

Taller 8

Juan Pablo Alfonso

10/5/2021

Definición de variables

- tasa_fatal: número de fatalidades por millón de millas transitables en el estado
- tasa_cint: tasa de uso del cinturón
- vel_65: binaria 1 para estados con límite de velocidad 65 millas/hora
- vel_70: binaria 1 para estados con límite de velocidad 70 millas/hora
- niv_alc08: binaria 1 para estados en donde el nivel alcohólico máximo es .08%
- ed_alc21: binaria 1 para estados en donde la edad mínima consumo alcohol 21
- ingreso: ingreso per cápita en el estado
- edad: media de la edad en el estado
- primario: dummy 1 si en el estado la policía puede detener por no uso cinturón
- secundario: dummy 1 sin el estado la policía solo detiene por otras violaciones
- mmt: millones de millas transitables
- cod: código numérico del estado
- ano: año de la observación
- estado: abreviación postal de cada estado

PUNTO 1

Punto 1a

Inspirados en el paper de Cohen y Einav, 2003, harémos un análisis de la efectividad de reglas e incentivos sobre la conducción en carretera para disminuir los accidentes en estados norteamericanos, usaremos una serie de tiempo segregada por estados desde 1983 hasta 1997.

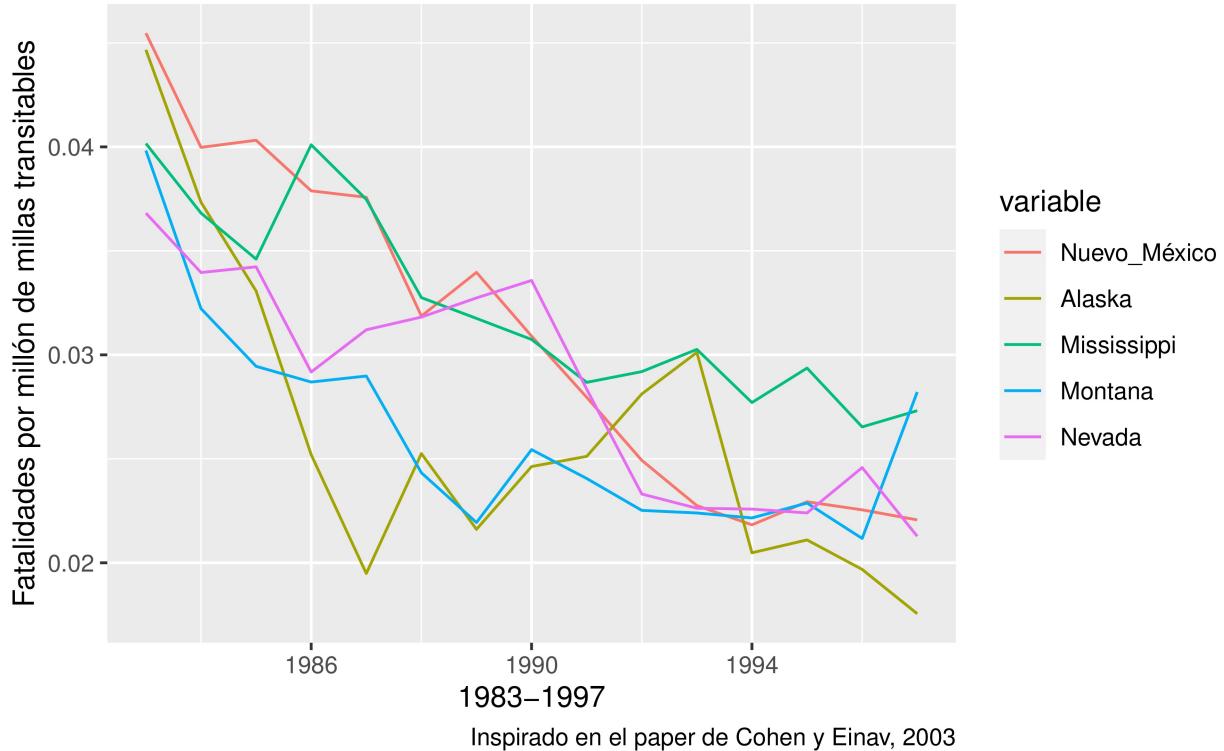
Primero revisaremos gráficamente la evolución de muertes de los 5 estados que en 1983 tuvieron el mayor número de muertes en carretera son:

```
##   estado  ano tasa_fatal
## 1      NM 1983  0.0454701
## 2      AK 1983  0.0446694
## 3      MS 1983  0.0401640
## 4      MT 1983  0.0398273
## 5      NV 1983  0.0368161
```

- NM:Nuevo México
- AK: Alaska
- MS: Mississippi
- MT: Montana
- NV: Nevada

Proyectando los datos para estos 5 estados desde 1983 hasta 1997 encontramos:

Evolución de las muertes en carretera De los 5 estados en U.S que tuvieron la mayor cantidad de muertos en 1983



La cantidad de muertes disminuyó para todos los que más muertes tenían en carretera, entre los factores que pudieron influir se encuentran los límites de velocidad, establecer límites de velocidad y respectivas multas, edad mínima de consumo de alcohol y límite máximo en pruebas de alcoholemia. Revisaremos el impacto de cada una de estas variables sobre la disminución de muertes que acabamos de ver gráficamente.

Punto 1b

Comenzaremos evaluando el impacto de la tasa del uso de cinturón para todos los estados en todos los años sobre la fatalidad carretera.

Modelo 1:

$$tasa_{fatal} = B_0 + B_1 tasacint_i + E_i$$

```
##  
## Modelo 1  
## =====
```

```

##                               tasa_fatal
## -----
## Tasa uso del cinturón      -0.012*** 
##                                         (0.001)
## 
## Constant                   0.026*** 
##                                         (0.001)
## 
## N                          556
## R2                         0.162
## Adjusted R2                0.161
## Residual Std. Error       0.005
## F Statistic                 107.238*** 
## -----
## Notes:                      ***Significant at the 1 percent level.
##                               **Significant at the 5 percent level.
##                               *Significant at the 10 percent level.

```

Nuestras hipótesis son:

$$h_o : B_1 = 0$$

$$h_o : B_1 \neq 0$$

Del primer modelo inferimos que:

- En promedio cuando en ningún estado se usa el cinturón de seguridad hay 0.026 muertes por millón de millas
- En promedio entre más personas usan el cinturón, las muertes por millón de millas disminuyen en todos los estados

Por ende rechazamos la hipótesis nula, con una confianza del 99% la tasa de uso de cinturón afecta la variabilidad de la cantidad de muertos en los estados de forma negativa.

En promedio el 16.2% ($r^2: 0.162$) de la variabilidad de los muertos por millón de millas en carreteras son explicados por la tasa de uso de cinturón.

Sin embargo podríamos caer en un sesgo de variable omitida, hay qué ver cómo cambia con efectos fijos.

Realizaremos el test de durbin watson, este nos permitirá medir la autocorrelación serial de primer orden:

```

## 
## Durbin-Watson test
## 
## data: lm1
## DW = 0.33962, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0

```

La forma de interpretar Durbin Watson es:

2: No auto correlación

0 y 2: auto correlación positiva:

2 y 4: auto correlación negativa

De acuerdo a esto podemos decir que no estamos bien de auto correlación, es bastante grande, es decir que los efectos por año pasan de año en año, lo que implica que podríamos afirmar que si ayer hubo accidentes; hoy también habrán en nuestro modelo 1.

Ahora exploraremos qué ocurre si incluimos más variables a nuestro modelo

Punto 1c

Modelo 2:

```
##  
## Modelo 2  
## ======  
## tasa_fatal  
## -----  
## Tasa de uso cinturon      0.004***  
##                               (0.001)  
##  
## vel_65                     0.0001  
##                               (0.0004)  
##  
## vel_70                     0.002***  
##                               (0.001)  
##  
## niv_alc08                 -0.002***  
##                               (0.0004)  
##  
## ed_alc21                  0.0001  
##                               (0.001)  
##  
## edad                       -0.00001  
##                               (0.0001)  
##  
## ln_ingreso                 -0.018***  
##                               (0.001)  
##  
## Constant                   0.197***  
##                               (0.008)  
##  
## N                          556  
## R2                         0.549  
## Adjusted R2                0.544  
## Residual Std. Error       0.003  
## F Statistic                 95.407***  
## ======  
## Notes:                      ***Significant at the 1 percent level.  
##                               **Significant at the 5 percent level.  
##                               *Significant at the 10 percent level.
```

De modelo 2 inferimos que el intercepto sigue muy significativo (***), y se capturó la diferencia por estados de velocidad límite de 65 ya que este límite 65 no fue significativo en este segundo modelo

El intercepto estaba sobre estimado, pasamos de (0.026) a (**0.197 con significancia del 99%**) *La tasa de cinturón estaba sub estimado, (-0.012* Modelo 1) a (0.004*** Modelo 2) tasa_cint estaba sub estimado.

Las variables que no fueron significativas en este modelo son vel_6, ed_alc21 y edad es decir que ni el límite de velocidad de 65 ni que la edad mínima de conducción de 21 años son estadísticamente significativas para afirmar que afectan la variabilidad de la cantidad de muertos por milla de carreterra.

Al hacer el tesde de Durbin Watson, encontramos:

```
##  
## Durbin-Watson test  
##  
## data: lm2  
## DW = 0.456, p-value < 2.2e-16  
## alternative hypothesis: true autocorrelation is greater than 0
```

Cuando segregamos por más variables encontramos que la autocorrelación disminuye desde 0.33962 hacia 0.456, cada vez más demostramos que pueden haber variables omitidas o diferencias idiosincráticas para demostrarlo construiremos un modelo 3 con efectos fijos para los estados.

Punto 1D

Efectos fijos por estado sin intercepto

Modelo 3

Ahora usaremos cada uno de los 51 estados como una variable y así segregaremos las diferencias idiosincráticas de los estados:

```
##  
## modelo 3  
## -----  
## tasa_fatal  
## -----  
## Tasa de uso cinturon      -0.017***  
##                               (0.001)  
##  
## factor(estado)AK          0.034***  
##                               (0.001)  
##  
## factor(estado)AL          0.032***  
##                               (0.001)  
##  
## factor(estado)AR          0.034***  
##                               (0.001)  
##  
## factor(estado)AZ          0.034***  
##                               (0.001)  
##  
## factor(estado)CA          0.030***  
##                               (0.001)  
##  
## factor(estado)CO          0.027***  
##                               (0.001)  
##  
## factor(estado)CT          0.023***  
##                               (0.001)  
##  
## factor(estado)DC          0.027***
```

```

##                               (0.001)
##                               0.027***  

##                               (0.001)
##                               0.034***  

##                               (0.001)
##                               0.028***  

##                               (0.001)
##                               0.031***  

##                               (0.001)
##                               0.031***  

##                               (0.001)
##                               0.033***  

##                               (0.001)
##                               0.027***  

##                               (0.001)
##                               0.027***  

##                               (0.001)
##                               0.029***  

##                               (0.001)
##                               0.029***  

##                               (0.001)
##                               0.033***  

##                               (0.001)
##                               0.019***  

##                               (0.001)
##                               0.028***  

##                               (0.001)
##                               0.023***  

##                               (0.001)
##                               0.028***  

##                               (0.001)
##                               0.023***  

##                               (0.001)
##                               0.031***  

##                               (0.001)
##                               0.035***
```

```

##                               (0.001)
##                               0.035***  

##                               (0.001)
##                               0.034***  

##                               (0.001)
##                               0.021***  

##                               (0.001)
##                               0.027***  

##                               (0.001)
##                               0.023***  

##                               (0.001)
##                               0.023***  

##                               (0.001)
##                               0.037***  

##                               (0.001)
##                               0.036***  

##                               (0.001)
##                               0.029***  

##                               (0.001)
##                               0.026***  

##                               (0.001)
##                               0.026***  

##                               (0.001)
##                               0.032***  

##                               (0.001)
##                               0.028***  

##                               (0.001)
##                               0.018***  

##                               (0.001)
##                               0.034***  

##                               (0.001)
##                               0.027***  

##                               (0.001)
##                               0.032***  

##                               (0.001)
##                               0.031***
```

```

##                               (0.001)
##                               0.027***  

##                               (0.001)
##                               0.026***  

##                               (0.001)
##                               0.026***  

##                               (0.001)
##                               0.027***  

##                               (0.001)
##                               0.025***  

##                               (0.001)
##                               0.033***  

##                               (0.001)
##                               0.031***  

##                               (0.001)
##  

## N                               556
## R2                               0.991
## Adjusted R2                     0.990
## Residual Std. Error             0.002
## F Statistic                     1,027.060***  

## ======  

## Notes: ***Significant at the 1 percent level.  

##          **Significant at the 5 percent level.  

##          *Significant at the 10 percent level.

```

Del modelo 3 se afirma que todos los estados son significativos para explicar los cambios en las muertes, es decir que existen diferencias idiosincráticas entre estados.

Haciendo un resumen del los 3 R^2 tenemos que

R^2 *Modelo 1:* 0.162

R^2 *Modelo 2:* 0.549

R^2 *Modelo 3:* 0.991

La variabilidad es mucho mejor explicada en el modelo 3, esto ocurre ya que se capturan las diferencias entre cada uno de los estados y estos jalan todos los efectos hacia cada una de las variables dummy haciendo que sean mucho menos volátiles y tengan menos ruido sus errores.

Frente a la tasa del uso de cinturon podemos decir que:

Estimación de la tasa en cada modelo:

Tasa modelo 1: -0.012***

Tasa modelo 2: 0.004***

Tasa modelo 3: -0.017***

Aunque todos son significativos, el modelo 3 logró aislar las diferencias de sesgo por var. omitida, y aún así tuvo un coeficiente negativo, superior al del modelo 1 y 2, sí se puede decir que en promedio las muertes en carretera disminuyen -0.017 cada vez que la proporción de personas que usa el cinturón aumenta.

Realizando el test de Durbin Watson, encontramos que:

```
##  
## Durbin-Watson test  
##  
## data: lmestado  
## DW = 1.2551, p-value < 2.2e-16  
## alternative hypothesis: true autocorrelation is greater than 0
```

El modelo 3, mejoró e hizo que la autocorrelación sea menor sea menor, haciendo que no caigamos en el error de series de tiempo donde lo que ocurrió el año pasado, ocurrirá también en el año presente.

Punto e

Modelo 4

Efectos fijo por estado y por año (efectos del tiempo), eliminando intercepto

```
##  
## modelo 3  
## ======  
## tasa_fatal  
## -----  
## Tasa de uso cinturon      -0.004***  
##                               (0.001)  
##  
## factor(estado)AK          0.031***  
##                               (0.001)  
##  
## factor(estado)AL          0.029***  
##                               (0.001)  
##  
## factor(estado)AR          0.032***  
##                               (0.001)  
##  
## factor(estado)AZ          0.030***  
##                               (0.001)  
##  
## factor(estado)CA          0.024***  
##                               (0.001)  
##  
## factor(estado)CO          0.025***  
##                               (0.001)  
##  
## factor(estado)CT          0.019***  
##                               (0.001)  
##  
## factor(estado)DC          0.024***  
##                               (0.001)  
##  
## factor(estado)DE          0.024***
```

```

##                               (0.001)
##                               0.030***  

##                               (0.001)
##                               0.025***  

##                               (0.001)
##                               0.024***  

##                               (0.001)
##                               0.026***  

##                               (0.001)
##                               0.030***  

##                               (0.001)
##                               0.023***  

##                               (0.001)
##                               0.023***  

##                               (0.001)
##                               0.025***  

##                               (0.001)
##                               0.028***  

##                               (0.001)
##                               0.030***  

##                               (0.001)
##                               0.017***  

##                               (0.001)
##                               0.023***  

##                               (0.001)
##                               0.022***  

##                               (0.001)
##                               0.024***  

##                               (0.001)
##                               0.020***  

##                               (0.001)
##                               0.027***  

##                               (0.001)
##                               0.035***  

##                               (0.001)
##                               0.030***
```

```

##                               (0.001)
##                               0.028***  

##                               (0.001)
##                               0.021***  

##                               (0.001)
##                               0.024***  

##                               (0.001)
##                               0.020***  

##                               (0.001)
##                               0.020***  

##                               (0.001)
##                               0.032***  

##                               (0.001)
##                               0.032***  

##                               (0.001)
##                               0.024***  

##                               (0.001)
##                               0.022***  

##                               (0.001)
##                               0.024***  

##                               (0.001)
##                               0.026***  

##                               (0.001)
##                               0.024***  

##                               (0.001)
##                               0.017***  

##                               (0.001)
##                               0.031***  

##                               (0.001)
##                               0.027***  

##                               (0.001)
##                               0.030***  

##                               (0.001)
##                               0.026***  

##                               (0.001)
##                               0.024***
```

```

##                               (0.001)
##                               0.021***  

##                               (0.001)
##                               0.023***  

##                               (0.001)
##                               0.022***  

##                               (0.001)
##                               0.022***  

##                               (0.001)
##                               0.031***  

##                               (0.001)
##                               0.0003  

##                               (0.001)
##                               0.0002  

##                               (0.001)
##                               0.001  

##                               (0.001)
##                               0.0003  

##                               (0.001)
##                               -0.0002  

##                               (0.001)
##                               -0.002*  

##                               (0.001)
##                               -0.003**  

##                               (0.001)
##                               -0.004***  

##                               (0.001)
##                               -0.005***  

##                               (0.001)
##                               -0.005***  

##                               (0.001)
##                               -0.005***  

##                               (0.001)

```

```

##                               (0.001)
##                               -0.006*** 
##                               (0.001)
##                               -0.006*** 
##                               (0.001)
##                               556
##                               0.994
##                               0.993
##                               0.002
##                               1,263.652*** 
## =====
## Notes: ***Significant at the 1 percent level.
##          **Significant at the 5 percent level.
##          *Significant at the 10 percent level.

```

Los años del 84 al 88 no tienen significancia, no podemos afirmar que el efecto del tiempo afecte la tasa de accidentes en la vía para este periodo

Punto 1F

Suma de residuales m1:

```
## [1] 0.01176252
```

Suma de residuales m2:

```
## [1] 0.006327758
```

parámetros:

El modelo 1 (restricted) tiene:

1 regresor: tasa_cint

1 intercepto

por lo tal tiene 2 parámetros; k1=2

El modelo 2 (unrestricted) tiene: 2 parámetros del modelo anterior +:

vel_65, vel_70, niv_alc08, ed_alc21, ln(ingreso) y edad = 6 regresores

Por lo cual tiene 8 parámetros; k2=8

Ftest manual:

```
## [1] 108.3615
```

Comprobación modelo profesor:

```
## [1] 108.3615
```

Valor crítico para la prueba F:

```
## [1] 3.674587
```

Teniendo en cuenta nuestras hipótesis:

$$h_0 : B_1 = 0$$

$$h_1 : B_1 \neq 0$$

Valor critico: 3.674587

Prueba F: 108.3615

$3.674587 < 108.3615$, rechazo la hipótesis nula; el modelo 2 es mejor que el modelo 1

Punto 1G

PUNTO 2

Cambiar variable de Estado:

Cambiar variable Delito:

Punto 2A

¿Cuántos de los individuos en la muestra violaron los términos de la libertad condicional? En número y como porcentaje de la muestra.

```
##   incump    n
## 1      0 597
## 2      1  78
```

Recordando, si incump es 1 es porque se violaron los términos, de los presos, 78 de ellos violaron los terminos.

```
## [1] 675
```

De 675 presos, 78 violaron, es decir que:

```
## [1] 0.1155556
```

El 11.55% de los presos violaron los términos de libertad

```
library(caTools)
set.seed(144)
split = sample.split(liber$incump, SplitRatio = 0.7)
train = subset(liber, split == TRUE)
test = subset(liber, split == FALSE)
#Comprobando cantidad de split:
nrow(train)/nrow(liber)
```

```
## [1] 0.7007407
```

```
nrow(test)/nrow(liber)
```

```
## [1] 0.2992593
```

Comprobando a través de incump:

Punto 2B

```
mean(liber$incump)
```

```
## [1] 0.1155556
```

```
mean(train$incump)
```

```
## [1] 0.1162791
```

```
mean(test$incump)
```

```
## [1] 0.1138614
```

Sigue quedando aproximadamente el 11% en la base de datos principal, en el train set y en el test set.

Punto 2c

Entrene el modelo 1 de regresión logística:

```
##  
## Modelo GML1  
## =====  
##           incump  
## -----  
## hombre          0.387  
##                 (0.438)  
##  
## raza            -0.887**  
##                 (0.395)  
##  
## edad            -0.0002  
##                 (0.016)  
##  
## estadoFlorida   -4.232***  
##                 (0.626)  
##  
## estadoOhio      -0.392  
##                 (0.572)  
##  
## estadootro     -0.835  
##                 (0.556)  
##  
## tiemp_reclu    -0.124  
##                 (0.120)  
##  
## sent_max        0.080  
##                 (0.055)  
##  
## concurso       1.612***  
##                 (0.385)  
##  
## delitoOtros    0.012  
##                 (0.571)
```

```

## 
## delitoTráfico drogas      -0.266
##                                         (0.641)
## 
## delitoTránsito            0.695
##                                         (0.671)
## 
## Constant                  -1.645
##                                         (1.262)
## 
## N                          473
## Log Likelihood           -125.742
## Akaike Inf. Crit.        277.485
## =====
## Notes:                   ***Significant at the 1 percent level.
##                         **Significant at the 5 percent level.
##                         *Significant at the 10 percent level.

```

Interpretación punto C:

Comencemos por raza con un coeficiente de -0.887, si este regresor es de 1, es porque el preso es blanco, es decir que si efectivamente es blanco su probabilidad de incumplir los términos de la libertad disminuye, si es 0; de otra raza, esta probabilidad no disminuye.

La siguiente variable significativa es “estado” cuando es Florida (4.232***), es decir que cuando los presos son de florida su probabilidad de incurrir en la violación del término de la libertad aumenta.

y finalmente la última variable que influye significativamente es Concurso (1 varios delitos,0 otro caso) (1.612***), es decir que, si cometió varios delitos, su probabilidad de incurrir en la violación del término de la libertad también aumenta.

Cuantitativamente podemos afirmar que los presos que se elijan para su libertad no deberían ser de Florida, ser de piel diferente a blanca y haber cometido varios delitos.

Punto D

Mantiendo las demás variables constante, menos Concurso, la probabilidad de que incumpla los términos de libertad, se daría por el siguiente modelo:

$$incump = B_0 + B_1 \text{Concurso}$$

Reemplazando coeficientes:

$$incump = -1.645 + 1.612(1)$$

```
## [1] -0.033
```

$$incump = -0.033$$

La probabilidad de que la persona incumpla los términos de libertad condicional es de -0.033, es decir que solo si solo se tiene en cuenta si ha cometido varias veces delitos; eso solo variaría la probabilidad de violar los términos en -0.033.

Mantiendo lo mismo para cuando el condenado es de raza blanca:

Modelo:

$$incump = B_0 + B_2 Raza$$

Reemplazando los coefficients de Stargazer:

$$incump = -1.645 - 0.887(1)$$

```
print(-1.645-0.0887)
```

```
## [1] -1.7337
```

Cuando la persona es de raza blanca y esto es la única variable que se tiene en cuenta, se afirma que ser blanco hace que la probabilidad de que viole los términos de la libertad condicional varie -1.7337, verlo de esta forma implica bias de racismo, pero es estadísticamente significativo.

Punto E

Juez usando el modelo

Características:

- B1: Hombre: 1
- B2: Raza blanca: 1
- B3: 30 años:
- B4: Estado Florida: 4
- B5: Tiempo Recluido: 4 meses
- B6: Sentencia máxima 10 meses
- B7: Concurso: no cometió varios delitos: 0
- B8: Delito relacionado con drogas: 3

¿Cuáles son las probabilidades de que viole la libertad condicional?

El modelo sería:

$$incump = B_0 + B_1 Hombre + B_2 Raza + B_3 Edad + B_4 EstadoFlorida + B_5 TiempoReclu + B_6 SentMax + B_7 Concurso + B_9 Delito$$

Reemplazando:

$$incump = -1.645 + 0.387(1) - 0.887(1) - 0.0002(30) - 4.232(1) - 0.124(4) + 0.080(10) + 1.612(0) - 0.266(1)$$

```
Caso1= (-1.645 + 0.387 -0.887-0.0002*(30)-4.232-0.124*(4) + 0.080*(10)- 0.266*(1))
Caso1
```

```
## [1] -6.345
```

La probabilidad del caso 1 con un hombre de estas caracteristicas disminuye 6.345%

TRAMPA DE LAS DUMMYS:

En el modelo 1 GLM1, no nos preocupaba la trampa de las dummies porque, para el caso 1, donde se evalua al hombre, se tiene que la primer variable en orden alfabetico que el One Hot Encoding Elimina no afecta el B7 (Delito, la cual no era desde un inicio una variable dummy sino un factor), sin embargo, para el siguiente caso B7, toma valor de Hurto, es decir que es la primera que se elimina porque se convierte en la base: como los coeficientes de los regresores son constantes pero el intercepto varia entre individuos se preferirá usar el truco de ponerle -1 al modelo, de esta forma, eliminamos el intercepto pero pasa a ser el coeficiente de B7 en el factor Hurto (que se me eliminó porque Hurto>Otros>Tráfico de Drogas alfabeticamente H>O>T)

El intercepto anterior mente era

Caso 2:

- B1: Mujer: 0
- B2: Raza Negra: 0
- B3: 40 años:
- B4: Estado Ohio: 2
- B5: Tiempo Recluida: 8 meses
- B6: Sentencia máxima: 6 meses
- B7: Concurso: no cometió varios delitos: 0
- B8: Delito relacionado con hurto: 2 (Para hacer delito=Hurto, se eliminó el intercepto con(-1) en el modelo, esto se puede hacer porque los regresores son constantes pero el intercepto varia entre individuos)

$$incump = B_1 \text{Hombre} + B_2 \text{Raza} + B_3 \text{Edad} + B_4 \text{EstadoOhio} + B_5 \text{TiempoReclu} + B_6 \text{SentMax} + B_7 \text{Concurso} + B_8 \text{DelitoHurto}$$

Reemplazando :

$$incump = 0.387(0) - 0.887(0) - 0.0002(40) - 0.392(1) - 0.124(8) + 0.080(6) + 1.612(0) - 1.645(1)$$

Operando:

```
probmujer<- -0.0002*(40) -0.392*(1) -0.124 *(8) + 0.080*(6)-1.645*(1)  
probmujer
```

```
## [1] -2.557
```

Entonces, la probabilidad de que una mujer con dichas caracteristicas viole la libertad condicional cae -2.557

Punto G: Probabilidades de cada individuo

```
##          2           3           8          10          19          27  
## 0.034060620 0.109658354 0.047885619 0.264862333 0.139226332 0.063119920  
##          35          46          47          51          52          60  
## 0.111177427 0.498369283 0.285341389 0.275211367 0.268418421 0.507402756  
##          66          70          71          73          75          76
```

```

## 0.166716296 0.064215738 0.134356330 0.146372309 0.385801268 0.072555606
##     80          81          85          89          98         104
## 0.062730405 0.076644455 0.078933293 0.109105843 0.083040866 0.104630267
##    106         109         110         112         115         117
## 0.144107237 0.074197474 0.053469844 0.099624954 0.070025632 0.178695818
##    120         122         125         126         128         130
## 0.092774926 0.118655852 0.055312584 0.110266521 0.167940884 0.210589088
##    131         133         138         141         151         164
## 0.162603177 0.071609759 0.427987643 0.537078357 0.049084455 0.357512845
##    166         167         171         176         177         182
## 0.085018146 0.085018146 0.069878246 0.097913445 0.285623029 0.091163793
##    188         191         192         194         195         196
## 0.228599050 0.305995742 0.574717882 0.583895320 0.584079722 0.537654788
##    199         200         201         202         204         205
## 0.584467753 0.800515946 0.570610752 0.387010359 0.544539762 0.573805356
##    206         208         210         211         216         218
## 0.504402064 0.570640871 0.846513772 0.584313404 0.758704213 0.633010852
##    219         226         228         229         230         232
## 0.124313200 0.682107068 0.306053156 0.847968157 0.748509699 0.379954029
##    236         240         242         254         257         259
## 0.907279069 0.537948775 0.588012456 0.557534155 0.197161510 0.303060221
##    260         264         268         269         271         272
## 0.211191492 0.097689637 0.048480471 0.092958753 0.107748142 0.130316937
##    276         278         282         286         296         299
## 0.285696058 0.071760820 0.041355665 0.199754598 0.059332686 0.191673654
##    301         304         308         318         319         321
## 0.025580207 0.042271847 0.083356403 0.128435689 0.067518531 0.058976424
##    324         327         329         332         334         336
## 0.084775449 0.119525815 0.151691118 0.107852268 0.083780932 0.111871719
##    340         341         345         347         348         349
## 0.077827502 0.110383720 0.003896322 0.003079504 0.016395512 0.026620797
##    354         359         362         364         368         369
## 0.016314464 0.043202439 0.048766322 0.018742518 0.002773754 0.058903408
##    370         373         374         375         376         377
## 0.016568436 0.017490734 0.041265470 0.013967621 0.030782072 0.073940772
##    379         380         389         390         393         394
## 0.038766000 0.019898267 0.027694496 0.007358163 0.009100897 0.008596758
##    395         403         406         407         414         416
## 0.026245807 0.028947618 0.005173238 0.016572280 0.012015091 0.056907422
##    421         428         432         445         448         452
## 0.017496864 0.004280348 0.038502370 0.017831333 0.051980578 0.038133416
##    456         460         461         462         465         466
## 0.011150959 0.042636344 0.021270249 0.014485086 0.031328474 0.038739864
##    469         472         474         477         480         484
## 0.017570865 0.013296622 0.003788245 0.041369787 0.005294832 0.012742494
##    488         491         495         498         502         504
## 0.035895075 0.027625012 0.039237602 0.005454823 0.035977999 0.018016245
##    507         509         511         519         520         524
## 0.016361583 0.074849108 0.035466999 0.028521105 0.037164077 0.032317099
##    525         526         527         528         529         534
## 0.016350860 0.007459125 0.043163282 0.015198906 0.010002962 0.007975122
##    535         538         540         554         556         564
## 0.042159035 0.014985943 0.022210826 0.034494571 0.022438983 0.013138276
##    567         573         576         587         589         590

```

```

## 0.002334403 0.031803191 0.022092215 0.023175432 0.047758466 0.038807280
##      596       601       614       618       619       621
## 0.069322487 0.013676942 0.010439569 0.028642415 0.035927485 0.054843706
##      623       629       630       635       638       641
## 0.002352864 0.027870030 0.032253521 0.044640361 0.015714397 0.015543545
##      647       649       655       657       661       665
## 0.071693206 0.002469374 0.041288133 0.002342076 0.016258314 0.003514337
##      669       672       673       674
## 0.031237469 0.047409154 0.047600750 0.147811407

```

Aquí mostramos las probabilidades de cada persona de que pertenezca al grupo de los que posiblemente violen la libertad condicional.

La probabilidad más baja:

```
min(predictions)
```

```
## [1] 0.002334403
```

La probabilidad más alta, hay un individuo que con el 90% de seguridad no respetará la libertad condicional

```
max(predictions)
```

```
## [1] 0.9072791
```

En términos de ODDS

```

##      2
## 0.03526165

```

Por cada punto de incumplir, tiene 0.035 de ser bueno y cumplir la libertad condicional.

Punto I, Matriz de confusión

```

## pred_0_1
##      0   1
##  0 167 12
##  1  11 12

```

TN FP FN TP

- Entonces, habrán 167 falsos negativos, personas que no fue son bueno respetando pero que el modelo interpreto que sí -Hay 12 falsos positivos, presos a los que no les dieron libertad porque el modelo los juzgó mal

PUNTO J Características del modelo, sensibilidad, especificidad, accuracy. Usando un threshold de 0.5.

```

#Sensitivity TP/(TP+FN)
12/(12+11)

```

```
## [1] 0.5217391
```

```
#Specificity  $TN/(TN+FP)$   
167/(167+12)
```

```
## [1] 0.9329609
```

```
#Accuracy  $(TP+TN)/(TP+TN+FP+FN)$   
(12+167)/(12+167+12+11)
```

```
## [1] 0.8861386
```

```
accurate <- test$incump == pred_0_1  
mean(accurate)
```

```
## [1] 0.8861386
```

La sensibilidad del modelo: 0.5217391 La especificidad del modelo: 0.9329609 y la precisión del modelo: 0.8861386 el cual también es el promedio de la precisión.

Punto K

AUC del Modelo:

```
## Warning: package 'ROCR' was built under R version 4.0.5
```

```
## [1] 0.8945834
```

Es un buen valor para clasificación: 0.8945834 es un clasificador fuerte