

PROYECTO DE ANÁLISIS DE DATOS I

Maestría en Estadística Aplicada – Cohorte 2020

Leiza Camilo Caro – María Eugenia Martinez – Martín Rodriguez

Análisis de Caso 1

Un estudio transversal fue diseñado con el objetivo de estimar las prevalencias de diabetes mellitus (DM) en la población de Telde (Gran Canaria) e identificar los factores asociados con ella. En el estudio se incluyeron 1030 sujetos con edades comprendidas entre los 30 y 82 años, que respondieron a un cuestionario diseñado para obtener información sobre algunas características personales: edad, sexo, historia personal y familiar de diabetes mellitus y estilos de vida. Además, se realizaron mediciones antropométricas y de presión arterial y, con muestras de sangre, se obtuvieron mediciones de glucemia, lípidos, marcadores de inflamación, hemostasia, etc.

- EDAD: está expresada en años
- SEXO (codificada con 0 para Hombres y 1 para Mujeres)
- PESO: expresado en kilogramos
- TALLA: expresada en cm
- DM: Indicadora de presencia de Diabetes Mellitus
- SEDENT: indicadora de ser o no sedentario
- TAS: Presión arterial Sistólica;
- TAD: Presión arterial Diastólica;
- HTA_OMS: indicadora de si se Padece de Hipertensión arterial (según lo definido por la OMS).
- ECV_B: indicadora de padecer Enfermedades Cardiovasculares
- TABACO indicadora de ser fumador
- ALCOHOL indicadora de beber alcohol regularmente
- STATIN: indicadora de consumo de Estanina (medicamento indicado para bajar colesterol y triglicéridos)
- OBCENT_ATP: indicadora de Obesidad Central.
- CINTURA: perímetro expresado en cm.
- CADERA: perímetro expresado en cm.
- COLESTEROL: colesterol total (en md/dL)

PARA LA SITUACIÓN 1: DOS VARIABLES BINARIAS. ESCRIBE BREVEMENTE EL OBJETIVO QUE PLANTEA USANDO EL MODELO INDICADO.

Se tiene como objetivo el evaluar la relación existente entre la posibilidad de padecer hipertensión arterial (HTA_OMS) de acuerdo con si la persona es fumadora o no y si bebe alcohol o no.

REALIZA UN ANÁLISIS DESCRIPTIVO DE LA INFORMACIÓN QUE INCLUYA TABLAS, GRÁFICOS Y SUS INTERPRETACIONES RELACIONADAS AL PROBLEMA.

Al analizar los datos obtenidos de la encuesta a personas de 30 a 82 años, se obtiene lo siguiente:

Tabla 1: Distribución de 80 personas encuestadas entre 30 y 82 años de acuerdo si presentan o no hipertensión arterial y teniendo en cuenta si son fumadores y bebedores de alcohol.

TABACO	ALCOHOL	HIRTENSIÓN ARTERIAL		TOTAL
		NO	SI	
NO	NO	30	15	45
	SI	12	8	20
SI	NO	7	2	9
	SI	5	1	6
TOTAL		54	26	80

Tabla 2: Distribución en proporción respecto a las 80 personas encuestadas entre 30 y 82 años de acuerdo si presentan o no hipertensión arterial y teniendo en cuenta si son fumadores y bebedores de alcohol.

TABACO	ALCOHOL	HIRTENSIÓN ARTERIAL		TOTAL
		NO	SI	
NO	NO	0,38	0,19	0,56
	SI	0,15	0,10	0,25
SI	NO	0,09	0,03	0,11
	SI	0,06	0,01	0,08
TOTAL		0,68	0,33	1,00

Teniendo en cuenta la información obtenida podemos decir que, en la muestra de 80 encuestas, la mayor proporción de los encuestados no presentan HTA y que de ellos el 55% no fuma y no consume alcohol. En el total de los encuestados esto representa el 38% y es el mayor porcentaje. La mayoría de los encuestados no fuman (81%) y de ellos sólo el 32% es hipertenso y ese porcentaje disminuye aún más (21%) en aquellos que si fuman. El comportamiento que se observa en los que no consumen alcohol, que son el 67% de los encuestados, de ellos el 28% sólo es hipertenso y de los que consumen alcohol el 35% lo es. Esto podría estar indicando una asociación o dependencia de la presencia de HTA con el consumo de tabaco y/o alcohol.

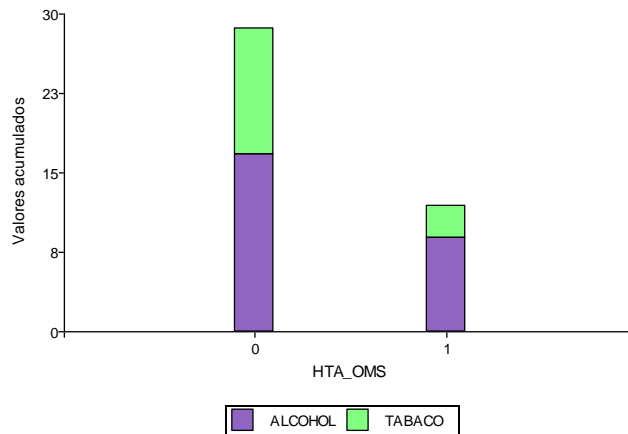


Figura 1: Distribución de personas encuestadas entre 30 y 82 años que no presentan (indicador "0") o presentan (indicador "1") hipertensión arterial (HTA_OMS)

ESCRIBE EL MODELO Y LAS HIPÓTESIS A PROBAR (ESTADÍSTICAMENTE Y EN PALABRAS EN FUNCIÓN DEL PROBLEMA).

1º) Modelo

$$\text{logit}(p) = \beta_0 + \beta_1 * x + \beta_2 * z$$

$$p = \frac{\exp(\beta_0 + \beta_1 * x + \beta_2 * z)}{1 + \exp(\beta_0 + \beta_1 * x + \beta_2 * z)}$$

donde:

p: es la probabilidad de que la persona presente hipertensión arterial (HTA)

β_0 : tasa de cambio de la probabilidad de presentar HTA cuando no se presenta tabaquismo ni consumo de alcohol.

β_1 : tasa de cambio de la probabilidad de presentar HTA cuando se presenta tabaquismo y consumo de alcohol está ausente.

x : indicadora de tabaquismo (0=ausencia; 1=presencia)

β_2 : tasa de cambio de la probabilidad de presentar HTA cuando se presenta consumo de alcohol y el tabaquismo está ausente.

z : indicadora de consumo de alcohol (0=ausencia; 1=presencia)

2º) Planteo de hipótesis

$$\text{A) } \begin{cases} H_0: \beta_0 = 0 \\ H_1: \beta_0 \neq 0 \end{cases} \quad \text{ó}$$

$\{H_0: \text{La probabilidad de presentar HTA depende sólo del tabaquismo y el consumo de alcohol}$
 $\{H_1: \text{La probabilidad de presentar HTA no sólo depende del tabaquismo y el consumo de alcohol}$

$$\text{B) } \begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases} \quad \text{ó}$$

$\{H_0: \text{La probabilidad de presentar HTA no depende del tabaquismo}$
 $\{H_1: \text{La probabilidad de presentar HTA depende del tabaquismo}$

$$\text{C) } \begin{cases} H_0: \beta_2 = 0 \\ H_1: \beta_2 \neq 0 \end{cases} \quad \text{ó}$$

$\{H_0: \text{La probabilidad de presentar HTA no depende del consumo de alcohol}$
 $\{H_1: \text{La probabilidad de presentar HTA depende del consumo de alcohol}$

D)

$\{H_0: \text{La probabilidad de presentar HTA es independiente del tabaquismo y del consumo de alcohol}$
 $\{H_1: \text{La probabilidad de presentar HTA no es independiente del tabaquismo y del consumo de alcohol}$

III) USANDO UN SOFTWARE Y MOSTRANDO LA SALIDA DE ESTE

```
> HTA_NO=c(30,12,7,5)
> HTA_SI=c(15,8,2,1)
> TABACO=c(0,0,1,1) #0=no; 1=si
> ALCOHOL=c(0,1,0,1) #0=no; 1=si
> modelo1=glm(cbind(HTA_SI,HTA_NO)~TABACO+ALCOHOL,family = binomial)
> summary(modelo1)
```

Call:

```
glm(formula = cbind(HTA_SI, HTA_NO) ~ TABACO + ALCOHOL, family = binomial)
```

Deviance Residuals:

```
1          2          3          4
```

```

-0.09812    0.14276    0.26004   -0.31751

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.6622    0.3068   -2.158   0.0309 *
TABACO       -0.8033    0.6984   -1.150   0.2501
ALCOHOL       0.1914    0.5124    0.374   0.7087
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1.74478  on 3  degrees of freedom
Residual deviance: 0.19844  on 1  degrees of freedom
AIC: 17.976

Number of Fisher Scoring iterations: 4

> fitted.values(modelo1)
      1      2      3      4
0.3402520 0.3844329 0.1876287 0.2185569
> exp(coef(modelo1))
(Intercept)    TABACO    ALCOHOL
  0.5157303  0.4478392  1.2109397
> coef(modelo1)
(Intercept)    TABACO    ALCOHOL
-0.6621713 -0.8033211  0.1913967
> 1-pchisq(0.19844,1)
[1] 0.655983

```

1. ESCRIBIR EL MODELO ESTIMADO E INTERPRETAR PARA EL PROBLEMA LOS COEFICIENTES.

$$\text{logit}(\hat{p}) = -0,6622 - 0,8033 * x + 0,1914 * z$$

$$\hat{p} = \frac{\exp(-0,6622 - 0,8033 * x + 0,1914 * z)}{1 + \exp(-0,6622 - 0,8033 * x + 0,1914 * z)}$$

\hat{p} : es la probabilidad de que la persona presente hipertensión arterial (HTA)

$\hat{\beta}_0 = -0,6622$: La probabilidad de presentar HTA disminuye cuando no se presenta tabaquismo y consumo de alcohol.

$\hat{\beta}_1 = -0,8033$: La probabilidad de presentar HTA disminuye cuando se presenta tabaquismo y consumo de alcohol está ausente.

x : indicadora de tabaquismo (0=ausencia; 1=presencia)

$\hat{\beta}_2 = 0,1914$ La probabilidad de presentar HTA aumenta cuando se presenta consumo de alcohol y el tabaquismo está ausente.

z : indicadora de consumo de alcohol (0=ausencia; 1=presencia)

2. PROBAR LA SIGNIFICACIÓN DE LOS PARÁMETROS DEL MODELO.

La muestra de encuestas realizadas a personas de 30 a 82 años, tiene suficiente evidencia con una significancia del 5% para decir que la probabilidad de presentar HTA depende no sólo del tabaquismo y el consumo de alcohol (P-valor=0,0309). En cambio, no es suficiente para decir que depende del tabaquismo o del consumo de alcohol (p-valor=0,2501 y p-valor=0,7087, respectivamente)

De acuerdo con la evaluación del modelo, se acepta que la presencia de HTA es independiente del tabaquismo y del consumo de alcohol. (p-valor=0,655983)

3. CALCULA EL ODDS RATIO (OR) E INTERPRETARLO.

$$odds_T = \frac{\hat{p}_T}{1 - \hat{p}_T} = \frac{0,1876}{0,8124} = 0,2309 \quad odds_A = \frac{\hat{p}_A}{1 - \hat{p}_A} = \frac{0,3844}{0,6156} = 0,6245$$

El $odds_T$ nos indica que existe 4 veces más probabilidad de no presentar HTA debida al tabaco que si presentarla. Y análogamente, el $odds_A$ indica que es 1,5 más probable no presentar HTA debido al consumo de alcohol.

$$OR = \frac{odds_A}{odds_T} = \frac{0,6245}{0,2309} = 2,70$$

Teniendo en cuenta los odds para consumo de alcohol y tabaquismo, la probabilidad de presentar HTA aumenta casi 3 veces si se consume alcohol respecto a los fumadores.

4. ¿A QUÉ CONCLUSIÓN LLEGA?

De acuerdo con la información obtenida, la HTA no depende del tabaquismo y del consumo de alcohol, pero el consumo de alcohol aumenta más las posibilidades de presentarla que el ser fumador.

PARA LA SITUACIÓN 2: TRES VARIABLES BINARIAS (LAS DOS DE LA SITUACIÓN 1 Y UNA MÁS).

ESCRIBE BREVEMENTE EL OBJETIVO QUE PLANTEA USANDO EL MODELO INDICADO.

Se tiene como objetivo el evaluar la relación existente entre la posibilidad de padecer hipertensión arterial (HTA_OMS) de acuerdo con si la persona es fumadora, si consume alcohol y/o si es sedentaria.

REALIZA UN ANÁLISIS DESCRIPTIVO DE LA INFORMACIÓN QUE INCLUYA TABLAS, GRÁFICOS Y SUS INTERPRETACIONES RELACIONADAS AL PROBLEMA.

Teniendo en cuenta las 80 encuestas realizadas a personas entre 30 y 82 años, se obtienen los siguientes resultados:

Tabla 3: Distribución de 80 personas encuestadas entre 30 y 82 años de acuerdo si presentan o no hipertensión arterial y teniendo en cuenta si son fumadores, si consumen alcohol y si son sedentarios.

TABACO	ALCOHOL	SEDENT	NO HTA	SI HTA	TOTAL
NO	NO	NO	19	5	24
		SI	11	10	21
	SI	NO	6	5	11
		SI	6	3	9
SI	NO	NO	3	0	3
		SI	4	2	6
	SI	NO	3	1	4
		SI	2	0	2
TOTAL			54	26	80

Se observa que la mayor cantidad de personas encuestadas (19/80) no son hipertensas, no fuman, no consumen alcohol y no son sedentarias. Y la menor cantidad de personas encuestadas (1/80) presenta HTA y es fumadora y consume alcohol. La mayoría No consumen alcohol ni fuman, sin embargo, las mayores frecuencias de personas con HTA se encuentran en esta categoría.

De los encuestados, sólo el 32% presenta HTA y son la mitad de los que no la tienen. De ellos el 38% no fuma, no consume alcohol, pero es sedentaria. Si observamos la figura 2, los que presentan HTA en su mayoría son sedentarios. En la figura 3 se observa que, paralelamente, las personas que no presentan HTA son más los fumadores que los que presentan HTA. Y respecto al sedentarismo, los que presentan HTA son más sedentarios que los que no presentan HTA. Esto podría indicar una cierta dependencia del sedentarismo en la presencia de la enfermedad.

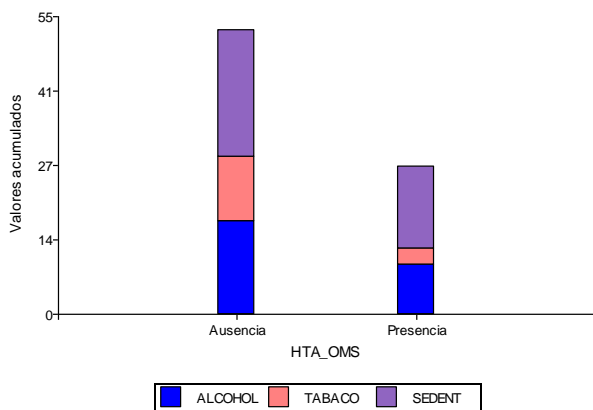


Figura 2: Distribución de personas encuestadas entre 30 y 82 años que no presentan o presentan hipertensión arterial (HTA_OMS)

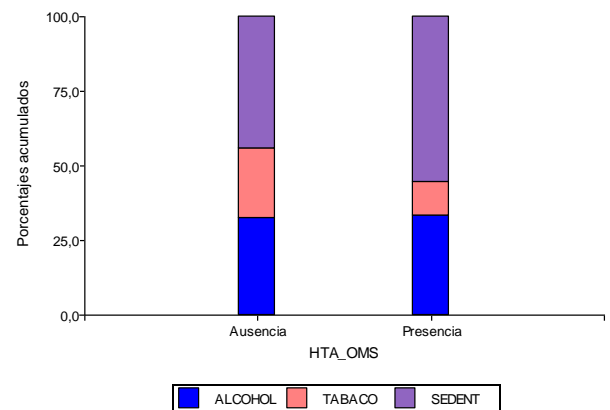


Figura 3: Distribución de porcentajes de personas encuestadas entre 30 y 82 años que no presentan o presentan hipertensión arterial (HTA_OMS), de acuerdo con el consumo de alcohol, el tabaquismo y el sedentarismo.

ESCRIBE EL MODELO Y LAS HIPÓTESIS A PROBAR (ESTADÍSTICAMENTE Y EN PALABRAS EN FUNCIÓN DEL PROBLEMA).

1º) Modelo

$$\text{logit}(p) = \beta_0 + \beta_1 * x + \beta_2 * z + \beta_3 * w$$

$$p = \frac{\exp(\beta_0 + \beta_1 * x + \beta_2 * z + \beta_3 * w)}{1 + \exp(\beta_0 + \beta_1 * x + \beta_2 * z + \beta_3 * w)}$$

donde:

p: es la probabilidad de que la persona presente hipertensión arterial (HTA)

β_0 : tasa de cambio de la probabilidad de presentar HTA cuando no se presenta tabaquismo ni consumo de alcohol ni es sedentaria.

β_1 : tasa de cambio de la probabilidad de presentar HTA cuando se presenta tabaquismo y consumo de alcohol está ausente y no es sedentaria.

x: indicadora de tabaquismo (0=ausencia; 1=presencia)

β_2 : tasa de cambio de la probabilidad de presentar HTA cuando se presenta consumo de alcohol y el tabaquismo está ausente y no es sedentaria.

z: indicadora de consumo de alcohol (0=ausencia; 1=presencia)

β_3 : tasa de cambio de la probabilidad de presentar HTA cuando la persona es sedentaria y no presenta consumo de alcohol ni de tabaquismo.

w: indicadora de sedentarismo (0=ausencia; 1=presencia)

USANDO UN SOFTWARE Y MOSTRANDO LA SALIDA DEL MISMO

```
HTA_NO=c(19,11,6,6,3,4,3,2)
> HTA_SI=c(5,10,5,3,0,2,1,0)
> TABACO=c(0,0,0,0,1,1,1,1) #0=no; 1=si
> ALCOHOL=c(0,0,1,1,0,0,1,1) #0=no; 1=si
> SEDENT=c(0,1,0,1,0,1,0,1) #0=no; 1=si
> modelo2=glm(cbind(HTA_SI,HTA_NO)~TABACO+ALCOHOL+SEDENT,family = binomial)
> summary(modelo2)
```

Call:
glm(formula = cbind(HTA_SI, HTA_NO) ~ TABACO + ALCOHOL + SEDENT, family = binomial)

Deviance Residuals:

	1	2	3	4	5	6	7
8	-0.6827	0.5429	0.9431	-0.8714	-0.9279	0.5665	0.4380
9							-1.138

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.0024	0.4072	-2.462	0.0138 *
TABACO	-0.8665	0.7053	-1.229	0.2192
ALCOHOL	0.2401	0.5203	0.461	0.6445
SEDENT	0.6691	0.4912	1.362	0.1732

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 8.5126 on 7 degrees of freedom
Residual deviance: 5.0806 on 4 degrees of freedom
AIC: 29.278

Number of Fisher Scoring iterations: 4

```
> PVALOR_MODELO=1-pchisq(modelo2$deviance,modelo2$df.residual)
> PVALOR_MODELO
[1] 0.2791245
```

ESCRIBIR EL MODELO ESTIMADO E INTERPRETAR PARA EL PROBLEMA LOS COEFICIENTES (LOS NO EXPLICADOS ANTERIORMENTE).

$$\text{logit}(\hat{p}) = -1,0024 - 0,8665 * x + 0,2401 * z + 0,6691 * w$$

$$\hat{p} = \frac{\exp(-1,0024 - 0,8665 * x + 0,2401 * z + 0,6691 * w)}{1 - \exp(-1,0024 - 0,8665 * x + 0,2401 * z + 0,6691 * w)}$$

donde:

$\hat{\beta}_3 = 0,6691$ La probabilidad de presentar HTA aumenta cuando se presenta sedentarismo mientras el consumo de alcohol y el tabaquismo están ausentes.

w: indicadora de sedentarismo (0=ausencia; 1=presencia)

PROBAR LA SIGNIFICACIÓN DE LOS PARÁMETROS DEL MODELO EXPLICANDO LO QUE ÉSTA SIGNIFICA PARA EL PROBLEMA.

Para el modelo, sólo el parámetro del intercept es significativo. Esto quiere decir que, la probabilidad de presentar hipertensión arterial no se ve afectada significativamente por la presencia o ausencia de tabaquismo, consumo de alcohol y sedentarismo, de acuerdo con la muestra de las 80 personas entre 30 y 82 años encuestadas.

Esto es confirmado por la prueba de Walt, la cual determina que existe independencia entre la presencia de HTA y el tabaquismo, consumo de alcohol y sedentarismo, suponiendo la independencia entre las 3 variables de clasificación, con un p-valor=0,2791.

AJUSTAR DIFERENTES MODELOS (DESDE EL MÁS COMPLEJO AL MÁS SIMPLE) ANALIZANDO LA BONDAD DEL AJUSTE CON EL AIC. ¿CUÁL DE LOS MODELOS RESULTA MEJOR?

```
> modelo3=glm(cbind(HTA_SI,HTA_NO)~TABACO+ALCOHOL,family = binomial)
> summary(modelo3)
```

Call:
glm(formula = cbind(HTA_SI, HTA_NO) ~ TABACO + ALCOHOL, family = binomial)

Deviance Residuals:

1	2	3	4	5	6	7	8
-1.4217	1.2832	0.4735	-0.3181	-1.1166	0.8470	0.1497	-0.9932

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.6622	0.3068	-2.158	0.0309 *
TABACO	-0.8033	0.6984	-1.150	0.2501
ALCOHOL	0.1914	0.5124	0.374	0.7087

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 8.5126 on 7 degrees of freedom
Residual deviance: 6.9663 on 5 degrees of freedom
AIC: 29.164

Number of Fisher Scoring iterations: 4

```
>
> modelo4=glm(cbind(HTA_SI,HTA_NO)~TABACO+SEDENT,family = binomial)
> summary(modelo4)
```

Call:
glm(formula = cbind(HTA_SI, HTA_NO) ~ TABACO + SEDENT, family = binomial)


```

Deviance Residuals:
    1      2      3      4      5      6      7      8
-0.8596  0.3884  1.1909 -0.6175 -0.9728  0.4723  0.5450 -1.0660

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9193     0.3617  -2.542   0.011 *
TABACO       -0.8476     0.7043  -1.204   0.229
SEDENT        0.6539     0.4894   1.336   0.181
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.5126  on 7  degrees of freedom
Residual deviance: 5.2924  on 5  degrees of freedom
AIC: 27.49

Number of Fisher Scoring iterations: 4

>
> modelo5=glm(cbind(HTA_SI,HTA_NO)~SEDENT+ALCOHOL,family = binomial)
> summary(modelo5)

Call:
glm(formula = cbind(HTA_SI, HTA_NO) ~ SEDENT + ALCOHOL, family = binomial)

Deviance Residuals:
    1      2      3      4      5      6      7      8
-0.4630  0.8847  1.1821 -0.5804 -1.3089 -0.2438 -0.1626 -1.4945

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.1073     0.4013  -2.759  0.00579 **
SEDENT        0.6230     0.4852   1.284  0.19916
ALCOHOL       0.1936     0.5131   0.377  0.70595
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.5126  on 7  degrees of freedom
Residual deviance: 6.7637  on 5  degrees of freedom
AIC: 28.961

Number of Fisher Scoring iterations: 4

>
> modelo6=glm(cbind(HTA_SI,HTA_NO)~TABACO,family = binomial)
> summary(modelo6)

Call:
glm(formula = cbind(HTA_SI, HTA_NO) ~ TABACO, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.55585 -0.99784  0.05686  0.70608  1.14939

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.6022     0.2594  -2.321  0.0203 *
TABACO       -0.7841     0.6957  -1.127  0.2597

```

```

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

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.5126  on 7  degrees of freedom
Residual deviance: 7.1051  on 6  degrees of freedom
AIC: 27.302

Number of Fisher Scoring iterations: 4

>
> modelo7=glm(cbind(HTA_SI,HTA_NO)~ALCOHOL,family = binomial)
> summary(modelo7)

Call:
glm(formula = cbind(HTA_SI, HTA_NO) ~ ALCOHOL, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5061  -1.2008  -0.2481   0.2581   1.5389

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.7777     0.2930  -2.654  0.00795 **
ALCOHOL         0.1417     0.5058   0.280  0.77932
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.5126  on 7  degrees of freedom
Residual deviance: 8.4344  on 6  degrees of freedom
AIC: 28.632

Number of Fisher Scoring iterations: 4

>
> modelo8=glm(cbind(HTA_SI,HTA_NO)~SEDENT,family = binomial)
> summary(modelo8)

Call:
glm(formula = cbind(HTA_SI, HTA_NO) ~ SEDENT, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4172  -0.7962  -0.3459   0.1483   1.3725

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.0361     0.3510  -2.952  0.00315 **
SEDENT         0.6086     0.4830   1.260  0.20764
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.5126  on 7  degrees of freedom
Residual deviance: 6.9053  on 6  degrees of freedom
AIC: 27.103

Number of Fisher Scoring iterations: 4

```

```

>
> modelo9=glm(cbind(HTA_SI,HTA_NO)~1,family = binomial)
> summary(modelo9)

Call:
glm(formula = cbind(HTA_SI, HTA_NO) ~ 1, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5357  -1.2577  -0.1421   0.2632   1.4356

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.7309      0.2387  -3.062   0.0022 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 8.5126  on 7  degrees of freedom
Residual deviance: 8.5126  on 7  degrees of freedom
AIC: 26.71

Number of Fisher Scoring iterations: 4

```

Si planteamos todos los modelos desde el más complejo al más simple; el modelo 9 (modelo nulo) es el que presenta el índice AIC más bajo, por lo tanto, es el modelo que mejor ajusta. Esto está en concordancia con el resto de los análisis realizados ya que el intercept es el coeficiente siempre significativo sin importar las variables de clasificación que se presenten en el modelo.

¿A QUÉ CONCLUSIÓN LLEGA?

En conclusión, la hipertensión arterial es independiente de que la persona sea fumadora no y/o consume o no alcohol y/o sea o no sedentaria.