estime los efectos marginales en la media, interprete 3 de ellos.

In [22]:

```
#5. Efecto marginales en las medias
medias<-colMeans(cereales[, c("grasa", "azucar", "vitaminas", "carbo", "fibra",
                              "proteina", "tazas", "calorias")])

(medias)
EM_medias_clogit<-margins(cereales_mclogit,
                           at=list(grasa=medias["grasa"],
                                   azucar=medias["azucar"],
                                   carbo=medias["carbo"],
                                   fibra=medias["fibra"],
                                   proteina=medias["proteina"],
                                   tazas=medias["tazas"],
                                   vitaminas=medias["vitaminas"])))

summary(EM_medias_clogit)
```

```
grasa      2.97916666666667
azucar     11.7708333333333
vitaminas  30.1666666666667
carbo      27.6041666666667
fibra      1.5
proteina   6.02083333333333
tazas      2
calorias   126.979166666667
```

factor	grasa	azucar	carbo	fibra	proteina	tazas	vitaminas	AME	SE	
azucar	2.979167	11.77083	27.60417	1.5	6.020833	2	30.16667	-0.007430775	0.007969235	-0.
calorias	2.979167	11.77083	27.60417	1.5	6.020833	2	30.16667	0.001463894	0.001117757	1.
carbo	2.979167	11.77083	27.60417	1.5	6.020833	2	30.16667	0.004756407	0.011224164	0.
fibra	2.979167	11.77083	27.60417	1.5	6.020833	2	30.16667	-0.087401279	0.063405794	-1.
grasa	2.979167	11.77083	27.60417	1.5	6.020833	2	30.16667	-0.036173952	0.031741928	-1.
proteina	2.979167	11.77083	27.60417	1.5	6.020833	2	30.16667	-0.030053945	0.025888348	-1.
tazas	2.979167	11.77083	27.60417	1.5	6.020833	2	30.16667	-0.002010565	0.090017685	-0.
vitaminas	2.979167	11.77083	27.60417	1.5	6.020833	2	30.16667	0.009260206	0.009489758	0.

No se obtienen resultados significativos sobre este modelo por lo que no se puede concluir sobre

las características en la probabilidad de elección de un individuo promedio. Si así fuera se tendría que un aumento de una unidad en grasa y azúcar tendría un aumento de probabilidad de elegir un cereal en 3.62 y en 0.74 puntos porcentuales respectivamente para un individuo promedio. Mientras que un aumento en una unidad en vitaminas aumentaría la probabilidad de escogencia en 0.93 puntos porcentuales para el individuo promedio.

1. ¿Qué es la prueba de Hausman?, ¿Por qué es importante utilizarla en este caso? Explique en detalle.

Se toma el supuesto que existe independencia en razón de probabilidad de las variables por lo que estas no se ven afectadas si se considera una opción alternativa. Si las probabilidades cambian con un aumento de alternativas entonces no existe independencia de alternativas irrelevantes. La prueba de Hausmann muestra si existe o no esta condición. Es importante utilizar la prueba en este y modelos de logit condicional para asegurarse que no cuenta con este sesgo y que y que la incorporación de más opciones no afectaría de manera significativa la elección de un individuo.

1. Estime la prueba de Hausman en este caso, interprete su resultado.

In [25]:

```
#7. Prueba de Hausman

B<-as.vector(cereales_mclogit$coefficients)

VB<-as.matrix(cereales_mclogit$covmat)

cereales1<-cereales[which(cereales$identif != 1),]
cereal_mclogitrestr<-mclogit(cbind(eleccion,individuos)~grasa+azucar+vitaminas+proteina
                             tazas+fibra+carbo,data=cereales1)
b<-as.vector(cereal_mclogitrestr$coefficients)
Vb<-as.matrix(cereal_mclogitrestr$covmat)

B1<-B

VB1<-VB

H<-t(b-B1)%*%solve(Vb-VB1)%*%(b-B1)
H

pvalueH<-1-pchisq(H,8)
pvalueH
```

```
Iteration 1 - deviance = 11.48059 - criterion = 0.4620389
Iteration 2 - deviance = 11.25912 - criterion = 0.01949704
Iteration 3 - deviance = 11.24874 - criterion = 0.0009145294
Iteration 4 - deviance = 11.24867 - criterion = 6.0803e-06
Iteration 5 - deviance = 11.24867 - criterion = 2.952622e-10
converged
4.470013
```

0.8124257

El P-Value es 0.8124257. El H calculado resulta mayor que pchisq con k=8 por lo que hay evidencia para rechazar la hipótesis nula de que haya independencia de alternativas irrelevantes. Por tanto, se recomienda utilizar un modelo logit anidado para este caso.

**Cuarta parte:** Datos de cuenta Deb y Trivedi analizan 4406 individuos de 66 o más años, quienes reciben Medicare en los Estados Unidos. La base de datos forma parte del paquete MixAll de R, su nombre es DebTrivedi. En esta pregunta se trabajará con dicha base.

In [26]:

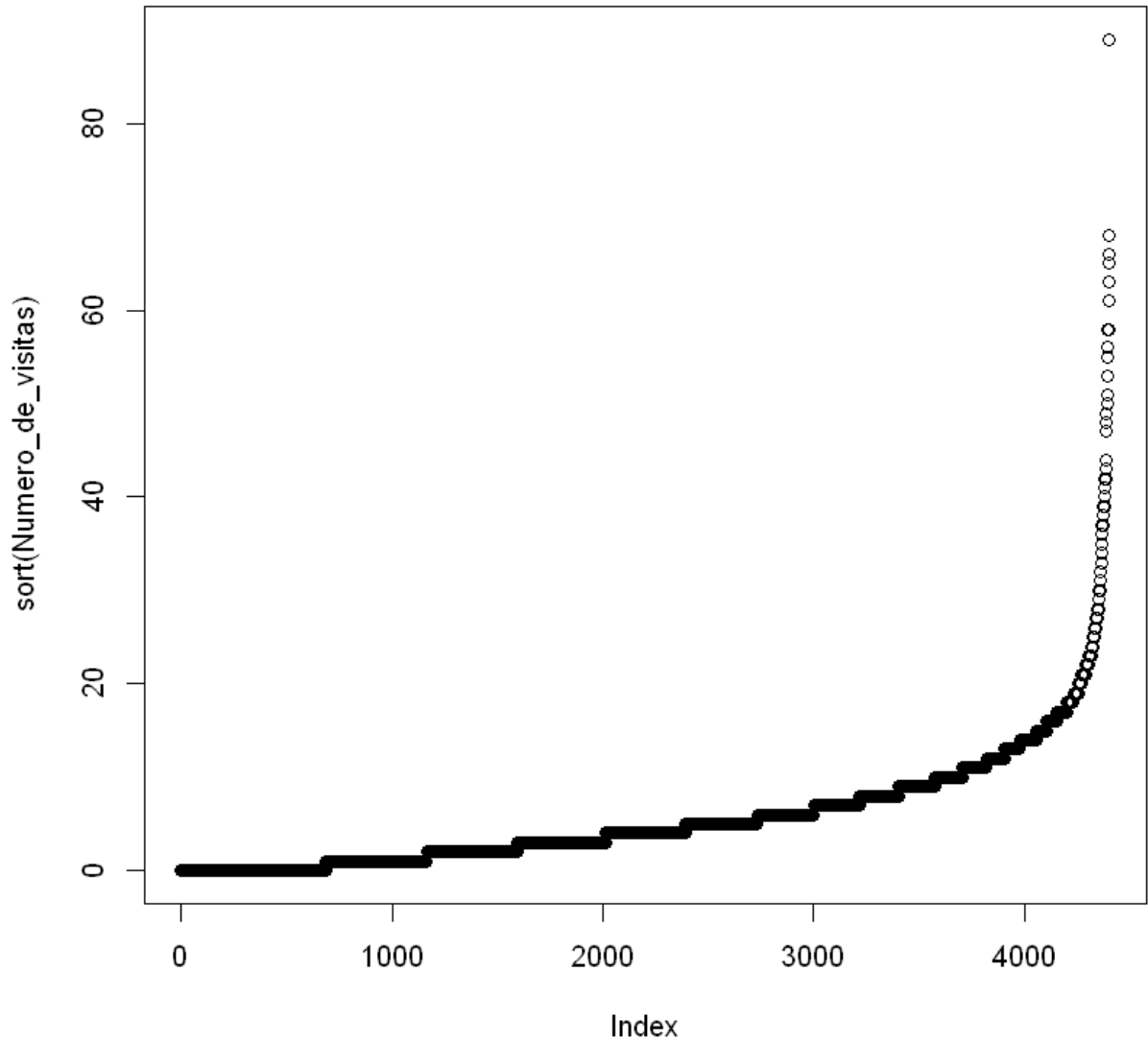
```
# 0- DATOS
data(DebTrivedi)
summary(DebTrivedi)
```

ofp	ofnp	opp	opnp	
Min. : 0.000	Min. : 0.000	Min. : 0.0000	Min. : 0.0000	
1st Qu.: 1.000	1st Qu.: 0.000	1st Qu.: 0.0000	1st Qu.: 0.0000	
Median : 4.000	Median : 0.000	Median : 0.0000	Median : 0.0000	
Mean : 5.774	Mean : 1.618	Mean : 0.7508	Mean : 0.5361	
3rd Qu.: 8.000	3rd Qu.: 1.000	3rd Qu.: 0.0000	3rd Qu.: 0.0000	
Max. : 89.000	Max. : 104.000	Max. : 141.0000	Max. : 155.0000	
emer	hosp	health	numchron	adldiff
Min. : 0.0000	Min. : 0.000	poor : 554	Min. : 0.000	no : 3507
1st Qu.: 0.0000	1st Qu.: 0.000	average : 3509	1st Qu.: 1.000	yes : 899
Median : 0.0000	Median : 0.000	excellent : 343	Median : 1.000	
Mean : 0.2635	Mean : 0.296		Mean : 1.542	
3rd Qu.: 0.0000	3rd Qu.: 0.000		3rd Qu.: 2.000	
Max. : 12.0000	Max. : 8.000		Max. : 8.000	
region	age	black	gender	married
midwest:1157	Min. : 6.600	no : 3890	female:2628	no : 2000
noreast: 837	1st Qu.: 6.900	yes: 516	male : 1778	yes:2406
other :1614	Median : 7.300			
west : 798	Mean : 7.402			
	3rd Qu.: 7.800			
	Max. : 10.900			
school	faminc	employed	privins	medicaid
Min. : 0.00	Min. : -1.0125	no : 3951	no : 985	no : 4004
1st Qu.: 8.00	1st Qu.: 0.9122	yes: 455	yes:3421	yes: 402
Median : 11.00	Median : 1.6982			
Mean : 10.29	Mean : 2.5271			
3rd Qu.: 12.00	3rd Qu.: 3.1728			
Max. : 18.00	Max. : 54.8351			

1. El propósito de los autores es estudiar la demanda de servicios médicos, capturan esta demanda mediante el número de consultas que han hecho los individuos estudiados a un consultorio médico. Grafique el número de visitas a un consultorio médico. ¿Qué se puede decir al respecto? ¿Qué comportamiento se observa?

In [27]:

```
#1- COMPORTAMIENTO
Numero_de_visitas <- DebTrivedi$ofp
## Plot
plot(sort(Numero_de_visitas))
```



```
In [28]: count(DebTrivedi, vars= ofp)
```

vars	n
0	683
1	481
2	428
3	420
4	383
5	338
6	268
7	217
8	188

vars	n
9	171
10	128
11	115
12	86
13	73
14	76
15	53
16	47
17	48
18	30
19	24
20	16
21	18
22	16
23	10
24	12
25	3
26	9
27	7
28	4
29	3
30	4
31	4
32	1
33	1
34	2
35	1
36	1
37	3
38	2
39	5
40	2
41	1
42	4

vars	n
43	2
44	1
47	1
48	1
49	1
50	1
51	1
53	2
55	1
56	1
58	2
61	1
63	1
65	1
66	1
68	1
89	1

Mediante este gráfico se pueden notar algunos puntos interesantes, cuando se ordenan las observaciones por las visitas al médico de menor a mayor, se nota que en más 4000 observaciones, el número de visitas oscila entre cero y veinte, pero para el resto de observaciones que son muy pocas, se nota que el número de visitas aumenta muy exponencialmente. Esto brinda indicios de que muy pocas observaciones acuden más de 20 veces a un consultorio y existen muchos valores relativamente bajos positivos o cero. De hecho el cero es el valor que más se repite (683 veces).

1. Después de observar la distribución de las visitas al médico, ¿se debe proceder con Mínimos Cuadrados Ordinarios? ¿Si o no? ¿Por qué? ¿Qué errores se pueden observar?

Para este análisis, no se puede proceder a estimar mediante MCO debido a que el MCO asume errores distribuidos normalmente, suponiendo que la variable dependiente fuera continua, pero por la naturaleza de los datos, si el número de visitas al consultorio médico es la variable dependiente, la continuidad no se cumple. Si se efectuara la estimación con MCO, se observan algunos errores por los que tener especial cuidado:

1. La regresión podría predecir valores negativos, a diferencia de un modelo de Poisson pues es  $\exp()$  y no resulta en valores negativos, nos asegura que siempre es positiva.
2. Pueden haber errores asociados a la heterocedasticidad, pues resulta que todas las distribuciones estándar de datos de conteo como esta exhiben heterocedasticidad.
3. Una variable de conteo (con un alto número de ceros y números bajos) no puede tener una distribución normal (debido a que la distribución normal es para las variables continuas que

puedan asumir todos los valores) y si asume sólo unos pocos valores, la distribución puede ser muy diferente de la normal.

1. Corra el modelo Poisson donde el numero de visitas al medico sea la variable dependiente. Como variables independientes utilice el numero de estancias hospitalarias, el estado de salud auto percibido, el número de enfermedades crónicas, el género, los años de educación y la variable dicotómica relacionada al seguro privado. En total son 6 variables independientes. Interprete los coeficientes significativos de la regresión

In [29]:

```
#3- Poisson
model_poisson<-glm(ofp ~ hosp + health + numchron + gender + school + privins, data = D
model_poisson
#Coeficientes exponenciados
exp(coef(model_poisson))
```

Call: glm(formula = ofp ~ hosp + health + numchron + gender + school + privins, family = poisson, data = DebTrivedi)

Coefficients:

(Intercept)	hosp	healthpoor	healthexcellent
1.02887	0.16480	0.24831	-0.36199
numchron	gendermale	school	privinsyes
0.14664	-0.11232	0.02614	0.20169

Degrees of Freedom: 4405 Total (i.e. Null); 4398 Residual

Null Deviance: 26940

Residual Deviance: 23170 AIC: 35960

<b>(Intercept)</b>	2.79791415759674
<b>hosp</b>	1.17915418515169
<b>healthpoor</b>	1.28185336416883
<b>healthexcellent</b>	0.696287101681166
<b>numchron</b>	1.15793619969155
<b>gendermale</b>	0.893758281044707
<b>school</b>	1.02648771547188
<b>privinsyes</b>	1.22346485373964

Interpretación: El aumento de una unidad en las estancias hospitalarias aumenta el promedio de visitas al consultorio en 1,18 veces respecto al promedio original. Otra manera de decirlo es: el cambio porcentual en la media del número esperado de visitas al consultorio al aumentar en una unidad la estancia hospitalarias es de 18%.

"Health poor" en comparación con el estado de salud promedio aumenta las visitas al consultorio promedio en 1,28 veces respecto al promedio original. Otra manera de decirlo es: el cambio porcentual en la media del número esperado de visitas al consultorio cuando la observación se considera "Health poor" es de 28%.

Mientras que "health excellent" reduce el promedio de visitas al consultorio en 0,70 veces respecto al promedio original. Otra manera de decirlo es: el cambio porcentual en la media del número esperado de visitas al consultorio cuando la observación se considera "Health excellent" es de -31%.



El aumento en una unidad del número de enfermedades crónicas aumenta el promedio de visitas al consultorio en 1.1580 veces respecto al promedio original. Otra manera de decirlo es: el cambio porcentual en la media del número esperado de visitas al consultorio cuando aumenta en una unidad el número de enfermedades crónicas es de 15.80%.

"Gendermale" en comparación con ser mujer, reduce el promedio de visitas al consultorio en 0.8937 veces respecto al promedio original. Otra manera de decirlo es: el cambio porcentual en la media del número esperado de visitas al consultorio cuando la observación es "Gendermale" es de -10.63%.

El aumento en un año de educación aumenta el promedio de visitas al consultorio en 1.0265 veces respecto al promedio original. Otra manera de decirlo es: el cambio porcentual en la media del número esperado de visitas al consultorio cuando la observación aumenta un año de educación es de 2.65%.

Si la observación posee un seguro privado, en comparación con no poseerlo, aumenta el promedio las visitas al consultorio en 1.2235 veces respecto al promedio original. Otra manera de decirlo es: el cambio porcentual en la media del número esperado de visitas al consultorio cuando la observación posee un seguro privado es de 22.35%.

1. ¿Cuál es la probabilidad de que los primeros 10 individuos de la muestra visiten al medico entre 20 y 10 veces?

```
In [30]: #4- Probabilidad de ocurrencia
#Número esperado de eventos para cada individuo: parámetro lambda
pred_n_poisson<-model_poisson$fitted.values
#Probabilidad de que visiten al consultorio médico entre 10 y 20 veces.
prob_10_20_poi<-ppois(q=20,lambda=pred_n_poisson, lower.tail = TRUE)-ppois(q=9,lambda=p
```

```
In [31]: prob_10_20_poi[1:10]
```

1	0.0624208696426326
2	0.0812767623170312
3	0.836554278811941
4	0.14100788115603
5	0.0472780637888054
6	0.407621554574442
7	0.0114575087339053
8	0.0114575087339053
9	0.0114575087339053
10	0.0114575087339053

La probabilidad de que el individuo 1 visite al consultorio medico entre 20 y 10 veces es de 6.242087e-02

La probabilidad de que el individuo 2 visite al consultorio medico entre 20 y 10 veces es de 8.127676e-02

La probabilidad de que el individuo 3 visite al consultorio medico entre 20 y 10 veces es de 8.365543e-01

La probabilidad de que el individuo 4 visite al consultorio medico entre 20 y 10 veces es de 1.410079e-01

La probabilidad de que el individuo 5 visite al consultorio medico entre 20 y 10 veces es de 4.727806e-02

La probabilidad de que el individuo 6 visite al consultorio medico entre 20 y 10 veces es de 4.076216e-01

La probabilidad de que el individuo 7 visite al consultorio medico entre 20 y 10 veces es de 1.145751e-02

La probabilidad de que el individuo 8 visite al consultorio medico entre 20 y 10 veces es de 1.145751e-02

La probabilidad de que el individuo 9 visite al consultorio medico entre 20 y 10 veces es de 1.145751e-02

La probabilidad de que el individuo 10 visite al consultorio medico entre 20 y 10 veces es de 1.145751e-02

1. Interprete los efectos marginales promedio relacionados a coeficientes significativos.

In [32]:

```
#5- Efectos marginales promedio
EMprom_media_poi<-poissonmfx(model_poisson,data=DebTrivedi,atmean=FALSE)
```

In [33]:

```
EMprom_media_poi
```

Call:

```
poissonmfx(formula = model_poisson, data = DebTrivedi, atmean = FALSE)
```

Marginal Effects:

	dF/dx	Std. Err.	z	P> z	
hosp	0.951606	0.035141	27.0794	< 2.2e-16	***
healthpoor	1.558203	0.121673	12.8065	< 2.2e-16	***
healthexcellent	-1.789118	0.127329	-14.0512	< 2.2e-16	***
numchron	0.846754	0.026972	31.3933	< 2.2e-16	***
gendermale	-0.641105	0.073072	-8.7736	< 2.2e-16	***
school	0.150960	0.010686	14.1268	< 2.2e-16	***
privinsyes	1.099516	0.086818	12.6646	< 2.2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

dF/dx is for discrete change for the following variables:

```
[1] "healthpoor" "healthexcellent" "gendermale" "privinsyes"
```

El efecto marginal de una hospitalización adicional se observa como un cambio promedio de 0.9516 en el número esperado de visitas al consultorio.

El efecto marginal de que la observación se perciba "healthpoor" se observa como un cambio promedio de 1.558203 en el número esperado de visitas al consultorio.

El efecto marginal de que la observación se perciba "healthexcellent" se observa como un cambio promedio de -1.789118 en el número esperado de visitas al consultorio.

El efecto marginal de una enfermedad crónica adicional se observa como un cambio promedio de 0.846754 en el número esperado de visitas al consultorio.

El efecto marginal de que la observación sea hombre se observa como un cambio promedio de -0.641105 en el número esperado de visitas al consultorio.

El efecto marginal de un año adicional de educación se observa como un cambio promedio de 0.150960 en el número esperado de visitas al consultorio.

El efecto marginal de que la observación tenga seguro privado se observa como un cambio promedio de 1.099516 en el número esperado de visitas al consultorio.

1. Interprete los efectos marginales en la media relacionados a coeficientes significativos.

In [34]:

```
#6- Efectos marginales en la media
```

```
EMmedias_media_poi<-poissonmfx(formula = model_poisson, data = DebTrivedi)
EMmedias_media_poi
```

Call:

```
poissonmfx(formula = model_poisson, data = DebTrivedi)
```

Marginal Effects:

	dF/dx	Std. Err.	z	P> z	
hosp	0.8905263	0.0324017	27.4840	< 2.2e-16	***
healthpoor	1.4762513	0.1162612	12.6977	< 2.2e-16	***
healthexcellent	-1.6881006	0.1205984	-13.9977	< 2.2e-16	***
numchron	0.7924041	0.0244046	32.4695	< 2.2e-16	***
gendermale	-0.6007256	0.0684608	-8.7747	< 2.2e-16	***
school	0.1412706	0.0099218	14.2383	< 2.2e-16	***
privinsyes	1.0325141	0.0815608	12.6594	< 2.2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

dF/dx is for discrete change for the following variables:

```
[1] "healthpoor" "healthexcellent" "gendermale" "privinsyes"
```

El efecto marginal de una hospitalización adicional se observa como un cambio de 0.8905263 del número esperado de visitas al consultorio del individuo promedio.

El efecto marginal de que la observación se perciba "healthpoor" se observa como un cambio de 1.4762513 del número esperado de visitas al consultorio del individuo promedio.

El efecto marginal de que la observación se perciba "healthexcellent" se observa como un cambio de -1.6881006 del número esperado de visitas al consultorio del individuo promedio.

El efecto marginal de una enfermedad crónica adicional se observa como un cambio de 0.7924041 del número esperado de visitas al consultorio del individuo promedio.

El efecto marginal de que la observación sea hombre se observa como un cambio de -0.6007256 del número esperado de visitas al consultorio del individuo promedio.

El efecto marginal de un año adicional de educación se observa como un cambio de 0.1412706 del número esperado de visitas al consultorio del individuo promedio.

El efecto marginal de que la observación tenga seguro privado se observa como un cambio de 1.0325141 del número esperado de visitas al consultorio del individuo promedio.

1. ¿Cuál es el numero esperado de eventos para los primeros 10 individuos?

El número esperado de eventos para cada individuo está enumerado desde el primero hasta el décimo.

In [35]:

```
#7- Número esperado de eventos
model_poisson[["fitted.values"]][1:10]
```

1	5.65859171627547
2	5.9611864986961
3	13.729686201501
4	6.70633799914545
5	5.36930017164558
6	8.96545186711645
7	4.21945086772773
8	4.21945086772773
9	4.21945086772773
10	4.21945086772773

1. Un supuesto base de la regresión Poisson es:  $\mu = \sigma$  al cuadrado . Averigüe si se cumple este supuesto. Utilice la Prueba de bondad de ajuste, el Pearson goodness of fit, la prueba de sobredispersión de Cameron y Trivedi y la Prueba LR Poisson vs Negative Binomial. En cada una debe indicar si se rechaza o no la hipótesis nula, además de mencionar cual es dicha hipótesis nula. Concluya con respecto al supuesto base.

Prueba de bondad de ajuste de Poisson

In [36]:

```
# Estadístico
model_poisson$deviance
#P-value
pchisq(model_poisson$deviance, df=model_poisson$df.residual, lower.tail=FALSE)
```

23167.8062410322

0

La H0 es: Los datos se distribuyen como Poisson. Se rechaza la H0 que los datos se distribuyen como Poisson. Se concluye que los datos no cumplen con los criterios de la distribución de Poisson y no se ajustan a esta distribución.

Pearson Goodness of fit

```
In [37]: Estad <- sum(residuals(model_poisson, type = "pearson")^2)
        prob <- 1-pchisq(Estad,model_poisson$df.residual)
        Estad
        prob
```

29493.5867940311

0

Este test funciona para analizar si el modelo de Poisson es apropiado para los datos o si es conveniente usar un modelo Binomial Negativo de manera que se pueda decidir el modelo que se ajusta mejor. Si el test es significativo, el modelo de Poisson se considera inapropiado. La  $H_0$  es que los datos se distribuyen como Poisson. Se rechaza la  $H_0$ . Se concluye que es más apropiado usar el modelo Binomial Negativo.

Fuente: poisson postestimation — Postestimation tools for poisson, Stata.com

Prueba de sobredispersión de Cameron y Trivedi

```
In [38]: pr_ofp <- predict(model_poisson, type = 'response')
        data1 <- DebTrivedi %>%
          mutate(ys = ((ofp - pr_ofp)^2-ofp)/pr_ofp) %>%
          mutate(mutecho = pr_ofp)
        camerontriv <- lm(formula = ys ~ 0 + mutecho, data=data1)
        summary(camerontriv)
```

Call:

```
lm(formula = ys ~ 0 + mutecho, data = data1)
```

Residuals:

Min	1Q	Median	3Q	Max
-24.54	-5.35	-3.08	0.27	1168.82

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
mutecho	0.89539	0.07872	11.37	<2e-16 ***

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 32.92 on 4405 degrees of freedom

Multiple R-squared: 0.02853, Adjusted R-squared: 0.02831

F-statistic: 129.4 on 1 and 4405 DF, p-value: < 2.2e-16

La  $H_0$  es: la varianza es igual a la media. Se rechaza la  $H_0$  de que la varianza sea igual a la media. Se concluye que que existe sobredispersión (la varianza no es igual a la media).

Prueba LR Poisson vs Negative Binomial

```
In [39]: # Negative Binomial
        NB <- glm.nb(formula = ofp ~ hosp + health + numchron + gender + school + privins, data
        #LR Test
        P <- logLik(model_poisson)
        N <- logLik(NB)
        teststat <- -2 * (as.numeric(P)-as.numeric(N))
        pvalue_nb <- 1-pchisq(teststat,1)
        teststat
        pvalue_nb
```

11602.1184268837

0

La  $H_0$  es que  $\alpha$  es igual a cero. Se rechaza la hipótesis nula que  $\alpha$  es cero. Se concluye que hay evidencia de sobredispersión, lo que indica que no es adecuado usar un modelo de Poisson pues asume que no hay sobredispersión.

1. Corra el modelo de la binomial negativa, tome las mismas variables usó en el punto 3. ¿Cómo se comparan los resultados de este modelo con respecto al modelo Poisson? Compare coeficientes.

In [40]:

```
NB <- glm.nb(formula = ofp ~ hosp + health + numchron + gender + school + privins, data = DebTrivedi)
summary(NB)
```

Call:

```
glm.nb(formula = ofp ~ hosp + health + numchron + gender + school + privins, data = DebTrivedi, init.theta = 1.206603534, link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.0469	-0.9955	-0.2948	0.2961	5.8185

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.929257	0.054591	17.022	< 2e-16 ***
hosp	0.217772	0.020176	10.793	< 2e-16 ***
healthpoor	0.305013	0.048511	6.288	3.23e-10 ***
healthexcellent	-0.341807	0.060924	-5.610	2.02e-08 ***
numchron	0.174916	0.012092	14.466	< 2e-16 ***
gendermale	-0.126488	0.031216	-4.052	5.08e-05 ***
school	0.026815	0.004394	6.103	1.04e-09 ***
privinsyes	0.224402	0.039464	5.686	1.30e-08 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.2066) family taken to be 1)

Null deviance: 5743.7 on 4405 degrees of freedom  
 Residual deviance: 5044.5 on 4398 degrees of freedom  
 AIC: 24359

Number of Fisher Scoring iterations: 1

Theta: 1.2066  
 Std. Err.: 0.0336

2 x log-likelihood: -24341.1070

A nivel teórico: El modelo Binomial Negativa con respecto al modelo de Poisson se diferencia debido a que el modelo de Poisson asume que la varianza es igual a la media (no asume sobredispersión) sino se cumple este supuesto se altera la inferencia.

La Binomial Negativa incluye el supuesto de sobredispersión en caso de que lo exista en forma de un componente aleatorio  $V$ . Entonces la Binomial Negativa incluye un sistema mixto de dos distribuciones, una de Poisson y una distribución Gamma, si no existe sobredispersión, la Binomial

Negativa se transforma en una distribución de Poisson, de lo contrario (existe sobredispersión) esta BN incluye las dos distribuciones.

Por tanto, si no hay sobredispersión los coeficientes estimados entre la Poisson y la Binomial Negativa serán los mismos, si hay sobredispersión los coeficientes estimados serán diferentes por el efecto de otra distribución.

A nivel matemático: Para el coeficiente de "hosp" con la BN aumenta 0.0529 el valor respecto a Poisson.

Para el coeficiente de "healthpoor" con la BN aumenta 0.05676 el valor respecto a Poisson.

Para el coeficiente de "healthexcellent" con la BN aumenta 0.02018 el valor respecto a Poisson.

Para el coeficiente de "numchron" con la BN aumenta 0.028277 el valor respecto a Poisson.

Para el coeficiente de "gendermale" con la BN reduce 0.014168 el valor respecto a Poisson.

Para el coeficiente de "school" con la BN aumenta 21/31250 el valor respecto a Poisson.

Para el coeficiente de "privinsyes" con la BN aumenta 0.022715 el valor respecto a Poisson.

1. Estime el modelo en dos partes o "hurdle model", donde la especificación del modelo de cuenta es una binomial negativa (dist = binomial negativa). Puntos extra: Interprete los coeficientes del modelo.

In [41]:

```
#10- Hurdle Model
```

```
HURDLE <- hurdle(formula = ofp ~ hosp + health + numchron + gender + school + privins,
summary(HURDLE)
```

Call:

```
hurdle(formula = ofp ~ hosp + health + numchron + gender + school + privins,
data = DebTrivedi, dist = "negbin", zero.dist = "binomial", link = "logit")
```

Pearson residuals:

Min	1Q	Median	3Q	Max
-1.1718	-0.7080	-0.2737	0.3196	18.0092

Count model coefficients (truncated negbin with log link):

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.197699	0.058973	20.309	< 2e-16 ***
hosp	0.211898	0.021396	9.904	< 2e-16 ***
healthpoor	0.315958	0.048056	6.575	4.87e-11 ***
healthexcellent	-0.331861	0.066093	-5.021	5.14e-07 ***
numchron	0.126421	0.012452	10.152	< 2e-16 ***
gendermale	-0.068317	0.032416	-2.108	0.0351 *
school	0.020693	0.004535	4.563	5.04e-06 ***
privinsyes	0.100172	0.042619	2.350	0.0188 *
Log(theta)	0.333255	0.042754	7.795	6.46e-15 ***

Zero hurdle model coefficients (binomial with logit link):

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.043147	0.139852	0.309	0.757688
hosp	0.312449	0.091437	3.417	0.000633 ***
healthpoor	-0.008716	0.161024	-0.054	0.956833
healthexcellent	-0.289570	0.142682	-2.029	0.042409 *
numchron	0.535213	0.045378	11.794	< 2e-16 ***

```

gendermale      -0.415658    0.087608   -4.745 2.09e-06 ***
school          0.058541    0.011989    4.883 1.05e-06 ***
privinsyes      0.747120    0.100880    7.406 1.30e-13 ***

```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Theta: count = 1.3955
```

```
Number of iterations in BFGS optimization: 16
```

```
Log-likelihood: -1.209e+04 on 17 Df
```

Los coeficientes de un modelo de dos partes no son suficientes para mostrar el efectos de las variables significativa, por lo tanto, se requiere analizar los efectos marginales considerando lo siguiente:

Se asume que la variable ofp no es categórica, pues no lo es, por tanto, usar ggpredict o ggmemmeans son exactamente iguales. "If you have no categorical predictors as non-focal terms (i.e. no factor needs to be held constant), then - as shown above - ggpredict() and ggmemmeans() yield the same results."

Fuentes: 1.<https://cran.r-project.org/web/packages/ggeffects/readme/README.html> 2.[https://cran.r-project.org/web/packages/ggeffects/vignettes/technical\\_differencepredictmemmeans.html](https://cran.r-project.org/web/packages/ggeffects/vignettes/technical_differencepredictmemmeans.html)

```
In [42]: EM_atmeans_hurdle<-ggpredict(HURDLE)
```

```
In [43]: EM_atmeans_hurdle
```

\$hosp	x	predicted	std.error	conf.low	conf.high	group
	0	6.473406	0.04285508	5.951885	7.040624	hosp
	1	8.001272	0.04622345	7.308253	8.760009	hosp
	2	9.889749	0.05789853	8.828806	11.078185	hosp
	3	12.223948	0.07405036	10.572549	14.133290	hosp
	4	15.109069	0.09235937	12.607276	18.107319	hosp
	5	18.675143	0.11177045	15.001157	23.248938	hosp
	6	23.082888	0.13179756	17.828051	29.886595	hosp
	7	28.530959	0.15219772	21.172173	38.447429	hosp
	8	35.264894	0.17283887	25.131701	49.483829	hosp

\$health	x	predicted	std.error	conf.low	conf.high	group
	poor	6.473406	0.04285508	5.951885	7.040624	health
	average	4.719698	0.05793629	4.213071	5.287248	health
	excellent	3.386794	0.07446836	2.926855	3.919010	health

\$numchron	x	predicted	std.error	conf.low	conf.high	group
	0	5.704651	0.04691464	5.203496	6.254072	numchron



x	predicted	std.error	conf.low	conf.high	group
1	6.473406	0.04285508	5.951885	7.040624	numchron
2	7.345758	0.04221673	6.762412	7.979424	numchron
3	8.335667	0.04514496	7.629796	9.106843	numchron
4	9.458976	0.05102941	8.558700	10.453951	numchron
5	10.733662	0.05899199	9.561672	12.049304	numchron
6	12.180123	0.06830980	10.653842	13.925060	numchron
7	13.821508	0.07850174	11.850444	16.120415	numchron
8	15.684085	0.08926890	13.166592	18.682931	numchron

**\$gender**

x	predicted	std.error	conf.low	conf.high	group
female	6.473406	0.04285508	5.951885	7.040624	gender
male	6.045930	0.04710184	5.512771	6.630653	gender

**\$school**

x	predicted	std.error	conf.low	conf.high	group
0	5.155567	0.05549561	4.624220	5.747968	school
1	5.263364	0.05254615	4.748277	5.834327	school
2	5.373414	0.04983534	4.873381	5.924753	school
3	5.485766	0.04740416	4.999042	6.019879	school
4	5.600467	0.04529765	5.124680	6.120426	school
5	5.717566	0.04356294	5.249650	6.227188	school
6	5.837113	0.04224584	5.373267	6.341001	school
7	5.959160	0.04138625	5.494865	6.462687	school
8	6.083759	0.04101295	5.613862	6.592988	school
9	6.210963	0.04113917	5.729823	6.732505	school
10	6.340827	0.04176039	5.842509	6.881647	school
11	6.473406	0.04285508	5.951885	7.040624	school
12	6.608757	0.04438824	6.058100	7.209466	school
13	6.746938	0.04631633	6.161440	7.388075	school
14	6.888009	0.04859238	6.262270	7.576273	school
15	7.032029	0.05116996	6.360990	7.773858	school
16	7.179060	0.05400592	6.457994	7.980636	school
17	7.329165	0.05706175	6.553653	8.196446	school
18	7.482409	0.06030404	6.648300	8.421169	school

**\$privins**

x	predicted	std.error	conf.low	conf.high	group
---	-----------	-----------	----------	-----------	-------

	<b>x</b>	<b>predicted</b>	<b>std.error</b>	<b>conf.low</b>	<b>conf.high</b>	<b>group</b>
	no	6.473406	0.04285508	5.951885	7.040624	privins
	yes	7.155448	0.02567068	6.804338	7.524675	privins

A nivel de interpretación, las estimaciones de los coeficientes no se pueden interpretar directamente. Para obtener el efecto de una covariable en el modelo, necesitamos usar el comando `margins` para obtener los efectos marginales.

Cada tabla muestra los efectos marginales sobre la media de un aumento en una unidad de cada covariable, y para cada cambio en cada covariable categórica o dicotómica. Debido a la magnitud de las tablas, no se va a interpretar cada valor, pero a nivel general se debe interpretar como el cambio en la media del número de visitas al consultorio, asociado a un cambio de 1 unidad (o el cambio categórico) en la variable de interés, para el INDIVIDUO PROMEDIO. Fuentes bibliográficas para la interpretación:

1. <https://data.library.virginia.edu/getting-started-with-hurdle-models/>
2. <https://www.rdocumentation.org/packages/pscl/versions/1.5.5/topics/hurdle>
3. <https://www.stata.com/stata14/hurdle-models/>