

Universidad De El Salvador  
Facultad Multidisciplinaria De Occidente  
Departamento De Matemática  
Licenciatura En Estadística



Aplicación Del Análisis De Regresión Logística  
Multinomial Para Predecir Trastornos Del Sueño

Docente:

Licenciado Jaime Isaac Peña

Presentado por:

Jenniffer Beatriz Alemán Castillo

## Table of contents

Análisis De Regresión Logística Multinomial De Los Trastornos Del Sueño	3
Supuestos De La Regresión Logística Multinomial	4
Base De Datos Que Se Utilizará	4
Análisis Descriptivo De Los Datos	5
Outliers	7
Correlación Entre Las Variables	8
Aplicación De La Técnica En El Software SPSS	9
Descripción De Los Datos	11
Resumen De Procesamiento De Casos	12
Ajuste Del Modelo	12
Prueba De Bondad De Ajuste Del Modelo	13
Evaluación Del Ajuste Del Modelo	14
Pruebas De La Razón De Verosimilitud	14
Estimación De Los Parámetros	15
Clasificación	17
Aplicación de la Técnica En El Software RStudio	18
Exponentes De Los Coeficientes Estimado Del Modelo	20
Probabilidades Ajustadas Para Cada Observación En El Modelo	21
Predicción Y Precisión Del Modelo	22
Aplicación De La Técnica En El Software Python	23
Ajuste Del Modelo	24

# **Análisis De Regresión Logística Multinomial De Los Trastornos Del Sueño**

La regresión logística multinomial es una técnica de análisis de datos que se utiliza para predecir la probabilidad de que una variable dependiente categórica pertenezca a una categoría específica en función de varias variables independientes. Esta técnica se utiliza en diversas áreas, como la medicina, la psicología y la sociología, para analizar variables dependientes categóricas y relacionarlas con variables independientes. La regresión logística multinomial generaliza el método de regresión logística para problemas multiclase, es decir, con más de dos posibles resultados discretos. La regresión logística multinomial se basa en la relación entre una variable dependiente categórica y varias variables independientes. En esta investigación se realizará una aplicación del análisis de regresión logística multinomial para predecir si una persona tiene o no un trastorno del sueño, en este caso se tomará como variable dependiente el desorden del sueño que una persona puede o no presentar y se tendrán en cuenta las categorías apnea del sueño, el insomnio o ningún trastorno del sueño. Las variables independientes serán las características que presentan las personas, como la calidad del sueño de la persona, el nivel de actividad físico o el estrés que la persona presenta, etc. Al analizar estas características, se puede predecir si una persona presenta un trastorno del sueño. La aplicación de la regresión logística multinomial para la identificación de personas con trastorno de sueño puede ser de gran utilidad en la prevención de problemas de salud de una persona, en este caso en específico se tomará en cuenta las siguientes variables para realizar el análisis:

1. Genero
2. Edad
3. Horas de sueño
4. Calidad del sueño
5. Nivel de actividad física
6. Estrés
7. Índice de Masa Corporal
8. Ritmo cardiaco
9. Pasos diarios
10. Desorden del sueño

Los trastornos del sueño pueden afectar la salud de las personas, por lo que es importante identificar a las personas que están más propensas a presentar un trastorno del sueño y prevenir estos trastornos. La regresión logística multinomial es una técnica útil para este propósito, ya que permite evaluar la relación entre una variable dependiente categórica (en este caso, la presencia o ausencia de un trastorno del sueño) y varias variables independientes.

## Supuestos De La Regresión Logística Multinomial

Al realizar un análisis de regresión logística multinomial es importante tener en cuenta los supuestos que deben cumplir los datos para que sea viable realizar este tipo de análisis. Algunos de los supuestos que deben cumplirse para realizar un análisis de regresión logística multinomial son:

1. Independencia de las observaciones: las observaciones deben ser independientes entre sí, es decir, que los valores de una observación no deben estar relacionados con los valores de otra observación.
2. Ausencia de multicolinealidad: las variables independientes no deben estar altamente correlacionadas entre sí, ya que esto puede afectar la precisión de las estimaciones de los coeficientes.
3. Linealidad en los registros logarítmicos: la relación entre las variables independientes y la variable dependiente debe ser lineal en los registros logarítmicos. Ausencia de valores atípicos: los valores atípicos pueden afectar la precisión de las estimaciones de los coeficientes y deben ser identificados y tratados adecuadamente.

Es importante tener en cuenta que estos supuestos deben ser evaluados antes de realizar el análisis de regresión logística multinomial para garantizar la validez y precisión de los resultados.

## Base De Datos Que Se Utilizará

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

```
filter, lag
```

The following objects are masked from 'package:base':

```
intersect, setdiff, setequal, union
```

```
library(readxl)
suenio <- read_excel("D:/Beatriz/Ciclo II 2023/Proyectos de Estudios Estadísticos/REGRESIÓN")
```

```
base_suenio <- select(suenio,
                      "Genero",
                      "edad",
                      "horasSueno",
                      "calidadsueno",
                      "nivActividadFisica",
                      "estrés",
                      "CategoriaIMC",
                      "ritmocardiaaco",
                      "pasos diarios",
                      "desordeSuenio")

head(base_suenio)
```

```
# A tibble: 6 x 10
  Genero  edad horasSueno calidadsueno nivActividadFisica estrés CategoriaIMC
  <dbl> <dbl>    <dbl>        <dbl>          <dbl> <dbl>        <dbl>
1     1    27      6.1          6            42     6          2
2     1    28      6.2          6            60     8          1
3     1    28      6.2          6            60     8          1
4     1    28      5.9          4            30     8          3
5     1    28      5.9          4            30     8          3
6     1    28      5.9          4            30     8          3
# i 3 more variables: ritmocardiaaco <dbl>, `pasos diarios` <dbl>,
#   desordeSuenio <dbl>
```

## Análisis Descriptivo De Los Datos

Primero se debe saber la cantidad de filas y columnas por las que está formada la base de datos, además de conocer los nombres de las columnas:

```
# Filas y columnas:

dim(base_suenio)
```

```
[1] 374 10
```

```
# Nombre de las columnas
colnames(base_suenio)
```

```
[1] "Genero"          "edad"            "horasSueno"      "calidadsueno"
[5] "nivActividadFisica" "estrés"          "CategoriaIMC"    "ritmocardiaaco"
[9] "pasos diarios"    "desordeSuenio"
```

Se tiene 23 variables y 8124 observaciones.

```
str(base_suenio)
```

```
tibble [374 x 10] (S3: tbl_df/tbl/data.frame)
 $ Genero          : num [1:374] 1 1 1 1 1 1 1 1 1 1 ...
 $ edad            : num [1:374] 27 28 28 28 28 28 29 29 29 29 ...
 $ horasSueno      : num [1:374] 6.1 6.2 6.2 5.9 5.9 5.9 6.3 7.8 7.8 7.8 ...
 $ calidadsueno    : num [1:374] 6 6 6 4 4 4 6 7 7 7 ...
 $ nivActividadFisica: num [1:374] 42 60 60 30 30 30 40 75 75 75 ...
 $ estrés          : num [1:374] 6 8 8 8 8 8 7 6 6 6 ...
 $ CategoriaIMC     : num [1:374] 2 1 1 3 3 3 3 1 1 1 ...
 $ ritmocardiaaco   : num [1:374] 77 75 75 85 85 85 82 70 70 70 ...
 $ pasos diarios    : num [1:374] 4200 10000 10000 3000 3000 3000 3500 8000 8000 8000 ...
 $ desordeSuenio    : num [1:374] 1 1 1 2 2 3 3 1 1 1 ...
```

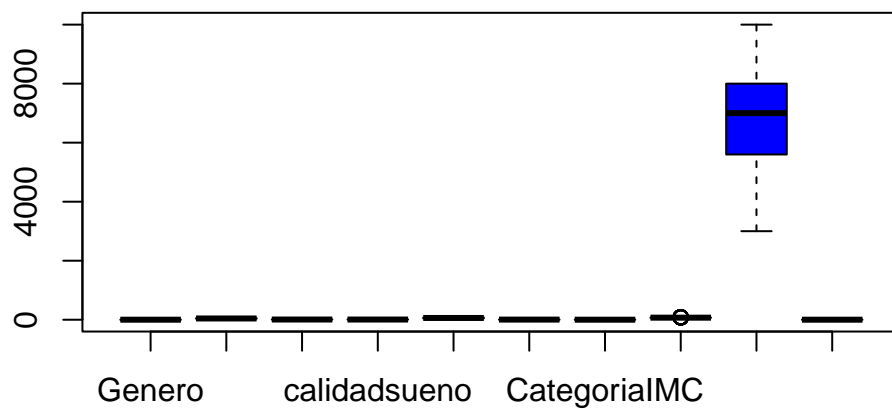
```
summary(base_suenio)
```

Genero	edad	horasSueno	calidadsueno
Min. :0.0000	Min. :27.00	Min. :5.800	Min. :4.000
1st Qu.:0.0000	1st Qu.:35.25	1st Qu.:6.400	1st Qu.:6.000
Median :1.0000	Median :43.00	Median :7.200	Median :7.000
Mean :0.5053	Mean :42.18	Mean :7.132	Mean :7.313
3rd Qu.:1.0000	3rd Qu.:50.00	3rd Qu.:7.800	3rd Qu.:8.000
Max. :1.0000	Max. :59.00	Max. :8.500	Max. :9.000
nivActividadFisica	estrés	CategoriaIMC	ritmocardiaaco
Min. :30.00	Min. :3.000	Min. :1.000	Min. :65.00
1st Qu.:45.00	1st Qu.:4.000	1st Qu.:1.000	1st Qu.:68.00
Median :60.00	Median :5.000	Median :1.000	Median :70.00
Mean :59.17	Mean :5.385	Mean :1.449	Mean :70.17
3rd Qu.:75.00	3rd Qu.:7.000	3rd Qu.:2.000	3rd Qu.:72.00
Max. :90.00	Max. :8.000	Max. :3.000	Max. :86.00

pasos diarios	desordeSuenio
Min. : 3000	Min. :1.00
1st Qu.: 5600	1st Qu.:1.00
Median : 7000	Median :1.00
Mean : 6817	Mean :1.62
3rd Qu.: 8000	3rd Qu.:2.00
Max. :10000	Max. :3.00

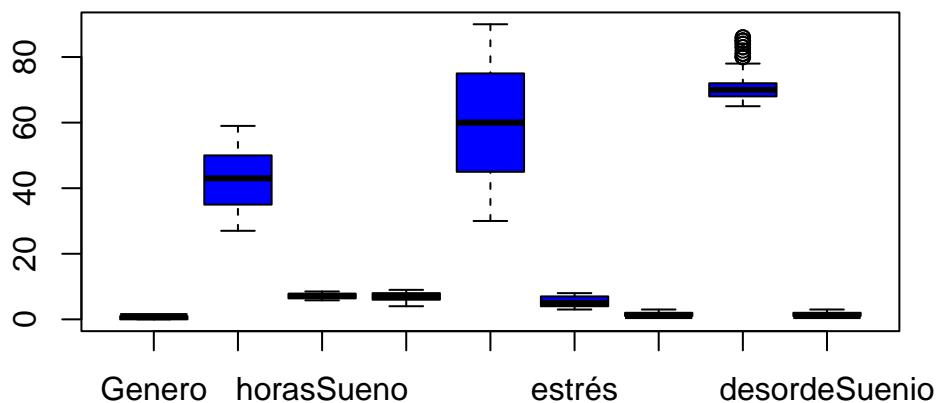
## Outliers

```
boxplot(base_suenio, col = "blue")
```



No se puede observar la salida de todas las variables debido a que la variable "pasos diarios", por lo que se procede a eliminar dicha variable para poder observar las demás salidas

```
sin_pasos <- base_suenio[, -c(9)]
boxplot(sin_pasos, col = "blue")
```



Se observa que la variable 'ritmo cardiaco' presenta valores atípicos en el conjunto de datos. Esto se refleja en una amplia variabilidad en los valores de ritmo cardiaco, con algunas personas exhibiendo ritmo cardiaco muy alto y otras presentando valores significativamente bajos. Estos valores extremos en el ritmo cardiaco podrían indicar la presencia de datos atípicos.

## Correlación Entre Las Variables

```
cor(base_suenio)
```

	Genero	edad	horasSueno	calidadsueno
Genero	1.000000000	-0.5963577	-0.12157854	-0.29136586
edad	-0.596357670	1.0000000	0.34470936	0.47373388
horasSueno	-0.121578538	0.3447094	1.00000000	0.88321300
calidadsueno	-0.291365855	0.4737339	0.88321300	1.00000000
nivActividadFisica	0.001454463	0.1789927	0.21236031	0.19289645
estrés	0.396018154	-0.4223445	-0.81102303	-0.89875203
CategoriaIMC	-0.242743191	0.4277036	-0.34960530	-0.32441269
ritmocardiaaco	0.217104841	-0.2256062	-0.51645489	-0.65986473
pasos diarios	-0.014509370	0.0579734	-0.03953254	0.01679141
desordeSuenio	-0.161130852	0.3039616	-0.38204517	-0.35747699



	nivActividadFisica	estrés	CategoriaIMC	ritmocardiaaco
Genero	0.001454463	0.39601815	-0.24274319	0.21710484
edad	0.178992720	-0.42234448	0.42770355	-0.22560619
horasSueno	0.212360315	-0.81102303	-0.34960530	-0.51645489
calidadsueno	0.192896455	-0.89875203	-0.32441269	-0.65986473
nivActividadFisica	1.000000000	-0.03413446	0.05207636	0.13697098
estrés	-0.034134464	1.00000000	0.15764925	0.67002646
CategoriaIMC	0.052076357	0.15764925	1.00000000	0.46989650
ritmocardiaaco	0.136970983	0.67002646	0.46989650	1.00000000
pasos diarios	0.772723050	0.18682895	-0.14278493	-0.03030858
desordeSuenio	-0.109224744	0.18129567	0.70756013	0.22091260
	pasos diarios	desordeSuenio		
Genero	-0.01450937	-0.1611309		
edad	0.05797340	0.3039616		
horasSueno	-0.03953254	-0.3820452		
calidadsueno	0.01679141	-0.3574770		
nivActividadFisica	0.77272305	-0.1092247		
estrés	0.18682895	0.1812957		
CategoriaIMC	-0.14278493	0.7075601		
ritmocardiaaco	-0.03030858	0.2209126		
pasos diarios	1.00000000	-0.1612745		
desordeSuenio	-0.16127452	1.0000000		

Se puede observar que hay algunas variables que presentan una correlación mayor a 0.5 y menor a -0.5, sin embargo, se procederá a seguir con el modelo.

## Aplicación De La Técnica En El Software SPSS

Anteriormente ya se especificaron las variables que se utilizaran en este análisis, sin embargo, es importante conocer la forma en que las variables están codificadas y lo que significan cada variable.

No	Variable	Descripción	Codificación
1	Género	Sexo de la persona	1 = Masculino, 2 = Femenino
2	Edad	La edad de la persona en años	
3	Horas de sueño	La cantidad de horas que la persona duerme por día	
4	Calidad del sueño	Una calificación subjetiva de la calidad del sueño	escala: 1-10

No	Variable	Descripción	Codificación
5	Nivel de actividad física	La cantidad de minutos que la persona realiza actividad física diariamente	escala: 1-10
6	Estrés	Una calificación del nivel de estrés experimentado por la persona	
7	Categoría de IMC	La categoría de IMC de la persona	1 = Peso normal, 2 = Sobrepeso, 3 = Obesidad
8	Ritmo cardíaco	La frecuencia cardíaca en reposo de la persona en latidos por minuto	
9	Pasos diarios	La cantidad de pasos que la persona da por día	
10	Desorden del sueño	La presencia o ausencia de un trastorno del sueño en la persona	1 = Ninguno, 2 = Insomnio, 3 = Apnea del Sueño

Una vez teniendo claras las variables y la codificación que estas tienen se procede a la importación de los datos en el software de SPSS:

## Descripción De Los Datos

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Genero	Número	1	0		{0, Femenin...	Ninguna	12	Derecha	Nominal	Entrada
2	edad	Número	2	0		Ninguna	Ninguna	12	Derecha	Escala	Entrada
3	horasSueno	Número	3	1		Ninguna	Ninguna	12	Derecha	Escala	Entrada
4	calidadsueno	Número	1	0		Ninguna	Ninguna	12	Derecha	Nominal	Entrada
5	nivActividadFi...	Número	2	0		Ninguna	Ninguna	12	Derecha	Escala	Entrada
6	estrés	Número	1	0		Ninguna	Ninguna	12	Derecha	Nominal	Entrada
7	CategoríaMC	Número	1	0		{1, Normal}...	Ninguna	12	Derecha	Nominal	Entrada
8	ritmocardiaa...	Número	2	0		Ninguna	Ninguna	12	Derecha	Escala	Entrada
9	pasosdianos	Número	5	0	pasos diarios	Ninguna	Ninguna	12	Derecha	Escala	Entrada
10	desordeSue...	Número	1	0		{1, Ninguno}...	Ninguna	12	Derecha	Nominal	Entrada

1: Genero	1	Visible: 1
-----------	---	------------

	Genero	edad	horasSueno	calidadsueno	nivActividadFisica	estrés	CategoríaMC	ritmocardiaco	pasosdianos	desordeSueno
1	Masculino	27	6.1	6	42	6	Exceso de peso	77	4200	Ninguno
2	Masculino	28	6.2	6	60	8	Normal	75	10000	Ninguno
3	Masculino	28	6.2	6	60	8	Normal	75	10000	Ninguno
4	Masculino	28	5.9	4	30	8	Obeso	85	3000	Insomnio
5	Masculino	28	5.9	4	30	8	Obeso	85	3000	Insomnio
6	Masculino	28	5.9	4	30	8	Obeso	85	3000	Apnea del Sueño
7	Masculino	29	6.3	6	40	7	Obeso	82	3500	Apnea del Sueño
8	Masculino	29	7.8	7	75	6	Normal	70	8000	Ninguno
9	Masculino	29	7.8	7	75	6	Normal	70	8000	Ninguno
10	Masculino	29	7.8	7	75	6	Normal	70	8000	Ninguno
11	Masculino	29	6.1	6	30	8	Normal	70	8000	Ninguno
12	Masculino	29	7.8	7	75	6	Normal	70	8000	Ninguno
13	Masculino	29	6.1	6	30	8	Normal	70	8000	Ninguno
14	Masculino	29	6.0	6	30	8	Normal	70	8000	Ninguno
15	Masculino	29	6.0	6	30	8	Normal	70	8000	Ninguno
16	Masculino	29	6.0	6	30	8	Normal	70	8000	Ninguno
17	Femenino	29	6.5	5	40	7	Normal	80	4000	Insomnio
18	Masculino	29	6.0	6	30	8	Normal	70	8000	Insomnio
19	Femenino	29	6.5	5	40	7	Normal	80	4000	Apnea del Sueño
20	Masculino	30	7.6	7	75	6	Normal	70	8000	Ninguno
21	Masculino	30	7.7	7	75	6	Normal	70	8000	Ninguno
22	Masculino	30	7.7	7	75	6	Normal	70	8000	Ninguno

Para realizar el análisis de regresión logístico multinomial en este software se debe seleccionar en la parte superior la opción Analizar y a continuación seleccionar la opción de Regresión y por último se selecciona la opción Logística Multinomial. A continuación, se seleccionan las variables independientes y la variable de pendiente, en este caso sería de la siguiente forma:

### Variable dependiente:

1. Desorden del sueño

### Variables independientes:

1. Genero
2. Edad
3. Horas de sueño
4. Calidad del sueño
5. Nivel de actividad física

6. Estrés
7. Índice de masa corporal
8. Ritmo cardíaco
9. Pasos diarios

Se desea investigar si las variables como género, horas de sueño, actividad física, etc. Tienen efecto en poseer algún tipo de desorden de sueño, en este caso esta variable tiene 3 categorías y se toma como referencia la primera categoría (ningún desorden de sueño). Además, no se incluirán las interacciones entre variables en el modelo. Una vez realizado esto se procede aceptar y analizar los resultados obtenidos.

## Resumen De Procesamiento De Casos

La tabla de resumen de procesamiento de casos es importante en un análisis de regresión logística multinomial porque proporciona información sobre el número de casos que se han utilizado en el análisis, incluyendo el número de casos que se han excluido debido a datos faltantes o valores atípicos.

Resumen de procesamiento de casos			
		N	Porcentaje marginal
desordeSuenio	Ninguno	219	58.6%
	Insomnio	78	20.9%
	Apnea del Sueño	77	20.6%
Válidos		374	100.0%
Perdidos		0	
Total		374	
Subpoblación		106 <sup>a</sup>	

a. La variable dependiente sólo tiene un valor observado en 84 (79.2%) subpoblaciones.

## Ajuste Del Modelo

La tabla de Información de ajuste de los modelos en SPSS es importante en un análisis de regresión logística multinomial porque proporciona información sobre la calidad del ajuste del modelo a los datos. Esta tabla incluye estadísticas como el Chi-cuadrado de Pearson, el logaritmo de la verosimilitud y el criterio de información de Akaike (AIC), que se utilizan

para evaluar la bondad de ajuste del modelo. Estas estadísticas permiten comparar diferentes modelos y seleccionar el modelo que mejor se ajuste a los datos.

<b>Información de ajuste de los modelos</b>						
Modelo	Criterios de ajuste de modelo			Pruebas de la razón de verosimilitud		
	AIC	normalizado	Logaritmo de la verosimilitud -2	Chi-cuadrado	gl	Sig.
Sólo intersección	640.754	648.602	636.754			
Final	237.266	315.751	197.266	439.488	18	.000

Se presenta la tabla donde realiza la prueba de razón de verosimilitud, en donde resulta ser significativa por lo que algún coeficiente de los regresores es significativamente distinto de cero, es decir que influye en la predicción de la variable dependiente.

## Prueba De Bondad De Ajuste Del Modelo

Las pruebas de bondad de ajuste del modelo y Pseudo R cuadrado son importantes en un análisis de regresión logística multinomial porque proporciona información sobre la bondad de ajuste del modelo a los datos. Las hipótesis de las pruebas de bondad de ajuste en el análisis de regresión logística multinomial son las siguientes:

$H_0$  :El modelo ajustado se ajusta bien a los datos.

$H_1$  :El modelo ajustado no se ajusta bien a los datos.

<b>Bondad de ajuste</b>			
	Chi-cuadrado	gl	Sig.
Pearson	486.939	192	.000
Desvianza	154.168	192	.979

Los resultados de la prueba de bondad de ajuste indican que el modelo no se ajusta bien a los datos. El valor Chi-cuadrado de Pearson es alto (486.939) y el valor de desviación también es alto (154.168), lo que indica que las probabilidades predichas se desvían significativamente de las probabilidades observadas. Además, el valor p para la prueba de bondad de ajuste es muy bajo (0.000), lo que indica que el modelo no se ajusta bien a los datos y que las probabilidades predichas se desvían significativamente de las probabilidades observadas.

## Evaluación Del Ajuste Del Modelo

El Pseudo R cuadrado es una medida de la bondad de ajuste del modelo, que indica la proporción de la variabilidad en la variable dependiente que es explicada por el modelo. Cuanto mayor sea el valor del Pseudo R cuadrado, mejor será el ajuste del modelo a los datos.

Pseudo R cuadrado	
Cox y Snell	.691
Nagelkerke	.808
McFadden	.608

Se presentan tres medidas de Pseudo R cuadrado: Cox y Snell, Nagelkerke y McFadden. Los valores de estas medidas oscilan entre 0 y 1, y cuanto más cercanos a 1 sean, mejor será el ajuste del modelo. En este caso, tomando a McFadden observamos que el modelo explica un 60% del resultado de la presencia o ausencia de un trastorno de sueño.

## Pruebas De La Razón De Verosimilitud

La prueba de razón de verosimilitud es importante en un análisis de regresión logística multinomial en SPSS porque se utiliza para evaluar la contribución de cada efecto al modelo. Para cada efecto, se calcula el logaritmo de la verosimilitud -2 y se compara con el logaritmo de la verosimilitud -2 del modelo sin ese efecto. Si la diferencia entre los dos logaritmos de la verosimilitud -2 es significativa, esto indica que el efecto tiene una contribución significativa al modelo. Las hipótesis de las pruebas de la razón de verosimilitud en el análisis de regresión logística multinomial son las siguientes:

$H_0$ : El efecto no tiene una contribución significativa al modelo.  $H_1$ : El efecto tiene una contribución significativa al modelo.

Pruebas de la razón de verosimilitud						
Efecto	Criterios de ajuste de modelo			Pruebas de la razón de verosimilitud		
	AIC de modelo reducido	BIC de modelo reducido	Logaritmo de la verosimilitud -2 de modelo reducido	Chi-cuadrado	gl	Sig.
Intersección	257.078	327.715	221.078	23.812	2	.000
Genero	244.195	314.832	208.195	10.929	2	.004
edad	246.070	316.707	210.070	12.804	2	.002
horasSueno	245.813	316.450	209.813	12.547	2	.002
calidadsueno	244.570	315.207	208.570	11.304	2	.004
nivActividadFisica	234.454	305.091	198.454	1.188	2	.552
estrés	234.481	305.118	198.481	1.215	2	.545
CategorialMC	247.391	318.028	211.391	14.125	2	.001
ritmocardiaaco	239.243	309.880	203.243	5.977	2	.050
pasos diarios	239.517	310.153	203.517	6.250	2	.044

El estadístico de chi-cuadrado es la diferencia de la log-verosimilitud -2 entre el modelo final y el modelo reducido.

El modelo reducido se forma omitiendo un efecto del modelo final. La hipótesis nula es que todos los parámetros de dicho efecto son 0.

De acuerdo con las hipótesis planteadas y los resultados obtenidos se puede decir que entre los predictores importantes están las variables género, edad, horas de sueño, calidad del sueño y categoría de peso, a continuación, se observan cuanto las variables independientes predicen la dependiente.

## Estimación De Los Parámetros

La estimación de los parámetros permite conocer los coeficientes de los predictores y su contribución al modelo. Los resultados se utilizan para interpretar la relación entre los predictores y la variable dependiente categórica.

		Estimaciones de parámetro							95% de intervalo de confianza para Exp(B)	
			Desv.						Límite inferior	Límite superior
desordeSuenio <sup>a</sup>	B	Error	Wald	gl	Sig.	Exp(B)				
Insomnio	Intersección	-33.553	13.329	6.337	1	.012				
	Genero	-2.106	.905	5.411	1	.020	.122	.021	.718	
	edad	.106	.060	3.140	1	.076	1.112	.989	1.250	
	horasSueno	2.739	1.006	7.419	1	.006	15.472	2.156	111.037	
	calidadsueno	-2.260	.818	7.630	1	.006	.104	.021	.519	
	nivActividadFisica	.005	.036	.020	1	.888	1.005	.936	1.079	
	estrés	-.485	.522	.863	1	.353	.616	.221	1.713	
	CategorialM C	2.950	.844	12.222	1	.000	19.115	3.656	99.940	
	ritmocardiaco	.263	.148	3.138	1	.076	1.301	.972	1.740	
	pasos diarios	.001	.000	2.101	1	.147	1.001	1.000	1.002	
Apnea del Sueño	Intersección	24.406	9.544	6.540	1	.011				
	Genero	.745	.770	.935	1	.333	2.106	.466	9.523	
	edad	.225	.067	11.146	1	.001	1.252	1.097	1.429	
	horasSueno	-1.450	1.031	1.976	1	.160	.235	.031	1.771	
	calidadsueno	-1.767	.650	7.394	1	.007	.171	.048	.611	
	nivActividadFisica	.039	.036	1.167	1	.280	1.040	.969	1.116	
	estrés	-.486	.518	.880	1	.348	.615	.223	1.699	
	CategorialM C	1.639	.793	4.271	1	.039	5.151	1.088	24.377	
	ritmocardiaco	-.137	.144	.905	1	.341	.872	.658	1.156	
	pasos diarios	-.001	.000	2.719	1	.099	.999	.998	1.000	

a. La categoría de referencia es: Ninguno.

La Salida anterior se hace la comparación en la primera sección de la tabla que la persona no tiene ningún desorden de sueño frente a que posee Apnea de sueño o Insomnio. El primer predictor es las horas de sueño, puesto que esta predice si la persona posee un trastorno o no de sueño. Con valor de Wald 7.419, el cual asigna la contribución individual de los regresaros a la variable dependiente, y además se observa que es estadísticamente significativo (0.006), luego su respectivo odd ratio (columna Exp(B)) es de 15.47 esto significa que a medida que aumenta las horas de sueño en una unidad, el cambio de probabilidad de una persona de tener insomnio en lugar de no tener ningún trastorno es de 15,47 veces más de probable, en otras palabras es más probable que se desarrolle insomnio si aumenta las horas de sueño. El siguiente predictor es la calidad de sueño, con valor de Wald de 7,63 y su p valor menor al de significancia, es decir que a medida que la calidad de sueno aumenta en una unidad, las probabilidades de tener insomnio en comparación de no tener ningún desorden son de 0.10, por lo que disminuye la probabilidad de tener un desorden de sueño. El tercer predictor es



categoríaIMC es decir el peso de la persona, se observa que es significativo en la predicción de la variable dependiente. Esto significa que si la persona aumenta de peso tiene 19,11 veces más de probabilidad de tener insomnio. Finalmente, el género de la persona, a medida que cambia de femenino a masculino la probabilidad de tener insomnio es de 0.12 veces más probable. En la segunda sección se compara apnea del sueño con ningún desorden de sueño, la edad resulta significativa en la predicción de la variable dependiente. Se observa que su p valor es de 0,01 que es menor al de significancia y su valor de Wald de 11,46. Entonces, a medida que aumenta la edad aumenta 1.252 veces la probabilidad de tener apnea en comparación de no tener ningún tipo de desorden. De la misma manera la calidad de Sueño, si disminuye es más probable que posea apnea, es decir que si a calidad de sueño aumenta en una unidad la probabilidad de que la persona tenga apnea es 0.171 veces. Finalmente, categoríaIMC si aumenta de peso en una unidad es 5.15 veces más probable que desarrolle apnea a que no posea ningún trastorno de sueño.

## Clasificación

Clasificación				
Observado	Pronosticado			Porcentaje correcto
	Ninguno	Insomnio	Apnea del Sueño	
Ninguno	204	10	5	93.2%
Insomnio	7	67	4	85.9%
Apnea del Sueño	10	3	64	83.1%
Porcentaje global	59.1%	21.4%	19.5%	89.6%

La tabla de clasificación muestra el número de observaciones que fueron clasificadas correctamente y el número de observaciones que fueron clasificadas incorrectamente para cada una de las categorías de la variable dependiente. En este caso, la tabla de clasificación muestra que el modelo clasificó correctamente 204 observaciones en la categoría 1, 67 observaciones en la categoría 2 y 64 observaciones en la categoría 3.

Además, el modelo clasificó incorrectamente 10 observaciones de la categoría 1 como categoría 2, 5 observaciones de la categoría 1 como categoría 3, 7 observaciones de la categoría 2 como categoría 1, 4 observaciones de la categoría 2 como categoría 3, 10 observaciones de la categoría 3 como categoría 1 y 3 observaciones de la categoría 3 como categoría 2.

La precisión del modelo es del 89.6% en la predicción de las categorías de la variable dependiente. Esto significa que el modelo clasificó correctamente el 89.6% de las observaciones en

los datos de entrenamiento. La precisión del modelo se utiliza para evaluar la capacidad del modelo para predecir la variable dependiente y para comparar diferentes modelos.

## Aplicación de la Técnica En El Software RStudio

### Librerías que se utilizarán

```
library(dplyr)
require(nnet)
```

Loading required package: nnet

```
library(pscl)
```

Classes and Methods for R developed in the  
Political Science Computational Laboratory  
Department of Political Science  
Stanford University  
Simon Jackman  
hurdle and zeroinfl functions by Achim Zeileis

Antes de aplicar `multinom()` se debe de tener los nombres de las columnas de la base de datos de forma adecuada, es decir, estos nombres no deben de tener espacios, por lo que hay que proceder a cambiar los nombres que generan problemas. En este caso unicamente la variable "pasos diarios" contiene espacios en su nombre.

```
# Cambiar el nombre de la columna "pasos diarios" a "pasos_diarios"
colnames(base_suenio)[colnames(base_suenio) == "pasos diarios"] <- "pasos_diarios"
```

Ahora se aplica `multinom()`.

La función `multinom()` se utiliza para ajustar un modelo de regresión logística multinomial a un conjunto de datos y puede tomar una fórmula que especifica la variable dependiente (`desordeSuenio`) y las variables independientes. El resultado de `multinom()` es un objeto de modelo que se puede utilizar para hacer predicciones y realizar otras operaciones en el modelo ajustado.

```
multinom.fit <- multinom(desordeSuenio ~
                          Genero +
```

```

edad +
horasSueno +
calidadsueno +
nivActividadFisica +
estrés +
CategoriaIMC +
ritmocardiaaco +
pasos_diarios
-1,
data = base_suenio)

```

```

# weights: 30 (18 variable)
initial value 410.880996
iter 10 value 239.990329
iter 20 value 153.790712
iter 30 value 153.330701
final value 153.330686
converged

```

El resultado que se muestra indica que el modelo ha convergido, lo que significa que el algoritmo utilizado para ajustar el modelo ha encontrado una solución que satisface los criterios de convergencia. El valor inicial y final que se muestran son los valores de la función de verosimilitud del modelo antes y después del ajuste, respectivamente. En este caso, el valor final es menor que el valor inicial, lo que indica que el modelo ajustado es mejor que el modelo inicial.

```
summary(multinom.fit)
```

Call:

```

multinom(formula = desordeSuenio ~ Genero + edad + horasSueno +
          calidadsueno + nivActividadFisica + estrés + CategoriaIMC +
          ritmocardiaaco + pasos_diarios - 1, data = base_suenio)

```

Coefficients:

	Genero	edad	horasSueno	calidadsueno	nivActividadFisica	estrés
2	-2.434412	0.1369599	2.206757	-3.24757	0.11683365	-0.3317447
3	0.864164	0.2084247	-1.272735	-1.10896	-0.04852001	-0.7161031

	CategoriaIMC	ritmocardiaaco	pasos_diarios
2	2.705224	-0.05803355	-0.0005715981
3	2.178973	0.11213840	0.0002847141

Std. Errors:

	Genero	edad	horasSueno	calidadsueno	nivActidadFisica	estrés
2	0.001568883	0.02629733	0.03158598	0.04055293	0.02111513	0.04408552
3	0.001066615	0.03081514	0.03373374	0.05190156	0.01751033	0.06324800

	CategoriaIMC	ritmocardiaaco	pasos_diarios
2	0.006147396	0.0158694	0.0002517732
3	0.006333034	0.0173187	0.0002197325

Residual Deviance: 306.6614

AIC: 342.6614

La tabla de coeficientes muestra los valores estimados para cada variable independiente en el modelo, así como el error estándar de cada estimación. Los coeficientes se interpretan como el cambio en la log-odds de la variable dependiente por cada unidad de cambio en la variable independiente correspondiente. Por ejemplo, el coeficiente para Genero=2 da un valor de -2.434412 lo que significa que por cada aumento de una unidad en Genero=2, el logaritmo de las probabilidades de desordeSuenio estar en el nivel 2 (en comparación con el nivel de referencia) disminuye en 2.434412, manteniendo constantes todas las demás variables predictoras. De manera similar, el coeficiente para Genero=3 es 0.864164, lo que significa que por cada aumento de una unidad en Genero=3, el logaritmo de las probabilidades de desordeSuenio estar en el nivel 3 (en comparación con el nivel de referencia) aumenta en 0.864164,

El AIC es un criterio de selección de modelos que se utiliza para comparar diferentes modelos. Un valor bajo de AIC indica que el modelo es mejor ajustado a los datos. En este caso, el valor de AIC es 342.6614, lo que indica que el modelo ajustado es relativamente bueno.

## Exponentes De Los Coeficientes Estimado Del Modelo

El código `exp(coef(multinom.fit))` en R se utiliza para calcular los exponentes de los coeficientes estimados en un modelo de regresión logística multinomial. Los exponentes de los coeficientes se interpretan como la razón de las probabilidades (odds ratio) de la variable dependiente para cada unidad de cambio en la variable independiente correspondiente. Al tomar el exponencial de los coeficientes, se convierten en odds ratios, lo que facilita la interpretación de los resultados.

```
exp(coef(multinom.fit))
```

	Genero	edad	horasSueno	calidadsueno	nivActidadFisica	estrés
2	0.08764924	1.146782	9.0862050	0.03886855	1.1239324	0.7176705
3	2.37302138	1.231736	0.2800647	0.32990175	0.9526383	0.4886528

	CategoriaIMC	ritmocardiaaco	pasos_diarios
2	14.957661	0.9436183	0.9994286
3	8.837229	1.1186677	1.0002848

En el resultado que se muestra, se presentan los coeficientes exponenciales para cada variable independiente en el modelo. Por ejemplo, para el nivel 2 de la variable dependiente, un aumento de una unidad en la variable Genero disminuye las probabilidades de desordeSuenio en un 91.7%, mientras que un aumento de una unidad en la variable edad aumenta las probabilidades de desordeSuenio en un 14.7% . Un aumento de una unidad en la variable horasSueno aumenta las probabilidades de desordeSuenio en un 809% , mientras que un aumento de una unidad en la variable calidadsueno disminuye las probabilidades de desordeSuenio en un 96.1% , y asi sucesivamente con los resultados para las demás variables.

## Probabilidades Ajustadas Para Cada Observación En El Modelo

El cálculo de las probabilidades ajustadas en un análisis de regresión logística multinomial es importante porque permite estimar la probabilidad de que una observación pertenezca a cada una de las categorías de la variable dependiente, dadas las variables independientes, lo que es útil para evaluar la precisión del modelo, hacer predicciones y tomar decisiones.

```
prob_ajust <- head(probability.table <- fitted(multinom.fit))
prob_ajust
```

	1	2	3
1	0.508278253	0.0186907883	0.47303096
2	0.954153280	0.0005772689	0.04526945
3	0.954153280	0.0005772689	0.04526945
4	0.007523527	0.3196301976	0.67284628
5	0.007523527	0.3196301976	0.67284628
6	0.007523527	0.3196301976	0.67284628

Para cada observación, se muestran las probabilidades ajustadas de pertenecer a cada una de las tres categorías de la variable dependiente (desordeSuenio). Por ejemplo, la primera observación tiene una probabilidad ajustada del 50.8% de pertenecer a la categoría 1, una probabilidad ajustada del 1.9% de pertenecer a la categoría 2 y una probabilidad ajustada del 47.3% de pertenecer a la categoría 3.

## Predicción Y Precisión Del Modelo

Ahora se procederá a predecir las categorías de la variable dependiente utilizando el modelo de regresión logística multinomial ajustado, crear una tabla de clasificación y calcular la precisión de las predicciones del modelo. Esto es útil para evaluar la precisión del modelo y para hacer predicciones precisas sobre la variable dependiente.

```
# Predecir
base_suenio$predicted <- predict(multinom.fit, newdata = base_suenio, "class")
```

```
# Tabla de clasificacion
ctable <- table(base_suenio$desordeSuenio, base_suenio$predicted)
```

```
ctable
```

	1	2	3
1	204	8	7
2	6	68	4
3	8	4	65

```
# Precisión de cálculo: suma de elementos diagonales dividida por el total de observaciones
round((sum(diag(ctable))/sum(ctable))*100,2)
```

```
[1] 90.11
```

El resultado muestra la tabla de clasificación de las predicciones realizadas en un modelo de regresión logística multinomial.

La tabla de clasificación muestra el número de observaciones que fueron clasificadas correctamente y el número de observaciones que fueron clasificadas incorrectamente para cada una de las categorías de la variable dependiente. En este caso, la tabla de clasificación muestra que el modelo clasificó correctamente 204 observaciones en la categoría 1, 68 observaciones en la categoría 2 y 65 observaciones en la categoría 3.

Además, el modelo clasificó incorrectamente 8 observaciones de la categoría 1 como categoría 2, 7 observaciones de la categoría 1 como categoría 3, 6 observaciones de la categoría 2 como categoría 1, 4 observaciones de la categoría 2 como categoría 3, 8 observaciones de la categoría 3 como categoría 1 y 4 observaciones de la categoría 3 como categoría 2.

La precisión del modelo es del 90.11% en la predicción de las categorías de la variable dependiente. Esto significa que el modelo clasificó correctamente el 90.11% de las observaciones en los datos de entrenamiento. La precisión del modelo se utiliza para evaluar la capacidad del modelo para predecir la variable dependiente y para comparar diferentes modelos.

## Aplicación De La Técnica En El Software Python

```
library(reticulate)
```

```
import pandas as pd
import numpy as np
```

```
# cargar la base de datos
suenio_py = pd.read_excel('basevida0s.xlsx')
columns_names = suenio_py.columns.values
columns_names
```

```
array(['Genero', 'edad', 'horasSueno', 'calidadsueno', 'nivActividadFisica',
      'estrés', 'CategoriaIMC', 'presion ', 'ritmocardiaaco',
      'pasos diarios ', 'desordeSuenio'], dtype=object)
```

```
suenio = suenio_py.drop(columns = 'presion ')
columns_names = suenio.columns.values
columns_names
```

```
array(['Genero', 'edad', 'horasSueno', 'calidadsueno', 'nivActividadFisica',
      'estrés', 'CategoriaIMC', 'ritmocardiaaco', 'pasos diarios ',
      'desordeSuenio'], dtype=object)
```

```
suenio.head()
```

	Genero	edad	horasSueno	...	ritmocardiaaco	pasos diarios	desordeSuenio
0	1	27	6.1	...	77	4200	1
1	1	28	6.2	...	75	10000	1
2	1	28	6.2	...	75	10000	1
3	1	28	5.9	...	85	3000	2
4	1	28	5.9	...	85	3000	2

```
[5 rows x 10 columns]
```

```
print("Número de observaciones por clase")
```

Número de observaciones por clase

```
print(suenio['desordeSuenio'].value_counts())
```

```
desordeSuenio
```

```
1      219
```

```
2       78
```

```
3       77
```

```
Name: count, dtype: int64
```

```
print("")
```

```
print("Porcentaje de observaciones por clase")
```

```
Porcentaje de observaciones por clase
```

```
print(100 * suenio['desordeSuenio'].value_counts(normalize=True))
```

```
desordeSuenio
```

```
1      58.556150
```

```
2      20.855615
```

```
3      20.588235
```

```
Name: proportion, dtype: float64
```

Las categorías que tiene esta variable son 1 = Ninguno, 2 = Insomnio, 3 = Apnea del Sueño. Esto quiere decir que el 58.56% de las personas no presentan un trastorno del sueño, mientras que el 20.86% presenta Insomnio y el 20.59% presenta Apnea del Sueño.

## Ajuste Del Modelo

Se ajusta un modelo de regresión logística múltiple con el objetivo de predecir si una persona presenta un desorden del sueño en función de todas las variables disponibles.

```
X = suenio.drop(columns = 'desordeSuenio')
```

```
y = suenio['desordeSuenio']
```

```
from sklearn.model_selection import train_test_split
```



```
X_train, X_test, y_train, y_test = train_test_split(
    X,
    y.values.reshape(-1,1),
    train_size = 0.8,
    random_state = 1234,
    shuffle = True
)
```

```
import statsmodels.api as sm
```

```
df = suenio_py.drop(columns = 'presion ')
df.head()
```

	Genero	edad	horasSueno	...	ritmocardiaaco	pasos diarios	desordeSuenio
0	1	27	6.1	...	77	4200	1
1	1	28	6.2	...	75	10000	1
2	1	28	6.2	...	75	10000	1
3	1	28	5.9	...	85	3000	2
4	1	28	5.9	...	85	3000	2

[5 rows x 10 columns]

```
df = df.rename(columns={'pasos diarios ': 'pasos_diarios'})
```

```
# Seleccionar las columnas correspondientes
```

```
X = df[['Genero', 'edad', 'horasSueno', 'calidadsueno', 'nivActividadFisica', 'estrés', 'Cat
y = df['desordeSuenio']
```

```
# Ajustar el modelo de regresión logística multinomial
```

```
model = sm.MNLogit(y, sm.add_constant(X))
```

```
result = model.fit()
```

Optimization terminated successfully.

Current function value: 0.378141

Iterations 8

```
# Imprimir un resumen de los resultados
```

```
print(result.summary())
```

# MNLogit Regression Results

Dep. Variable:	desordeSuenio	No. Observations:	374
Model:	MNLogit	Df Residuals:	354
Method:	MLE	Df Model:	18
Date:	lun., 02 oct. 2023	Pseudo R-squ.:	0.6084
Time:	22:54:12	Log-Likelihood:	-141.42
converged:	True	LL-Null:	-361.17
Covariance Type:	nonrobust	LLR p-value:	5.156e-82

desordeSuenio=2	coef	std err	z	P> z	[0.025	0.975]
const	-33.5526	13.329	-2.517	0.012	-59.676	-7.429
Genero	-2.1062	0.905	-2.326	0.020	-3.881	-0.332
edad	0.1060	0.060	1.772	0.076	-0.011	0.223
horasSueno	2.7390	1.006	2.724	0.006	0.768	4.710
calidadsueno	-2.2604	0.818	-2.762	0.006	-3.864	-0.656
nivActividadFisica	0.0051	0.036	0.141	0.888	-0.066	0.076
estrés	-0.4847	0.522	-0.929	0.353	-1.508	0.538
CategoriaIMC	2.9505	0.844	3.496	0.000	1.296	4.605
ritmocardiaaco	0.2628	0.148	1.771	0.076	-0.028	0.554
pasos_diarios	0.0007	0.000	1.450	0.147	-0.000	0.002

desordeSuenio=3	coef	std err	z	P> z	[0.025	0.975]
const	24.4057	9.544	2.557	0.011	5.700	43.111
Genero	0.7446	0.770	0.967	0.333	-0.764	2.254
edad	0.2249	0.067	3.339	0.001	0.093	0.357
horasSueno	-1.4496	1.031	-1.406	0.160	-3.471	0.572
calidadsueno	-1.7669	0.650	-2.719	0.007	-3.040	-0.493
nivActividadFisica	0.0390	0.036	1.080	0.280	-0.032	0.110
estrés	-0.4864	0.518	-0.938	0.348	-1.503	0.530
CategoriaIMC	1.6392	0.793	2.067	0.039	0.085	3.194
ritmocardiaaco	-0.1367	0.144	-0.951	0.341	-0.418	0.145
pasos_diarios	-0.0008	0.000	-1.649	0.099	-0.002	0.000

El Pseudo R-cuadrado es de 0.6084, lo que indica que el modelo explica el 60.84% de la varianza en la variable dependiente.

Además el coeficiente representa el logaritmo de la razón de probabilidad (odds) de que un individuo esté en la categoría "desordeSuenio=2" en comparación con la categoría de referencia (normalmente, la categoría 1) cuando todas las demás variables independientes son cero. En

este caso, si el valor de para la variable Genero es de -2.260, esto significa que un aumento de una unidad en la variable se asocia con una disminución de 2.260 unidades en el logaritmo de la razón de posibilidades de pertenecer a la categoría correspondiente, manteniendo todas las demás variables constantes.

```
from sklearn.model_selection import train_test_split

X = df[['Genero', 'edad', 'horasSueno', 'calidadsueno', 'nivActividadFisica', 'estrés', 'Cat
y = df['desordeSuenio']

# Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Ajustar el modelo de regresión logística multinomial
model = sm.MNLogit(y_train, sm.add_constant(X_train))
result = model.fit()
```

Optimization terminated successfully.  
Current function value: 0.373216  
Iterations 8

```
# Hacer predicciones con los datos de prueba
y_pred = result.predict(sm.add_constant(X_test))

print(y_pred)
```

	0	1	2
329	0.911404	0.051939	0.036657
33	0.857915	0.000266	0.141818
15	0.986661	0.001985	0.011354
325	0.911404	0.051939	0.036657
57	0.797282	0.000193	0.202525
..	...	...	...
94	0.981759	0.012761	0.005480
195	0.102012	0.025943	0.872044
312	0.928958	0.036238	0.034804
292	0.031045	0.939128	0.029826
372	0.130240	0.771293	0.098467

[75 rows x 3 columns]

La primera fila de la matriz muestra las probabilidades de clase para la observación número 329 en los datos de prueba. La probabilidad de pertenecer a la primera clase es 0.911404, la probabilidad de pertenecer a la segunda clase es 0.051939, y la probabilidad de pertenecer a la tercera clase es 0.036657. Esto significa que el modelo predice que la observación número 329 tiene una alta probabilidad de pertenecer a la primera clase y una baja probabilidad de pertenecer a la segunda y tercera clase.