Efectos de las Leyes Obligatorias del Uso del Cinturón de Seguridad en el Comportamiento de Conducción y las Muertes por Accidentes de Tránsito

Nelson Brayan Mamani Flores

27 de febrero, 2023

Contents

Introducción

Los accidentes de tráfico son la principal causa de muerte de los estadounidenses entre los 5 y los 32 años de edad. Es por ello que, mediante distintas políticas de gasto, el gobierno federal ha alentado a los estados a instituir normativas de obligatoriedad de uso del cinturón de seguridad para reducir el número de muertes y lesiones graves. En ese sentido, este trabajo investigará la eficacia de las leyes en el aumento del uso del cinturón de seguridad y la reducción de víctimas mortales. Cabe resaltar que los datos recogidos contienen un panel equilibrado de datos anuales sobre 50 estados de EE.UU. (más el distrito de Columbia) para el periodo 1983-1997.

Desarrollo

Claramente, el uso de los cinturones de seguridad en las carreteras ha demostrado ser una medida eficaz para reducir la cantidad de muertes y lesiones graves en los accidentes de tránsito. De hecho, según la World Health Organization (WHO, 2014), el uso del cinturón de seguridad reduce el riesgo de muerte en un 45% y el riesgo de lesiones graves en un 50%.

Asimismo, varios estudios han demostrado que la falta de uso del cinturón de seguridad está relacionada con una mayor tasa de lesiones graves y muertes en accidentes de tránsito. En consideración a ello, Perez-Nuñez et.al (2014), en su estudio, "El estado de las lesiones causadas por el tránsito en México: evidencias para fortalecer la estrategia mexicana de seguridad vial", demostró que los pasajeros que no usaban cinturones de seguridad en el momento de la colisión presentaban una mayor tasa de lesionabilidad.

Análisis de Datos

Al iniciar un script en R, es común llamar todas las librerías necesarias al principio del script mediante la función library(). La razón principal de esto es asegurarse de que todas las funciones y objetos definidos en esas librerías estén disponibles para su uso en el resto del script.

```
library(tidyverse)
library(lmtest)
library(plm)
library(modeest)
library(data.table)
library(dplyr)
library(stargazer)
library(tinytex)
```

Definición de la base

Los datos para el analisis provienen de la investigación realizada por el profesor Liran Einav de la Universidad de Stanford en su arituculo: "The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities," The Review of Economics and Statistics, 2003, Vol. 85, pp 828-843. (Eynav & Cohen, 2003)

El fichero de datos contiene las siguientes variables:

- name: nombre del estado, incluyendo el distrito de Columbia.
- fips: código estatal federal de procesamiento de información.
- fatalityrate: número de muertes por millón de millas conducidas.
- sb_usage: proporción de uso del cinturón de seguridad.
- vmt: millones de millas de tráfico por año.

- speed65: 1 si en el estado hay un límite de velocidad de 65 millas por hora, 0 en otro caso.
- speed70: 1 si en el estado hay un límite de velocidad de 70 millas por hora o superior, 0 en otro caso.
- ba08: 1 si en el estado hay un límite de alcohol en sangre menor o igual al 0,08%, 0 en otro caso.
- drinkage21: 1 si en el estado hay una prohibición de beber alcohol a menores de 21 años, 0 en otro caso.
- income: renta per cápita en el estado;
- age: edad media en el estado.
- primary: 1 si en el estado la ley de obligatoriedad de uso del cinturón de seguridad se aplica de forma básica, 0 en otro caso.
- secondary: 1 si en el estado la ley de obligatoriedad de uso del cinturón de seguridad se aplica de forma secundaria, 0 en otro caso.

Ahora bien, como parte del analisis de datos, realizamos la importación de nuestra data en R. Para ello utilizamos los siguientes comandos.

```
S6_cinturon <- read.csv("C:/Users/Nelson/Documentos/Projects/Modelo_Panel/base/S6_cinturon.csv", sep=";
```

Una vez realizada la importación de nuestros datos, estructuramos nuestra data en forma de tabla utilizando la siguiente función:

```
# Tabla1
knitr::kable(head((subset(S6_cinturon,
                          select = c(name,year,fatalityrate,sb_useage,
                                     drinkage21,income,age,
                                     speed65,speed70))),20))
```

name	year	fatalityrate	sb_useage	drinkage21	income	age	speed65	speed70
Alaska	1983	0.0446694	NA	1	17973	28.23497	0	0
Alaska	1984	0.0373363	NA	1	18093	28.34354	0	0
Alaska	1985	0.0330729	NA	1	18925	28.37282	0	0
Alaska	1986	0.0251996	NA	1	18466	28.39665	0	0
Alaska	1987	0.0194872	NA	1	18021	28.45325	0	0
Alaska	1988	0.0252538	NA	1	18447	28.85142	0	0
Alaska	1989	0.0216105	NA	1	19970	29.14895	0	0
Alaska	1990	0.0246293	0.45	1	21073	29.58628	0	0
Alaska	1991	0.0251181	0.66	1	21496	29.82771	0	0
Alaska	1992	0.0281177	0.66	1	22073	30.21070	1	0
Alaska	1993	0.0301174	0.69	1	22711	30.46439	1	0
Alaska	1994	0.0204819	0.69	1	23417	30.75657	1	0
Alaska	1995	0.0211011	0.69	1	23971	31.17860	1	0
Alaska	1996	0.0196841	0.69	1	24310	31.44535	1	0
Alaska	1997	0.0175519	0.69	1	24969	31.60147	1	0
Alabama	1983	0.0299691	NA	0	9505	34.20065	0	0
Alabama	1984	0.0282758	0.13	0	10417	34.42163	0	0
Alabama	1985	0.0251346	0.17	1	11133	34.60257	0	0
Alabama	1986	0.0317913	0.29	1	11736	34.76067	0	0
Alabama	1987	0.0296852	0.21	1	12394	34.95566	1	0

A continuación, realizamos la descripción basica de los datos de nuestras variables no categóricas. Los resultados son los siguientes:

```r
# Tabla 2
resumen_std <- summary(subset(S6_cinturon,
                              select = c(fatalityrate,sb_useage,
                                         income,l_income,vmt)))
knitr::kable(resumen_std)
```

fatalityrate	sb_useage	income	l_income	vmt
Min. :0.008327	Min. :0.0600	Min. : 8372	Min. : 9.033	Min. : 3099
1st Qu.:0.017341	1st Qu.:0.4200	1st Qu.:14266	1st Qu.: 9.566	1st Qu.: 11401
Median :0.021199	Median :0.5500	Median :17624	Median : 9.777	Median : 30319
Mean :0.021490	Mean :0.5289	Mean :17993	Mean : 9.762	Mean : 41448
3rd Qu.:0.024774	3rd Qu.:0.6500	3rd Qu.:21080	3rd Qu.: 9.956	3rd Qu.: 52312
Max. :0.045470	Max. :0.8700	Max. :35863	Max. :10.487	Max. :285612
NA	NA's :209	NA	NA	NA

Como se observa en los resultados, con respecto a la variable "sb_usage", se cumple que la media en proporción de uso de cinturon de seguridad por estado es del 52.89%. Asimimo, la variable "fatalityrate", indica que número de muertes por millón de millas conducidas es de 0.0215.

Ahora bien para realizar un analisis más eficiente, seleccionamos las variables pertinentes para nuestro modelo y le damos el formato de panel de datos.

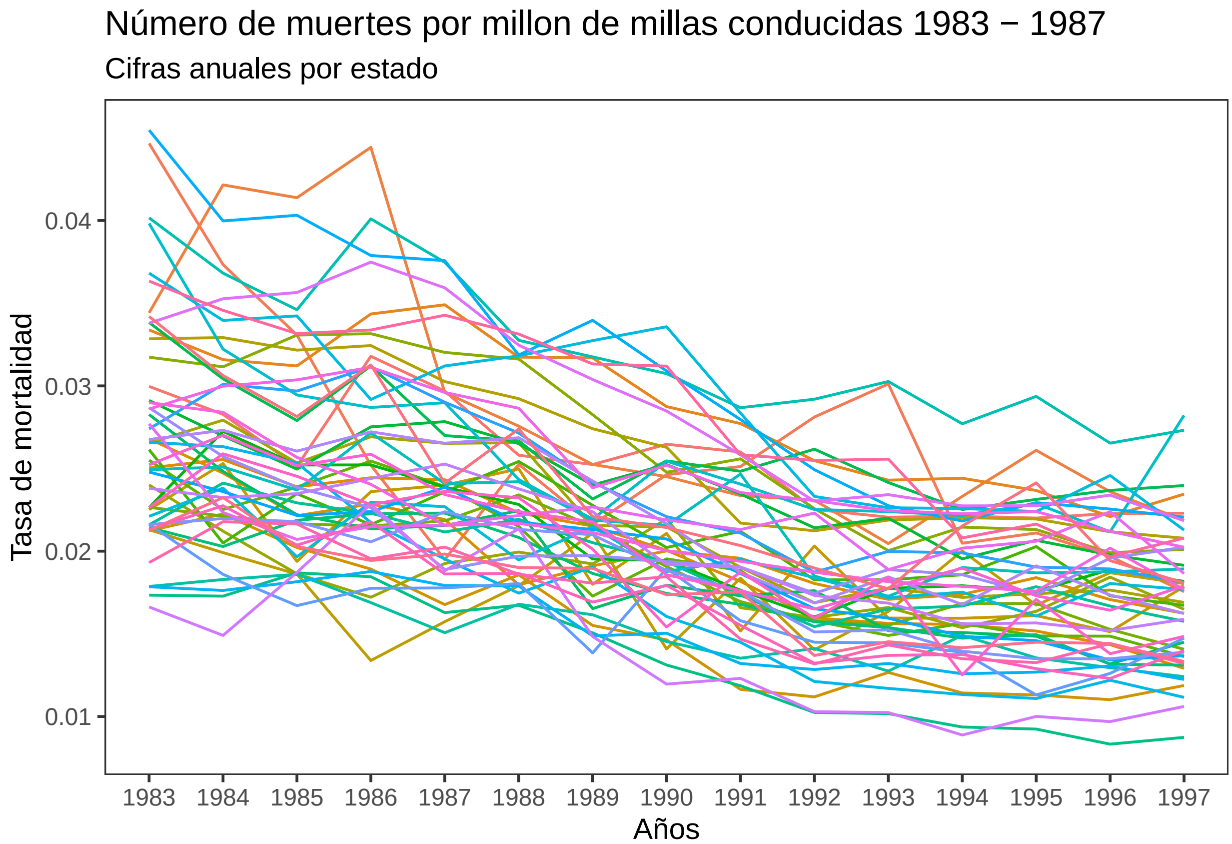
```r
# Selección de  variables de interes y especificación de data panel
S6_cinturon2 <- S6_cinturon %>%
  select(name, year, fatalityrate, sb_useage, speed65,
         speed70, ba08, drinkage21, l_income, age, primary, secondary)

S6_cinturon_pl <- pdata.frame(S6_cinturon2,
                              index=c("name","year"))
```

Visualización inicial

A continuación realizamos los graficos de tendencia para cada variable pertinente en nuestro modelo. Los resultados nos mostraran la evolución de una determinada variable a lo largo de los años. En este caso, en la figura 2, se presenta a la variable de accidentes fatales por millon de millas conducidas para cada estado (fatalityrate).

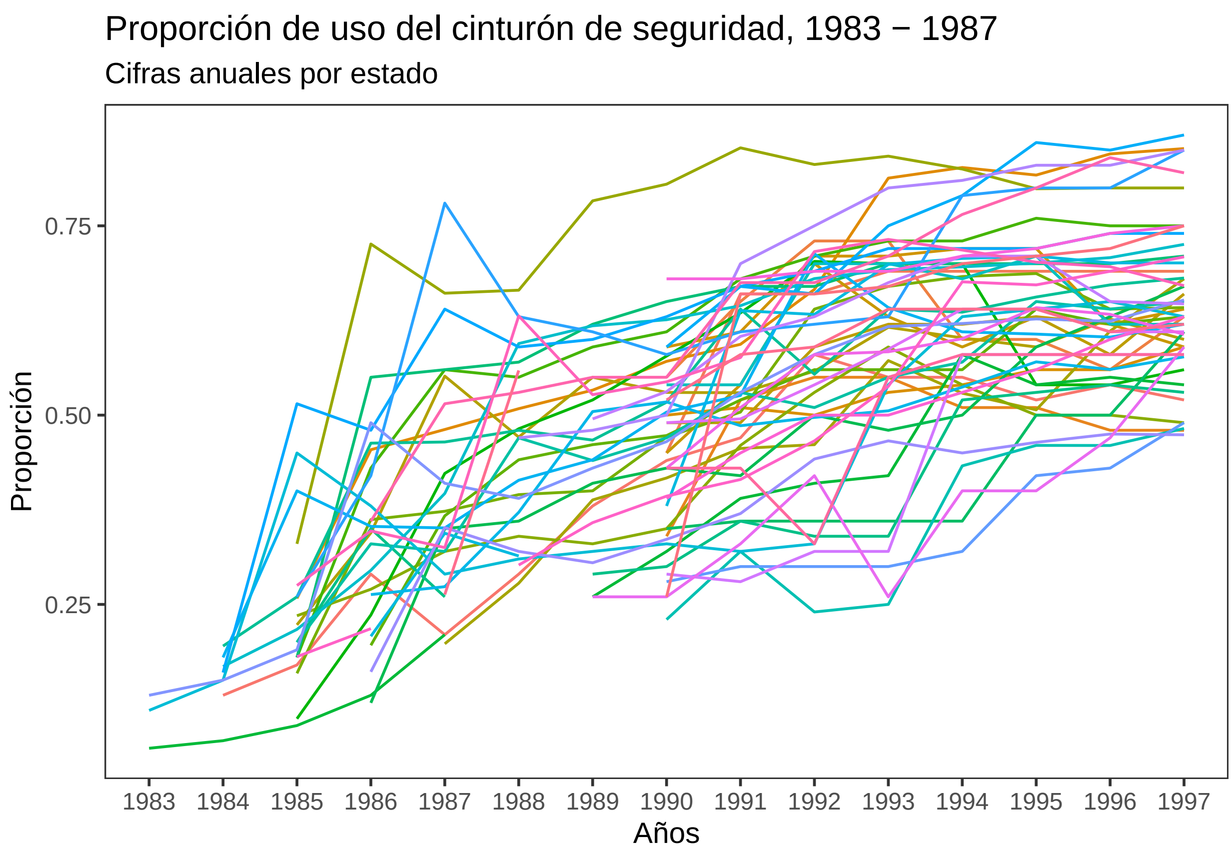
```r
# Gráfico 1
S6_cinturon_pl %>%
  ggplot(aes(x=year, y=fatalityrate, color=name,
      group=name))+
  geom_line()+
  labs(title ="Número de muertes por millon de millas conducidas 1983 - 1987",
       subtitle = "Cifras anuales por estado", x="Años",
       y="Tasa de mortalidad")+
  theme_bw() +
  theme(panel.grid.major = element_line(color = "white"),
        panel.grid.minor = element_line(color = "white"),
        panel.background = element_rect(fill = "white"))+
  theme(legend.position = "none")
```



Tal como se puede observar existe una disminución en la tendencia de la variable de muertes por millon de millas conducidas. Tal es asi que desde el año 1991 a 1997, el ratio se concentró en el intervalo de 0.01 a 0.03.

Seguidamente, analizamos la evolución de nuestra variable "sb_usage", que corresponde a la proporcion de uso de cinturon de seguridad.

```
# Gráfico 2
S6_cinturon_pl %>%
  ggplot(aes(x=year, y=sb_useage, color=name,
      group=name))+
  geom_line()+
  labs(title ="Proporción de uso del cinturón de seguridad, 1983 - 1987",
       subtitle = "Cifras anuales por estado", x="Años",
       y="Proporción")+
  theme_bw() +
  theme(panel.grid.major = element_line(color = "white"),
        panel.grid.minor = element_line(color = "white"),
        panel.background = element_rect(fill = "white"))+
  theme(legend.position = "none")
```



Tal como observamos en el gráfico 3, la proporción de uso de los cinturones de seguridad en los estados a ido aumentado. Esto corresponde a un mayor control de parte de las autoridades de transporte en los estados unidos.

Descripción estadística básica

A priori, los datos mostrados en la tabla 3 nos indican que Massachusetts es el estado con una menor tasa de fatalidad, correspondiente al 1.3% por millón de millas conducidas. Asimismo, el estado Mississippi es el que cuenta con una mayor tasa de fatalidad, la cual es un 3.2% por millón de millas conducidas.

```
# Tabla 3
sum <- S6_cinturon_pl %>%
  group_by(name) %>%
  summarize(Promedio = mean(fatalityrate),
            Desv_std=sd(fatalityrate),
            Mediana=median(fatalityrate),
            Minimo=min(fatalityrate),
            Maximo=max(fatalityrate))
knitr::kable(sum)
```

name	Promedio	Desv_std	Mediana	Minimo	Maximo
Alabama	0.0254442	0.0032859	0.0252422	0.0220036	0.0317913
Alaska	0.0262288	0.0074594	0.0251181	0.0175519	0.0446694
Arizona	0.0287392	0.0079733	0.0252470	0.0204598	0.0444297

name	Promedio	Desv_std	Mediana	Minimo	Maximo
Arkansas	0.0285822	0.0043967	0.0287468	0.0220905	0.0349066
California	0.0196591	0.0044684	0.0200521	0.0129126	0.0255423
Colorado	0.0199553	0.0030991	0.0191101	0.0162401	0.0267950
Connecticut	0.0152814	0.0040752	0.0146371	0.0110183	0.0222528
Delaware	0.0196884	0.0037027	0.0193849	0.0151317	0.0253017
District of Columbia	0.0176193	0.0025166	0.0180397	0.0133861	0.0212972
Florida	0.0262862	0.0049601	0.0262825	0.0207825	0.0329219
Georgia	0.0217134	0.0044190	0.0214720	0.0168797	0.0279285
Hawaii	0.0184845	0.0024848	0.0184309	0.0153749	0.0240082
Idaho	0.0264516	0.0053064	0.0255864	0.0199059	0.0331577
Illinois	0.0190023	0.0032125	0.0190678	0.0140658	0.0234063
Indiana	0.0192065	0.0042652	0.0188332	0.0136258	0.0255039
Iowa	0.0209563	0.0028859	0.0204908	0.0167238	0.0261431
Kansas	0.0206626	0.0035079	0.0194319	0.0160162	0.0272480
Kentucky	0.0238761	0.0033188	0.0240012	0.0191457	0.0291179
Louisiana	0.0262747	0.0033282	0.0254600	0.0225220	0.0338374
Maine	0.0188557	0.0043676	0.0176117	0.0131836	0.0282686
Maryland	0.0179283	0.0033601	0.0174413	0.0131091	0.0223056
Massachusetts	0.0133941	0.0038750	0.0131151	0.0083273	0.0186921
Michigan	0.0193444	0.0028952	0.0193733	0.0154359	0.0241216
Minnesota	0.0152268	0.0020806	0.0149133	0.0124095	0.0186001
Mississippi	0.0322316	0.0045730	0.0307402	0.0265362	0.0401640
Missouri	0.0218951	0.0031459	0.0218770	0.0172744	0.0271584
Montana	0.0262857	0.0050563	0.0243303	0.0211730	0.0398273
Nebraska	0.0196911	0.0023775	0.0194674	0.0160688	0.0238135
Nevada	0.0285766	0.0053153	0.0291761	0.0212766	0.0368161
New Hampshire	0.0170777	0.0057636	0.0160504	0.0110871	0.0265980
New Jersey	0.0151975	0.0025238	0.0148753	0.0122417	0.0187715
New Mexico	0.0308651	0.0079796	0.0309017	0.0218262	0.0454701
New York	0.0192185	0.0040258	0.0207386	0.0134186	0.0247902
North Carolina	0.0237686	0.0048128	0.0220868	0.0181090	0.0311542
North Dakota	0.0160394	0.0027640	0.0157957	0.0113063	0.0216297
Ohio	0.0181492	0.0035794	0.0188336	0.0134931	0.0223859
Oklahoma	0.0206513	0.0031982	0.0195805	0.0173816	0.0286826
Oregon	0.0220081	0.0045047	0.0216546	0.0162390	0.0273122
Pennsylvania	0.0200738	0.0037872	0.0192047	0.0151998	0.0252715
Rhode Island	0.0141392	0.0045758	0.0123042	0.0088795	0.0228403
South Carolina	0.0288307	0.0058434	0.0284792	0.0218470	0.0374867
South Dakota	0.0216569	0.0020773	0.0215816	0.0186445	0.0277030
Tennessee	0.0255013	0.0037780	0.0238393	0.0202392	0.0311227
Texas	0.0214062	0.0037608	0.0201849	0.0175763	0.0289878
Utah	0.0207944	0.0037727	0.0189733	0.0164287	0.0270131
Vermont	0.0193069	0.0042890	0.0187394	0.0125163	0.0258914
Virginia	0.0169127	0.0035622	0.0169203	0.0122998	0.0227502
Washington	0.0169473	0.0032407	0.0180649	0.0131819	0.0217823
West Virginia	0.0284784	0.0058695	0.0311973	0.0196688	0.0363372
Wisconsin	0.0174355	0.0031494	0.0175327	0.0133262	0.0232420
Wyoming	0.0239652	0.0052070	0.0220870	0.0177253	0.0341965

Por otro lado, con respecto a la variable "sb_useage", que corresponde a la proporción de uso del cinturon de seguridad, podemos notar en la tabla 4 que los estados que anteriormente tenian una mayor tasa de fatalidad, ahora muestran una menor proporción de uso de cinturon, a excepcion de Massachussets. Es asi que, tenemos indicios de una relación directa entre ambas variables, las cuales seran de gran ayuda para explicar nuestro modelo.

```r
# Tabla 4
sum2 <- S6_cinturon_pl %>%
  group_by(name) %>%
  na.omit() %>%
  summarise(Promedio = mean(sb_useage),
            Desv_std=sd(sb_useage),
            Mediana=median(sb_useage),
            Minimo=min(sb_useage),
            Maximo=max(sb_useage))
knitr::kable(sum2)
```

name	Promedio	Desv_std	Mediana	Minimo	Maximo
Alabama	0.4028571	0.1565704	0.45500	0.130	0.5800000
Alaska	0.6525000	0.0829372	0.69000	0.450	0.6900000
Arizona	0.6312500	0.0691660	0.61500	0.550	0.7300000
Arkansas	0.4439000	0.1207996	0.49500	0.198	0.5500000
California	0.6322538	0.1880705	0.59350	0.258	0.8520000
Colorado	0.5362500	0.0324863	0.53500	0.500	0.5900000
Connecticut	0.6650000	0.0552914	0.67500	0.590	0.7200000
Delaware	0.5937500	0.0738604	0.60500	0.450	0.7000000
District of Columbia	0.5850000	0.0634710	0.60500	0.490	0.6600000
Florida	0.5225000	0.1167421	0.55150	0.223	0.6270000
Georgia	0.4633636	0.1412121	0.46100	0.198	0.6800000
Hawaii	0.7476923	0.1398889	0.80000	0.330	0.8530000
Idaho	0.4196154	0.1157667	0.46000	0.235	0.5900000
Illinois	0.5096923	0.1655387	0.50400	0.159	0.6870000
Indiana	0.5023333	0.1271615	0.54000	0.196	0.6400000
Iowa	0.6176923	0.1654365	0.68000	0.180	0.7600000
Kansas	0.5166154	0.1793226	0.54000	0.099	0.7030000
Kentucky	0.3264286	0.1910540	0.35500	0.060	0.5800000
Louisiana	0.4550000	0.1464427	0.45500	0.120	0.6700000
Maine	0.4250000	0.0985611	0.36000	0.350	0.6100000
Maryland	0.6138462	0.1425096	0.67000	0.180	0.7100000
Massachusetts	0.3800000	0.1190874	0.34500	0.200	0.5400000
Michigan	0.5232500	0.1505777	0.53550	0.195	0.6810000
Minnesota	0.4869231	0.1376823	0.51000	0.200	0.6500000
Mississippi	0.3594000	0.1103653	0.37650	0.230	0.4823000
Missouri	0.6387500	0.0701911	0.65000	0.540	0.7100000
Montana	0.5544214	0.1967704	0.63525	0.168	0.7255000
Nebraska	0.4053333	0.1767511	0.33000	0.110	0.6500000
Nevada	0.5486364	0.1940774	0.63800	0.208	0.7100000
New Hampshire	0.4719500	0.1091288	0.50520	0.263	0.5770000
New Jersey	0.4974286	0.1472250	0.51500	0.180	0.7130001
New Mexico	0.7550000	0.1055597	0.77000	0.590	0.8700000
New York	0.6153571	0.1544029	0.65500	0.160	0.7400000
North Carolina	0.6446154	0.1673129	0.63000	0.260	0.8500000
North Dakota	0.3550000	0.0792825	0.31000	0.280	0.4900000
Ohio	0.4604000	0.1790135	0.49000	0.130	0.6520000
Oklahoma	0.3845417	0.0957476	0.40600	0.161	0.4750000
Oregon	0.7020000	0.1571128	0.77500	0.470	0.8500000

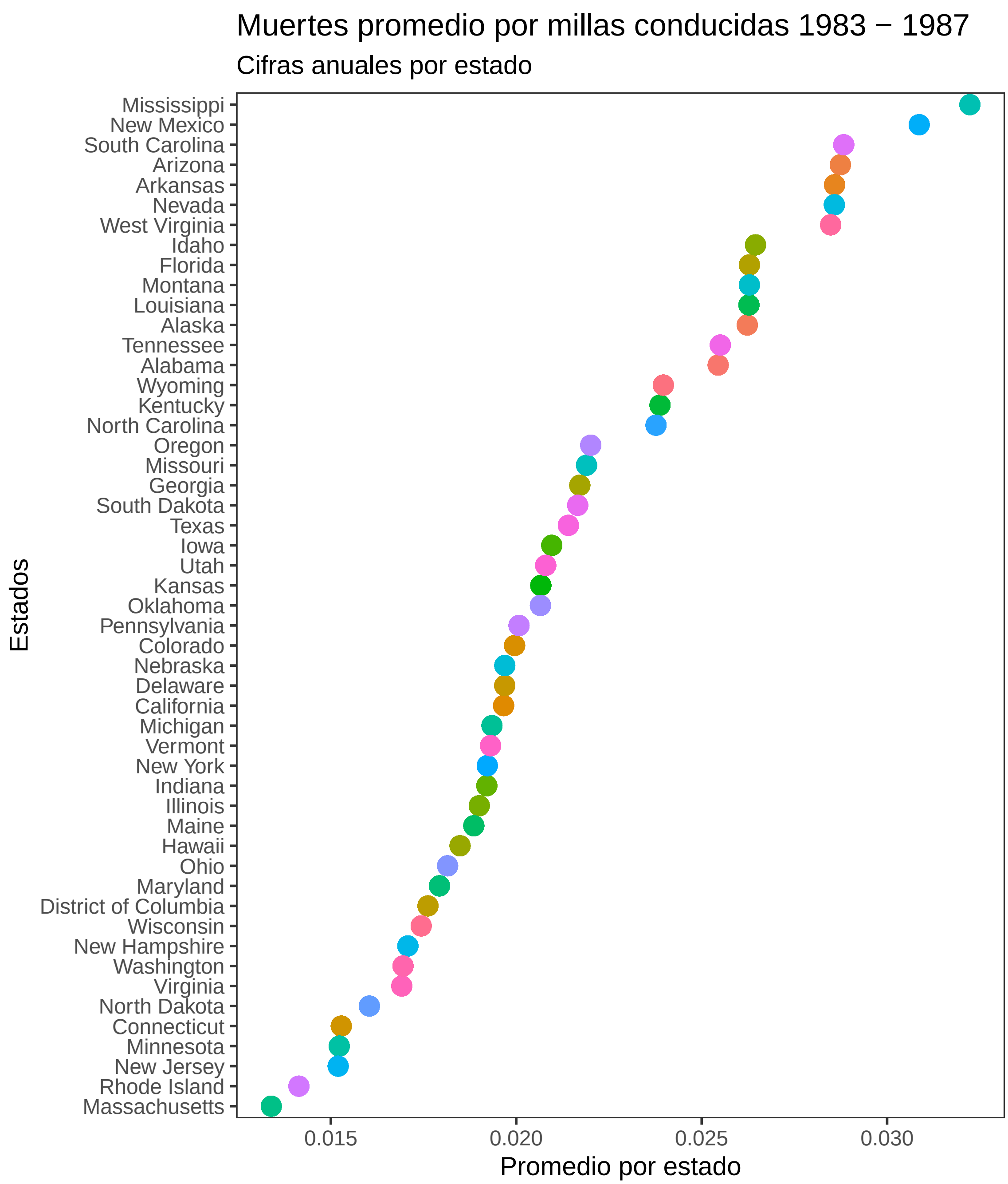
name	Promedio	Desv_std	Mediana	Minimo	Maximo
Pennsylvania	0.6280000	0.0743135	0.64700	0.495	0.7100000
Rhode Island	0.4412500	0.1489427	0.45000	0.280	0.5800000
South Carolina	0.5266667	0.1599523	0.58700	0.128	0.6400000
South Dakota	0.3766667	0.1119151	0.40000	0.260	0.5900000
Tennessee	0.5732875	0.0705632	0.59230	0.430	0.6419000
Texas	0.7075000	0.0271241	0.70000	0.680	0.7500000
Utah	0.5200000	0.0781939	0.51500	0.390	0.6300000
Vermont	0.4681667	0.1883198	0.44050	0.181	0.7090000
Virginia	0.5738461	0.1622328	0.63100	0.275	0.7320000
Washington	0.6491667	0.1484133	0.67500	0.360	0.8400000
West Virginia	0.5075000	0.0973579	0.56500	0.330	0.5800000
Wisconsin	0.5662000	0.1135418	0.60000	0.263	0.6400000
Wyoming	0.6412500	0.1572475	0.68500	0.260	0.7500000

Ahora bien, para un mejor analisis los nuestros datos obtenidos, ilustraremos nuestros resultados de manera gráfico tomando en cuenta nuestras 2 variables determinantes. Es asi que, generaramos una variable que sea el promedio por estado del ratio de fatalidad por millas conducidas de la siguiente manera.

```
# Generamos el promedio de la variable fatalityrate por estado ----
S6_cinturon_pl <- S6_cinturon_pl %>%
  group_by(name) %>%
  mutate(fatalidad_prom = mean(fatalityrate))
```

Seguidamente, generamos el gráfico correspondiente al promedio de ratio de fatalidad por millas conducidas por cada estado.

```
#Grafico 3
ggplot(S6_cinturon_pl, aes(x=fatalidad_prom, y=(reorder(name, fatalidad_prom)), color=name)) +
  geom_point(size=3.5) +
  labs(title ="Muertes promedio por millas conducidas 1983 - 1987 ",
       subtitle = "Cifras anuales por estado", x="Promedio por estado",
       y="Estados") +
  theme_bw() +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(colour="grey60",
                                          linetype="blank")) +
  theme(legend.position = "None")
```

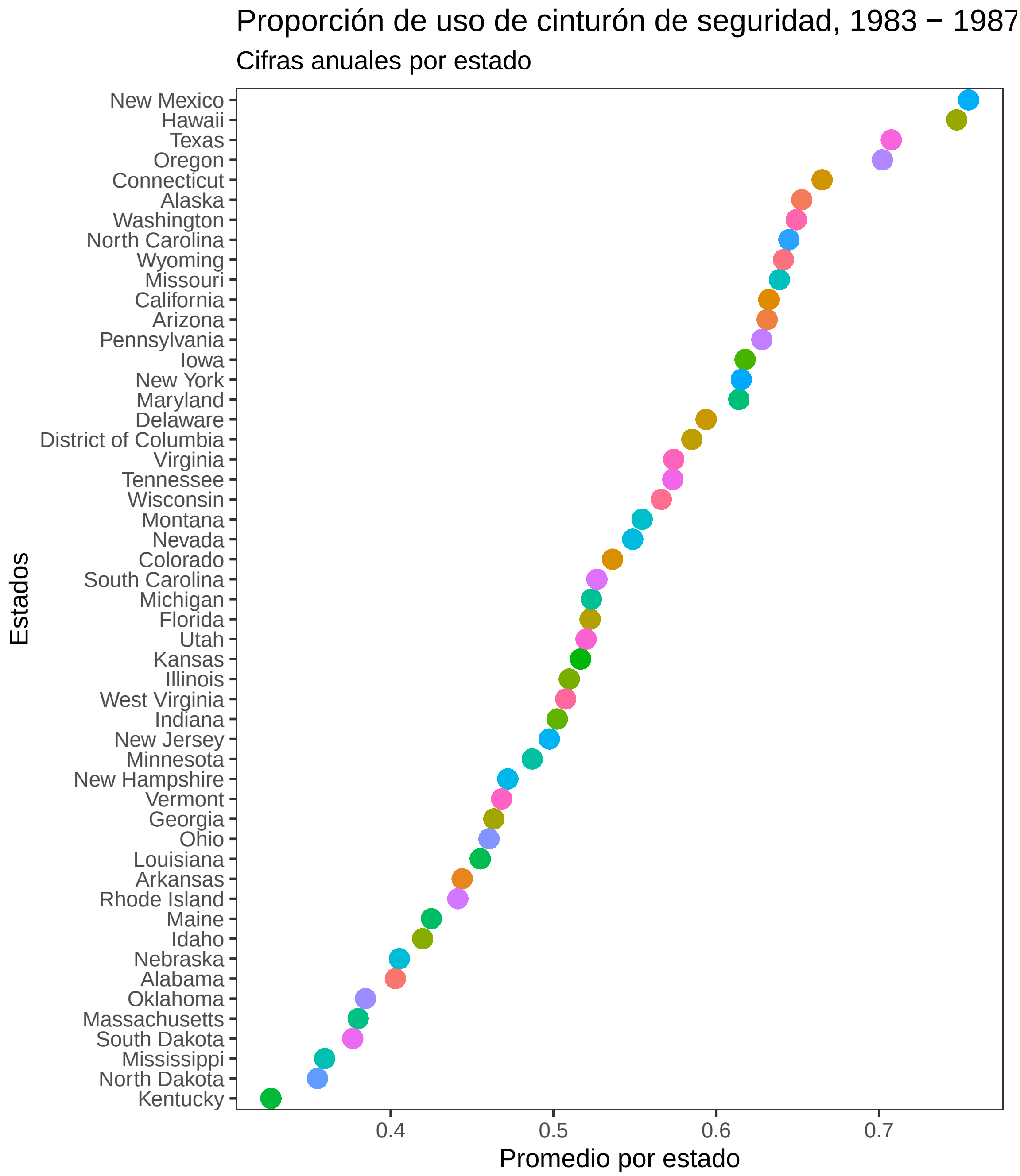


Tal como se puede observar en el gráfico 1, el estado de de Massachusetts presente el menor ratio de fatalidad por millon de millas conducidas correspondiente a menos 0.015. En cambio, el estado de Mississippi presente una mayor tasa de fatalidad promedio (0.03223) mayor a todos los estados.

A continuación, realizamos el mismo proceso para nuestra variable de proporción de uso de cinturón de seguridad.

```
# Generamos el promedio de la variable fatalityrate por estado ----
S6_cinturon_pl <- S6_cinturon_pl %>%
  group_by(name) %>%
  na.omit() %>%
  mutate(cinturon_prom = mean(sb_useage))
```

```r
#Grafico 4
ggplot(S6_cinturon_pl, aes(x=cinturon_prom, y=(reorder(name, cinturon_prom)), color=name)) +
  geom_point(size=3.5) +
  labs(title ="Proporción de uso de cinturón de seguridad, 1983 - 1987 ",
       subtitle = "Cifras anuales por estado", x="Promedio por estado",
       y="Estados") +
  theme_bw() +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(colour="grey60",
                                          linetype="blank")) +
  theme(legend.position = "None")
```



Los gráficos mostrados nos muestran una posible relación entre el ratio de fatalidad por millon de millas conducidas y la proporción de uso de cinturón de seguridad. Es uno de los resultados a priori que podemos concluir. Sin embargo, tenemos que tomar en cuenta nuestras demás variables y ver la significatividad de nuestro modelo para una conclusión más acertada, y libre de sesgo.

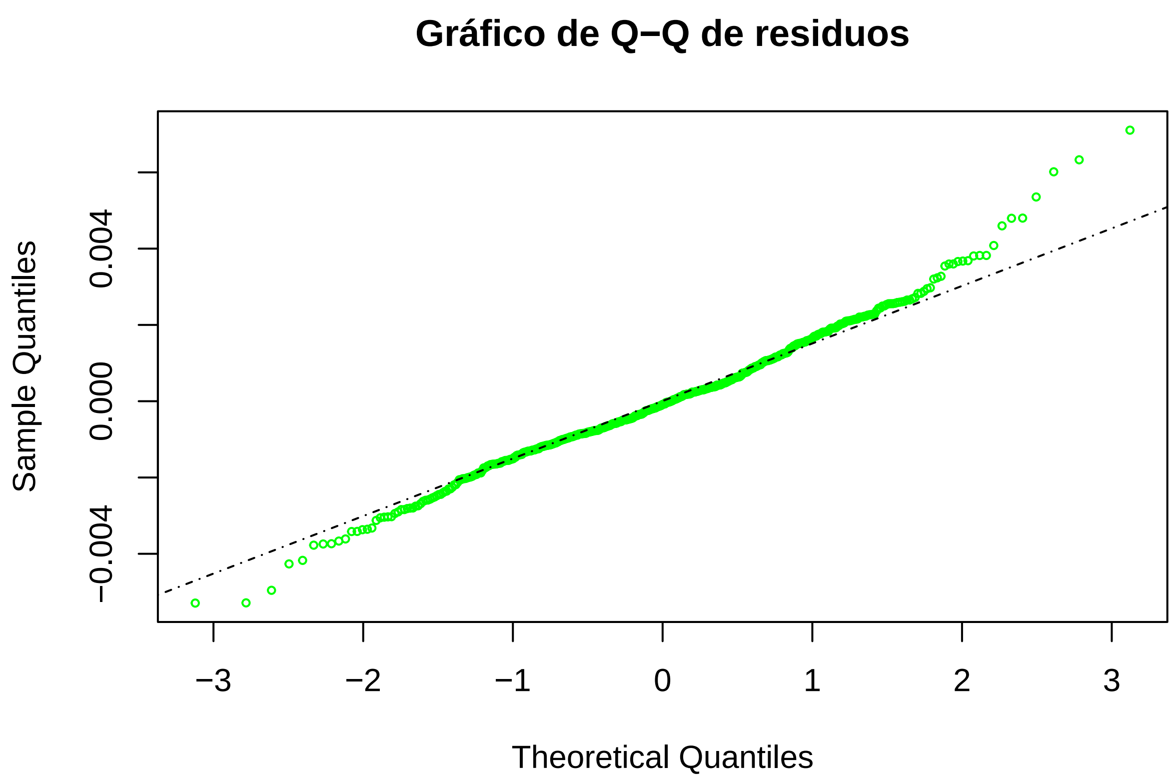
Análisis estadístico de la serie

En principio, realizaremos el análisis estadístico planteando un modelo de regresión panel con efectos fijos, básicamente porque una regresión de efectos fijos permite tener en cuenta variables omitidas que varían entre las entidades individuales pero que no cambian en el tiempo, que en nuestro caso podría ser la cultura de la bebida o cultura del alcohol en cada estado.

```r
# Relizamos la regresión panel con efectos fijos
reg_panel_ef <- plm(fatalityrate ~ sb_useage + speed65 +
                  speed70 + ba08 + drinkage21 + l_income + age,
                  data=S6_cinturon_pl,
                  model="within",
                  cluster="name")
residuos <- resid(reg_panel_ef)
```

Seguidamente, realizamos el gráfico Q-Q (quantile-quantile plot) de los residuos en nuestra regresión de datos panel con efectos fijos. La cual se utiliza para evaluar si los residuos siguen una distribución normal. En tal caso, si los residuos siguen una distribución normal, entonces los puntos en el gráfico Q-Q deben seguir aproximadamente una línea recta.

```r
# Gráfico 5
qqnorm(residuos, col = "green", cex = 0.5, lty = "dashed",
       main = "Gráfico de Q-Q de residuos")
qqline(residuos, lty="dotdash")
```



Podemos ver en nuestro gráfico Q-Q anterior que los residuos tienden a desviarse ligeramente de la línea de 45 grados, especialmente en los extremos de la cola. Esto podría ser una indicación de que posiblmente nuestro datos no esten normalmente distribuidos, quizá por la presencia de heterocedasticidad o algunos datos atípicos. Por lo pronto, es factible que sea necesario realizar ajustes adicionales tales como, utilizar técnicas de regresión robusta para hacer frente a estos inconvenientes. En ese sentido, al realizarlos ajustes al modelo podríamos mejorar su capacidad para explicar los datos.

Planteamiento del Modelo

En este caso, plantearemos inicialmente la estructura de nuestro modelo para su posterior sustenación téorica.

$$fatalityrate_{it} = \beta_0 + \beta_1 sb\_useage_{it} + \beta_2 l\_income_{it} + \beta_3 age_{it} + \\ \gamma_1 speed65_{it} + \gamma_2 speed70_{it} + \gamma_3 ba08_{it} + \gamma_4 drinkage_{it} + u_{it}$$

Donde:

- fatalityrate: número de muertes por millón de millas conducidas.
- sb_useage: proporción de uso del cinturón de seguridad.
- l_income: logaritmo de la renta per cápita.
- speed65: 1 si en el estado hay un límite de velocidad de 65 millas por hora, 0 en otro caso.
- speed70: 1 si en el estado hay un límite de velocidad de 70 millas por hora o superior, 0 en otro caso.
- ba08: 1 si en el estado hay un límite de alcohol en sangre menor o igual al 0,08%, 0 en otro caso.
- drinkage21: 1 si en el estado hay una prohibición de beber alcohol a menores de 21 años, 0 en otro caso.

Modelo de regresión agrupada de datos panel

En el modelo de regresión pooled, se combinan los datos de diferentes grupos o períodos en un solo conjunto de datos y se realiza un análisis de regresión en el conjunto de datos combinado. Según Hiestand(2005), el modelo de regresión agrupado es un tipo de modelo que tiene coeficientes constantes, tanto las intersecciones como a las pendientes. Básicamente, es un modelo que agrupa todas las variables en una sola regresión de tipo lineal. Este modelo se diferencia de otros modelos de regresión longitudinal, como el modelo de efectos aleatorios, en que no tiene en cuenta las diferencias entre los grupos o períodos individuales.

Montero(2011), estima el siguiente modelo:

$$y_{it} = \alpha_i + \beta_1 X_{it} + ... + \beta_n X_{it} + u_{it}$$

Donde, si no se disponen de todas las variables de influencia entonces $COV(X_{it}, \epsilon_{it}) \neq 0$, es decir los residuos no son independientes de las observaciones; por lo tanto, el modelo estara sesgado.

```
## Pooled Regression
reg_pool <- plm(fatalityrate ~ sb_useage + speed65 +
                  speed70 + ba08 + drinkage21 + l_income + age,
                data=S6_cinturon_pl,
                model = "pooling")
```

Para solucionar los problemas de la regresión agrupada se proponen modelos alternativos mediante el anidamiento de los datos: el de efectos fijos y el de efectos aleatorios.

Modelo de regresión de datos panel con efectos fijos

De acuerdo con Stock & Watson(2012), el modelo de regresión de efectos fijos es:

$$Y_{it} = \beta_1 X_{1,it} + ... + \beta_k X_{k,it} + \alpha_i + u_{it}$$

La regresión de efectos fijos es un método que permite tener en cuenta las variables omitidas en datos de panel cuando las variables omitidas varían entre las distintas entidades individuales (estados), pero no cambian en el tiempo. Asimismo, el modelo de regresión de efectos fijos presenta n interceptos diferentes, uno para cada entidad individual. Estos interceptos pueden representarse mediante un conjunto de variables binarias (o indicadores). Estas variables binarias absorben las influencias de todas las variables omitidas que difieren de una entidad individual a otra, pero son constantes en el tiempo.

```
## Fixed Effect Regression
reg_panel_ef <- plm(fatalityrate ~ sb_useage + speed65 +
                  speed70 + ba08 + drinkage21 + l_income + age,
                  model="within",
                  data=S6_cinturon_pl)
```

Modelo de regresión de datos panel con efectos Aleatorios

Montero (2012) indica que, el modelo de efectos aleatorios tiene la misma especificación que el de efectos fijos con la salvedad de que v_i, en lugar de ser un valor fijo para cada individuo y constante a lo largo del tiempo para cada individuo, es una variable aleatoria con un valor medio v_i y una varianza $Var(vi) \neq 0$. En otras palabras, no se tiene certeza del valor exacto en el origen (v_i) que pueda tener cada individuo sino que pensamos que este, probablemente gravitará en torno a un valor central. El modelo planteado es el siguiente:

$$Y_{it} = \alpha + \beta_1 X_{1,it} + ... + \beta_k X_{k,it} + v_i + u_{it}$$

```
#Ramndom Effect Regression
reg_panel_ea <- plm(fatalityrate ~ sb_useage + speed65 +
                      speed70+ba08+drinkage21+l_income+age,
                    data=S6_cinturon_pl,
                    model="random")
```

Correción de errores estándar agrupados

Los errores a heterocedasticidad y correlación consistentes son errores estándar robustos que se utilizan en modelos de regresión para corregir los problemas de heterocedasticidad y correlación serial en los datos. Según Hanck et al.(2021), al igual que para la heteroscedasticidad, la autocorrelación invalida las fórmulas de error estándar habituales, así como los errores estándar robustos a la heteroscedasticidad, ya que estos se derivan bajo el supuesto de que no hay autocorrelación. Cuando hay tanto heteroscedasticidad como autocorrelación, es necesario utilizar los llamados errores estándar de heteroscedasticidad y autocorrelación consistented (HAC).

Como se muestra a continuación, es bastante fácil especificar el uso de errores estándar agrupados en R. Convenientemente, vcovHC() reconoce los objetos del modelo de panel (objetos de la clase plm) y calcula los errores estándar agrupados de forma predeterminada para cada modelo de regresión.

```
rob_se <- list(sqrt(diag(vcovHC(reg_pool, type = "HC1"))),
               sqrt(diag(vcovHC(reg_panel_ef, type = "HC1"))),
               sqrt(diag(vcovHC(reg_panel_ea, type = "HC1"))))
```

Luego de realizar los ajustes necesarios a nuestro modelo, presentamos las regresiones robustas en la tabla 5, a través de los siguientes comandos:

```
st <- stargazer(reg_pool,reg_panel_ef,reg_panel_ea,
          digits = 5,
          header = FALSE,
```

```
          type = "html",
          se = rob_se,
          title = "Modelos de Regresión",
          model.numbers = FALSE,
          column.labels = c("Pooled", "fixed","Random"),
          out = "tabla.html" )
```

```
# Tabla 5
tabla <- readLines("tabla.html")
knitr::asis_output(tabla)
```

Modelos de Regresión

Dependent variable:

fatalityrate

Pooled

fixed

Random

sb_useage

0.00407*

-0.00577***

-0.00444***

(0.00237)

(0.00165)

(0.00163)

speed65

0.00015

-0.00043

-0.00034

(0.00060)

(0.00045)

(0.00044)

speed70

0.00240***

0.00123***

0.00134***

(0.00069)

(0.00034)

(0.00035)

ba08

-0.00192***

-0.00138***

-0.00137***

(0.00073)

(0.00037)

(0.00037)

drinkage21

0.00008

0.00075

0.00076

(0.00126)

(0.00071)

(0.00070)

l_income

-0.01814***

-0.01351***

-0.01263***

(0.00241)

(0.00236)

(0.00196)

age

-0.00001

0.00098

0.00021

(0.00044)

(0.00074)

(0.00051)

Constant

0.19655***

0.13866***

(0.01964)

(0.01674)

Observations

556

556

556

R2

0.54929

0.68679

0.66819

Adjusted R2

0.54353

0.65094

0.66395

F Statistic

95.40697*** (df = 7; 548)

155.99930*** (df = 7; 498)

1,081.33200***

Note:

p<0.1; ***p<0.05;*** p<0.01

Los resultados nos muestran las 3 regresiones realizadas para cada tipo de modelo (agrupado, fijo y aleatorio). A priori, podemos observar un modelo de datos agrupados (pooled) con resultados un poco usuales al sentido común; ya que el coeficiente positivo "sb usege"(Proporcion de uso de cinturon de seguridad por estado), estadísticamente significativo al 10%, básicamente nos indica que a un mayor uso del cinturón de seguridad más muertes por accidentes de tráfico.

Ahora bien, con respecto a las regresiones con efectos fijos y efectos aleatorios el resultado cambia nuestro analisis completamente ya que el efecto del uso de cinturon es negativo. Lo cual indica que a un mayor uso del cinturon de segurida menres muertes por acccidentes de tráfico. Sin embargo, en el análisis de regresión de datos panel, los modelos obtenidos se pueden derivar de varios métodos, por lo que se necesitan más pruebas para elegir el modelo correcto para predecir la regresión.

Comparación de modelos

Modelo de regresion agrupada vs con efectos fijos

- **Test de Chow:**

A continuación, realizamos las pruebas con respecto a la elección del modelo. Como primera prueba realizamos el test de Chow, el cual se lleva a cabo a traves del estadistico F, para probar el efecto fijo o la significación conjunta de los dummies (Yunitaningtyas & Indahwati, 2019). En otras palabras, la prueba F realiza una comparación del modelo de efectos fijos whitin y el modelo agrupado (pooling model).

H_0 : EL modelo pooled es mejor que el de efecto fijo

H_1 : El modelo de efectos fijos es mejor que Pooled

```
# Modelo pooling vs efectos fijos
pFtest(reg_panel_ef,reg_pool)
```

```
##
##  F test for individual effects
##
## data:  fatalityrate ~ sb_useage + speed65 + speed70 + ba08 + drinkage21 +  ...
## F = 29.666, df1 = 50, df2 = 498, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

Tal como podemos observar, con un p-value < 2.2e-16 se rechaza la hipótesis nula y se acepta la hipótesis alternativa, que indica que existe un efecto significativo con respecto alguna de nuestras variables ficticias en el modelo, lo que indica que el modelo de regresion de efectos fijos es mejor que el modelo pooled.

Modelo de regresion con efectos fijos vs efectos aleatorios

- **Test de Hausmann:**

Cuando la prueba de Chow muestra que el modelo correspondiente es de efectos fijos, el siguiente paso es realizar las pruebas de Hausmann. La prueba de Hausman es una prueba estadística para seleccionar si es más apropiado el modelo de efectos fijos o de efectos aleatorios. Con respecto a las pruebas de hipótesis, Woldridge (2010), menciona que un rechazo mediante la prueba de Hausman significa que el supuesto clave de efectos aleatorios es falso y por tanto, se usan las estimaciones de efectos fijos.

$H_0 : Corr(X_{it}, u_{it}) = 0$

$H_1 : Corr(X_{it}, u_{it}) \neq 0$

```r
# Modelo fijos vs efectos aleatorios
phtest(reg_panel_ef, reg_panel_ea)
```

```
##
##  Hausman Test
##
## data:  fatalityrate ~ sb_useage + speed65 + speed70 + ba08 + drinkage21 +  ...
## chisq = 31.953, df = 7, p-value = 4.144e-05
## alternative hypothesis: one model is inconsistent
```

Los resultados nos muestran que con un valor p = 4.144e-05 se rechaza la hipótesis nula de exogeneidad de los efectos fijos y se acepta la hipótesis alterna, que indica que existe correlación entre el error y los regresores, y es preferible elegir el modelo de efectos fijos.

Modelo de efectos fijos con efectos individuales y temporales

- **Efectos fijos individuales y temporales de la variable Falityrate:**

De acuerdo con, Stock & Watson(2012), si algunas variables omitidas son constantes en el tiempo pero varían entre los estados (como por ejemplo, las normas culturales), mientras que otras son constantes entre los estados pero varían en el tiempo (como por ejemplo, los estándares nacionales de seguridad), entonces resulta apropiado incluir efectos tanto individuales (para los estados) como temporales. En otras palabras, tener en cuentas estos efectos nos permiten eliminar el sesgo de variable omitida derivado de variables constantes para todas las entidades individuales correlacionados con los regresores.

En ese sentido, el modelo combinado de regresión de efectos fijos individuales y temporales es:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + ... + \gamma_n Dn_i + \delta_2 B2_t + ... + \delta_t BT_t + u_{it}$$

Ahora bien, ya que sabemos que nuestro modelo de efectos fijos es significativo, podemos ver como cambíam los resultados cuando se agregan efectos fijos temporales más los efectos fijos individuales de cada estado a nuestra regresión de efectos fijos.

```r
reg_panel_eft <- plm(fatalityrate ~ sb_useage + speed65 +
                  speed70 + ba08 + drinkage21 + l_income + age + factor(year),
                  model="within",
                  data=S6_cinturon_pl)
sum3 <- summary(reg_panel_eft, vcov = vcovHC(reg_panel_eft, type="HC1"))
sum3
```

```
## Oneway (individual) effect Within Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(reg_panel_eft, type = "HC1")
##
## Call:
## plm(formula = fatalityrate ~ sb_useage + speed65 + speed70 +
##     ba08 + drinkage21 + l_income + age + factor(year), data = S6_cinturon_pl,
##     model = "within")
##
## Unbalanced Panel: n = 51, T = 8-15, N = 556
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -0.00524805 -0.00082045 -0.00012876  0.00072230  0.00735905
##
## Coefficients:
##                     Estimate  Std. Error t-value Pr(>|t|)
## sb_useage        -0.00371856  0.00143713 -2.5875 0.009958 **
## speed65          -0.00078331  0.00057433 -1.3639 0.173243
## speed70           0.00080417  0.00045271  1.7763 0.076306 .
## ba08             -0.00082249  0.00043888 -1.8741 0.061525 .
## drinkage21       -0.00113368  0.00061590 -1.8407 0.066280 .
## l_income          0.00626437  0.00663292  0.9444 0.345418
## age               0.00131804  0.00068685  1.9189 0.055579 .
## factor(year)1984 -0.00043192  0.00136441 -0.3166 0.751716
## factor(year)1985 -0.00107066  0.00174666 -0.6130 0.540180
## factor(year)1986 -0.00057773  0.00198798 -0.2906 0.771475
## factor(year)1987 -0.00087222  0.00246919 -0.3532 0.724063
## factor(year)1988 -0.00188500  0.00284854 -0.6617 0.508450
## factor(year)1989 -0.00417658  0.00322420 -1.2954 0.195805
## factor(year)1990 -0.00526596  0.00350523 -1.5023 0.133668
## factor(year)1991 -0.00666225  0.00372216 -1.7899 0.074096 .
## factor(year)1992 -0.00851800  0.00394613 -2.1586 0.031375 *
## factor(year)1993 -0.00893993  0.00415746 -2.1503 0.032024 *
## factor(year)1994 -0.00962971  0.00454800 -2.1173 0.034740 *
## factor(year)1995 -0.01011234  0.00484768 -2.0860 0.037500 *
## factor(year)1996 -0.01107658  0.00515738 -2.1477 0.032232 *
## factor(year)1997 -0.01160753  0.00547938 -2.1184 0.034650 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    0.005078
## Residual Sum of Squares: 0.0012663
```

```
## R-Squared:      0.75062
## Adj. R-Squared: 0.71404
## F-statistic: 53.3537 on 21 and 50 DF, p-value: < 2.22e-16
```

Como vemos, cuando incluimos en la regresión los efectos temporales. Es decir, variables no observables, pero constantes para las entidades individuales que evolucionan en el tiempo, vemos que el efecto sigue del uso de cinturón de seguridad sigue siendo negativo con respecto a nuestra variable de fatalidad por millón de millas conducidas. Sin embargo, el efecto es algo menor y significativo solo al 5%.

- **Test de Wald:**

El contraste conjunto de Wald es una prueba estadística que se utiliza para evaluar la significancia conjunta de un conjunto de parámetros en un modelo de regresión. En el caso específico de los modelos de datos panel, el contraste conjunto de Wald se puede utilizar para evaluar la significancia conjunta de las variables ficticias temporales, las cuales se utilizan para controlar el efecto del tiempo en el modelo. Esto es importante porque las variables ficticias temporales añadidas a nuestro modelo se utilizan para controlar el efecto del tiempo en el modelo, y si no tienen un efecto conjunto significativo en el modelo, entonces es posible que el modelo con efectos fijos individuales y temporales no sea el correcto.

H_0 : Sin efectos temporales

H_1 : Con efectos temporales

```
waldtest(reg_panel_eft, formula ="factor(year)")
```

```
## Wald test
##
## Model 1: fatalityrate ~ sb_useage + speed65 + speed70 + ba08 + drinkage21 +
##     l_income + age + factor(year)
## Model 2: fatalityrate ~ sb_useage + speed65 + speed70 + ba08 + drinkage21 +
##     l_income + age
##   Res.Df  Df  Chisq Pr(>Chisq)
## 1    484
## 2    498 -14 123.89  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Como podemos observar, los efectos temporales son estadísticamente significativos. Por lo Tanto, la inclusión de efectos temporales a nuestro modelo de efectos fijos parece la más fiable.

- **Efectos fijos individuales y temporales de la variable sb_usage:**

Ahora con respecto a los efectos de las leyes obligatorios de uso de cinturon de seguridad tomamos que existen dos maneras de aplicar las leyes de obligatoriedad del uso del cinturón de seguridad: la primaria, mediante la cual la aplicación significa que un oficial de policía puede detener un coche y multar al conductor si el oficial observa que un ocupante no lleva puesto el cinturón de seguridad, y la secundaria, mediante la cual la aplicación significa que un oficial de policía puede poner una multa si un ocupante no lleva puesto el cinturón de seguridad, pero debe de existir otra razón para poder detener el coche. Esto está recogido en las variables dummy primary y secondary.

$$sb\_usage_{it} = \gamma_1 primary_{it} + \gamma_2 secondary_{it} + \gamma_1 speed65_{it} + \gamma_2 speed70_{it}$$
$$+\gamma_3 ba08_{it} + \gamma_4 drinkage_{it} + \beta_3 l\_incomeit + \beta_3 age_{it} + \alpha_{it} + factor(year)$$

A continuación, realizamos una regresión de la variable sb_usage sobre las variables primary, secondary, speed65, speed70, ba08, drinkage21, log(income) y age, incluyendo efectos fijos individuales de estado y temporales en la regresión.

El resultado de la estimación es:

```
reg_panel_eft_2 <- plm(sb_useage ~  + primary + secondary + speed65 +
                       speed70 + ba08 + drinkage21 + l_income +
                       age + factor(year),
                     model="within",
                     data=S6_cinturon_pl)
sum4 <- summary(reg_panel_eft_2, vcov = vcovHC(reg_panel_eft_2, type="HC1"))
sum4
```

```
## Oneway (individual) effect Within Model
##
## Note: Coefficient variance-covariance matrix supplied: vcovHC(reg_panel_eft_2, type = "HC1")
##
## Call:
## plm(formula = sb_useage ~ +primary + secondary + speed65 + speed70 +
##     ba08 + drinkage21 + l_income + age + factor(year), data = S6_cinturon_pl,
##     model = "within")
##
## Unbalanced Panel: n = 51, T = 8-15, N = 556
##
## Residuals:
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -0.2777235 -0.0306088 -0.0010068  0.0310643  0.2459545
##
## Coefficients:
##                   Estimate Std. Error t-value  Pr(>|t|)
## primary          0.2055968  0.0229495  8.9587 < 2.2e-16 ***
## secondary        0.1085184  0.0132763  8.1739 2.646e-15 ***
## speed65          0.0228486  0.0203142  1.1248   0.26125
## speed70          0.0120424  0.0203881  0.5907   0.55503
## ba08             0.0037584  0.0174433  0.2155   0.82950
## drinkage21       0.0107149  0.0269021  0.3983   0.69059
## l_income         0.0582733  0.2539047  0.2295   0.81857
## age              0.0138232  0.0229049  0.6035   0.54646
## factor(year)1984 0.0041175  0.0282650  0.1457   0.88424
## factor(year)1985 0.0575165  0.0426133  1.3497   0.17773
## factor(year)1986 0.1073522  0.0545727  1.9671   0.04974 *
## factor(year)1987 0.1240640  0.0763541  1.6249   0.10485
## factor(year)1988 0.1390916  0.0966453  1.4392   0.15074
## factor(year)1989 0.1702315  0.1118603  1.5218   0.12871
## factor(year)1990 0.1897742  0.1280015  1.4826   0.13883
## factor(year)1991 0.2370685  0.1357108  1.7469   0.08130 .
## factor(year)1992 0.2633958  0.1507081  1.7477   0.08115 .
## factor(year)1993 0.2824178  0.1618973  1.7444   0.08172 .
## factor(year)1994 0.2983707  0.1721161  1.7335   0.08364 .
## factor(year)1995 0.2959065  0.1834495  1.6130   0.10739
## factor(year)1996 0.2875623  0.1966614  1.4622   0.14433
## factor(year)1997 0.2977333  0.2082343  1.4298   0.15342
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    10.012
## Residual Sum of Squares: 1.5817
## R-Squared:      0.84201
## Adj. R-Squared: 0.81846
## F-statistic: 513.49 on 22 and 50 DF, p-value: < 2.22e-16
```

Como podemos observar, los efectos de las leyes obligarias de uso de cinturon de seguridad (primary y secondary) son positivos y significativos al 1%; sin embargo, el efecto de la aplicación primaria es aproximadamente el doble que de la aplicación secundaria. Lo cual nos indica que, hay un efecto mayor de la aplicación primaria con respecto al uso de cinturón de seguridad, en la que un oficial de policía podía detener un coche y multar al conductor si el oficial observaba que un ocupante no llevaba puesto el cinturón de seguridad.

Interpretación de la regresión

- **Efectos en la tasa de fatalidad**

En este apartado realizaremos la interpretación de la regresión de panel con efectos fijos y temporales que incluye a la variable dependiente "fatalityrate", que es el número de muertes por millón de millas conducidas y Los coeficientes de las variables explicativas que influyen en la tasa de mortalidad.

```
# Tabla 6
knitr::kable(
  sum3$coefficients,
  digits = 5,
  align = "c"
)
```

	Estimate	Std. Error	t-value	Pr(>|t|)
sb_useage	-0.00372	0.00144	-2.58750	0.00996
speed65	-0.00078	0.00057	-1.36387	0.17324
speed70	0.00080	0.00045	1.77633	0.07631
ba08	-0.00082	0.00044	-1.87406	0.06153
drinkage21	-0.00113	0.00062	-1.84068	0.06628
l_income	0.00626	0.00663	0.94444	0.34542
age	0.00132	0.00069	1.91895	0.05558
factor(year)1984	-0.00043	0.00136	-0.31656	0.75172
factor(year)1985	-0.00107	0.00175	-0.61298	0.54018
factor(year)1986	-0.00058	0.00199	-0.29061	0.77147
factor(year)1987	-0.00087	0.00247	-0.35324	0.72406
factor(year)1988	-0.00189	0.00285	-0.66174	0.50845
factor(year)1989	-0.00418	0.00322	-1.29538	0.19581
factor(year)1990	-0.00527	0.00351	-1.50231	0.13367
factor(year)1991	-0.00666	0.00372	-1.78989	0.07410
factor(year)1992	-0.00852	0.00395	-2.15857	0.03137
factor(year)1993	-0.00894	0.00416	-2.15033	0.03202
factor(year)1994	-0.00963	0.00455	-2.11735	0.03474
factor(year)1995	-0.01011	0.00485	-2.08602	0.03750
factor(year)1996	-0.01108	0.00516	-2.14772	0.03223

	Estimate	Std. Error	t-value	Pr(>|t|)
factor(year)1997	-0.01161	0.00548	-2.11840	0.03465

En primer lugar, podemos observar que el coeficiente de la variable "sb_useage" sugieren que un mayor uso del cinturón de seguridad está relacionado con una disminución en la tasa de mortalidad. Es decir, un aumento del 1% en la proporción de uso del cinturón de seguridad se asocia con una disminución de 0.0037 en la tasa de mortalidad.

Asimismo, el coeficiente de las variable "speed65" sugieren que no es significativamente estadísticos en la tasa de mortalidad. En cambio, el coeficiente de las variable "speed70", indica que si en el estado hay un límite de velocidad de 70 millas por hora o superior reduce en 0.0008 el ratio de fatalidad. Por otro lado, el coeficiente de la variable "ba08" indica que un límite de alcohol en sangre menor o igual al 0,08% se asocia con una disminución de la tasa de mortalidad, aunque la significancia estadística no es muy alta. Por su parte, el coeficiente de la variable "drinkage21" sugiere que una prohibición de beber alcohol a menores de 21 tiene una relación negativa con la tasa de mortalidad, que corresponde a un disminucion de 0.0011 en las muertes millón de millas conducidas.

Tambien se observa que, el coeficiente de la variable "l_income" no es significativo, lo que sugiere que el nivel de ingresos per cápita no tiene una relación significativa con la tasa de mortalidad. No así, los coeficientes de las variables "age" y los efectos temporales sugieren que hay ciertos años en los que la tasa de mortalidad es significativamente más alta que en otros años.

- **Efectos en el uso del cinturon de seguridad**

Seguidamente, realizamos la interpretación de la regresión de panel con efectos fijos y temporales que incluye a la variable dependiente "fatalityrate", que es la proporción de uso del cinturón de seguridad sobre las variables de aplicación primaria y secundaria, incluyendo a las demás variables.

```r
# Tabla 7
knitr::kable(
  sum4$coefficients,
  digits = 5,
  align = "c"
)
```

	Estimate	Std. Error	t-value	Pr(>|t|)
primary	0.20560	0.02295	8.95867	0.00000
secondary	0.10852	0.01328	8.17386	0.00000
speed65	0.02285	0.02031	1.12476	0.26125
speed70	0.01204	0.02039	0.59066	0.55503
ba08	0.00376	0.01744	0.21547	0.82950
drinkage21	0.01071	0.02690	0.39829	0.69059
l_income	0.05827	0.25390	0.22951	0.81857
age	0.01382	0.02290	0.60350	0.54646
factor(year)1984	0.00412	0.02827	0.14568	0.88424
factor(year)1985	0.05752	0.04261	1.34973	0.17773
factor(year)1986	0.10735	0.05457	1.96714	0.04974
factor(year)1987	0.12406	0.07635	1.62485	0.10485
factor(year)1988	0.13909	0.09665	1.43920	0.15074
factor(year)1989	0.17023	0.11186	1.52182	0.12871
factor(year)1990	0.18977	0.12800	1.48259	0.13883

	Estimate	Std. Error	t-value	Pr(>|t|)
factor(year)1991	0.23707	0.13571	1.74687	0.08130
factor(year)1992	0.26340	0.15071	1.74772	0.08115
factor(year)1993	0.28242	0.16190	1.74443	0.08172
factor(year)1994	0.29837	0.17212	1.73354	0.08364
factor(year)1995	0.29591	0.18345	1.61301	0.10739
factor(year)1996	0.28756	0.19666	1.46222	0.14433
factor(year)1997	0.29773	0.20823	1.42980	0.15342

Los resultados indican que la ley de obligatoriedad de uso del cinturón de seguridad de forma básica (primary) y secundaria (secondary), están significativamente asociadas con un mayor uso del cinturón de seguridad (sb_usage). Donde la aplicación primaria es la más determinante, de acuerdo con nuestro modelo. También podemos observar que la ley de límite de alcohol en sangre menor o igual al 0,08% (ba08) no está significativamente relacionada con el uso del cinturón de seguridad, mientras que la prohibición de beber alcohol a menores de 21 años (drinkage21) tampoco tiene una relación significativa en nuestro modelo.

Respecto a los límites de velocidad, el límite de velocidad de 65 millas por hora (speed65) no tiene un efecto significativo en el uso del cinturón de seguridad, mientras que el límite de velocidad de 70 millas por hora o superior (speed70) tampoco está significativamente relacionado con el uso del cinturón de seguridad.

Por último, en cuanto a los efectos temporales, los coeficientes indican que el uso del cinturón de seguridad ha aumentado significativamente a lo largo de los años de estudio, pero solo en los años 1986, 1991 y 1992.

Evaluación del modelo

Prueba de Normalidad

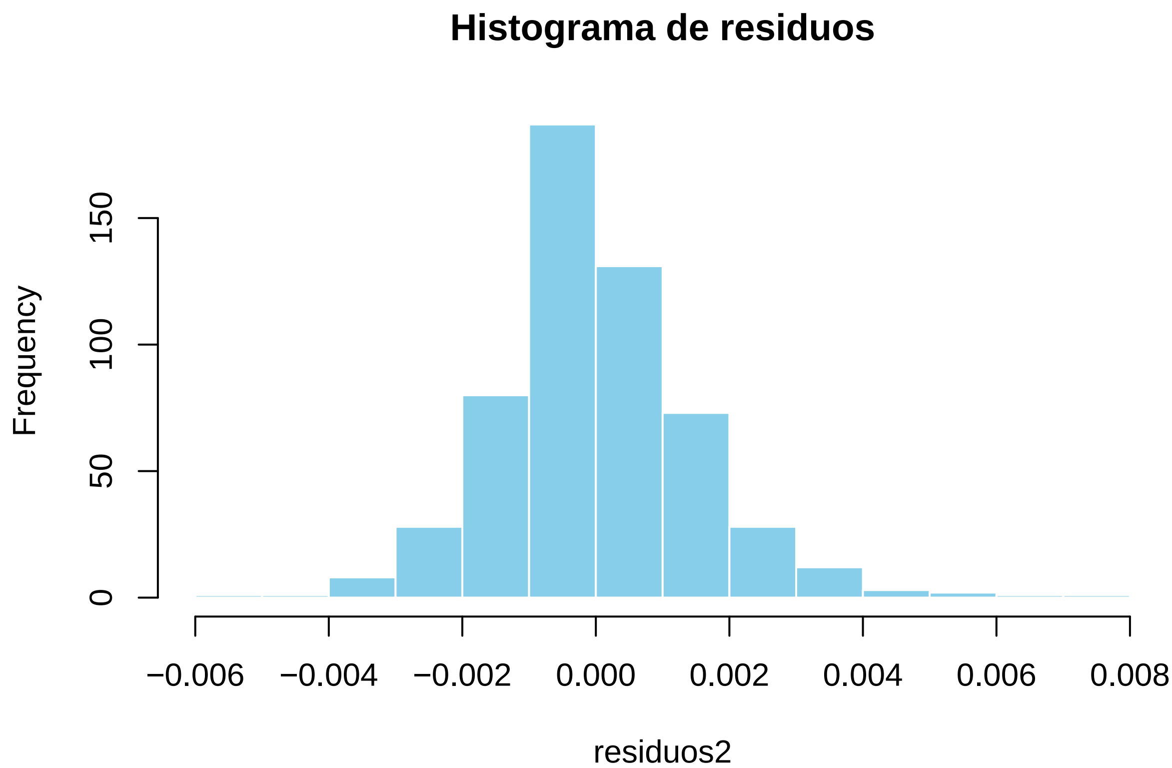
Para realizar la evaluación de nuestro modelo realizaremos las distintas pruebas validación, pero hay que tener en cuenta que los modelos de panel pueden ser más complicados que los modelos estándar de regresión.

Como primer punto realizamos la prueba de bondad de ajuste llamada prueba Kolmogorov-Smirnov, la cual se utiliza para evaluar si los residuos de un modelo se ajustan a una distribución normal. En este caso, la prueba sugiere que los residuos del modelo no se ajustan a una distribución normal.

```
residuos2 <- as.vector(resid(reg_panel_eft, vcov = vcovHC(reg_panel_eft, type="HC1")))
ks.test(residuos2, "pnorm")
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  residuos2
## D = 0.49791, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```
hist(residuos2, col = "skyblue", border = "white", main = "Histograma de residuos")
```



Estos problemas son usuales en los modelos de datos panel ya que, los efectos fijos son variables dummy que representan las diferencias entre las unidades (estados) que se mantienen constantes a lo largo del tiempo. Estas variables son específicas de cada unidad y no varían con el tiempo, lo que las hace diferentes de las variables independientes que se utilizan para explicar la variación en la variable dependiente. Asimismo, al incluir efectos fijos al modelo para controlar las características inobservables de las unidades que no cambian con el tiempo, tiene un impacto en la variable dependiente; ya que las variables independientes se correlacionan con los efectos fijos, lo que viola la suposición de normalidad de los residuos.

Pruebas de heterocedasticidad

Seguidamente, realizaremos la prueba de heterocedasticidad de Breusch-Pagan, la cual es una prueba estadística utilizada para detectar si los errores de un modelo de datos panel con efectos fijos presentan heterocedasticidad, es decir, si la varianza de los errores no es constante en todas las observaciones.

En general, para esta prueba se recomienda estandarizar los residuos antes de realizar la prueba de heterocedasticidad de Breusch-Pagan utilizando la opción "studentize = T" (que es la opción predeterminada en la función "bptest").

```
# Pruebas de Heterocedasticidad
#Pruebas de Heterocedasticidad
bptest(reg_panel_eft, studentize = T)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  reg_panel_eft
```

```
## BP = 25.549, df = 21, p-value = 0.2242
```

Ahora bien, dado que el valor de p es mayor que el nivel de significancia típico de 0.05, no se rechaza la hipótesis nula de que no hay heterocedasticidad en los residuos del modelo de regresión. En otras palabras, no hay suficiente evidencia estadística para afirmar que hay heterocedasticidad en los residuos del modelo. Por lo tanto, se puede asumir que la varianza de los errores es constante en todo el rango de valores de la variable independiente.

Conclusiones

Basado en la literatura y la evidencia empírica disponible, se ha demostrado que existe una relación inversa entre el uso de cinturón de seguridad y la tasa de accidentes automovilísticos. Los estudios indican que el uso del cinturón de seguridad reduce significativamente la probabilidad de sufrir lesiones graves o fatales en caso de accidente. Por lo tanto, se puede concluir que el uso del cinturón de seguridad es una medida efectiva para prevenir lesiones y salvar vidas en caso de accidentes automovilísticos. Es importante que los conductores y pasajeros adopten esta práctica como un hábito regular, ya que esto puede tener un impacto significativo en la seguridad en las carreteras.

En general, estos resultados sugieren que la promoción del uso del cinturón de seguridad puede ser una medida efectiva para reducir la tasa de mortalidad en accidentes de tránsito, mientras que los límites de velocidad más altos pueden tener un efecto opuesto. Asimismo, la obligatoriedad de uso del cinturón de seguridad de forma básica y secundaria está significativamente asociada con un mayor uso del cinturón de seguridad. Por último, es importante tener en cuenta que los resultados pueden variar dependiendo del contexto y otras variables no incluidas en el modelo.

Referencias

Eynav, H. & Cohen, J. (2003). The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities. The Review of Economics and Statistics, 2003, Vol. 85, pp 828-843.

Hanck, C., Arnold, M., & Schmelzer, M. (2021). Introduction to Econometrics with R.

Hiestand, T. (2005). Using Pooled Model, Random Model And Fixed Model Multiple Regression To Measure Foreign Direct Investment In Taiwan.Clute Institute.

Organización Mundial de la Salud. (2014). Road Traffic Accident Statistics. Geneva: WHO.

Montero, R. (2011). Efectos fijos o aleatorios: test de especificación. Universidad de Granada. España.

Stock, J. H., & Watson, M. W. (2012). Introducción a la Econometría (3.ª ed.). Pearson Educación.

Wooldridge, J. (2010). Introducción a la econometría. Un enfoque moderno, 4a. edición.

Yunitaningtyas, K & Indahwati, Y. (2019). A panel data analysis of tourism and economic development in Southeast Asian countries. Department of Statistics, IPB University.