

Econometría Avanzada

Primer Examen Parcial

Presentado por: Giovany Astroz y Javier leal

12 marzo 2021

1. Estimación de la demanda de cerveza

1.1. Análisis de regresión

a. Cree una nueva variable llamada “marca” que contenga la marca de cada UPC de cerveza (por ejemplo, Heineken).

Al analizar la base de datos se encontró que la marca de cada UPC está contenida dentro de la variable *descrip*, sin embargo, dentro de esta variable también se encuentran características adicionales que hacen referencia principalmente a la variedad, presentación y contenido. Ahora bien, estas especificaciones adicionales se encuentran siempre luego de la marca, razón por la cual se decidió que una primera aproximación para encontrar coincidencias sería extraer la primera palabra de la descripción de cada UPC. Siguiendo esta aproximación se encuentra que existen 172 palabras diferentes que clasifican *a priori* las 787 observaciones de la base de datos.

Claramente, este acercamiento no proporciona una clasificación perfecta, pues existen marcas distintas que empiezan por la misma palabra. Por lo tanto, se realizó una verificación manual para identificar las marcas cuyos UPC eran correctamente agrupados bajo este criterio y las que no. Durante este proceso se encontraron más problemas con los datos que se mencionarán más adelante.

Luego de reconocer todas las limitaciones y correcciones que se debían hacer, se creó la variable *marca* de la siguiente manera:

1. Se crea la variable *marca* extrayendo la primera palabra de la variable *descrip*. Aunque identificar la marca por la primera palabra no es suficiente en muchos casos, sí es una buena aproximación inicial.
2. Se diferencian las marcas que coinciden en la primera palabra. Por ejemplo, bajo la palabra “SAMUEL” quedaban agrupadas dos marcas diferentes: Samuel Smith y Samuel Adams.
3. Se unifican bajo el mismo nombre marcas que inicialmente quedaron separadas por omisión de caracteres especiales, abreviaturas o errores tipográficos en el nombre. Por ejemplo, la marca Murphy’s Irish Stout se encontraba escrita con y sin apóstrofo: “MURPHY’S” y “MURPHYS”; la marca Leinenkugel se escribía también bajo la abreviatura “LEINNKGL” y la marca Guinness se encontraba incorrectamente escrita como “GUINESS”.
4. Se cambia la primera palabra por el nombre completo para todas las marcas con al menos dos observaciones y para aquellas con una sola observación en que se consideró necesario. Por ejemplo, se sustituye la palabra “SIERRA” por “SIERRA NEVADA”. Si bien es cierto que estas marcas ya se encontraban correctamente identificadas, se prefirió poner el nombre completo para tener una base de datos más robusta.
5. Se eliminan de la base de datos los UPC que corresponden a marcas que no son de cerveza (vinos, sidra, bebidas sustitutas). Por ejemplo, se identifica que las variedades de la marca Sante (hardonnay, white, cabernet) corresponden a diferentes tipos de vino, y que la marca Hard Core produce sidra, una bebida alcohólica producto de la fermentación de la manzana. Específicamente, se eliminaron de la base los UPC asociados a las marcas Sante, Ariel, Sutter Home, St Regis, Hard Core, Hornsby’s, Santa, Zima, Woodchuck y Woodpecker.

6. En concordancia con el punto 1.1.c., se eliminan aquellos UPC para los cuales no es posible inferir la marca y además, aun ignorando el hecho de que no se conocía la marca, tampoco se podía determinar su “nacionalidad”. Por ejemplo, la descripción para el UPC 1820000875 es simplemente “WINTER BREW”, la cual hace referencia a una variedad de cerveza que es común en varias marcas. A modo de ilustración, esta podría pertenecer a la marca Pete’s Wicked (americana) o a Fuller’s (inglesa). Dado lo anterior, como tampoco se podía definir si se trataba de una cerveza nacional o importada se decidió eliminar esta observación. Así mismo, se eliminaron las observaciones con las descripciones “WIT”, “BEER LIMIT”, “BEERS OF THE WORLD” y “OS CLASSIC”, las cuales no aportaban información suficiente para contruir las variables *importada* y *artesanal*.

Es conveniente mencionar que existían casos que aplicaban a más de una de las correcciones indicadas en los pasos 2, 3 y 4. Por ejemplo, bajo la palabra “SAM” se encontraban dos marcas diferentes: Sam Smith y Sam Adams (paso 2); precisamente, “SAM SMITH” y “SAM ADAMS” correspondían a la abreviatura de Samuel Smith y Samuel Adams (paso 3); y finalmente, las marcas señalas contaban con nombres de dos palabras (paso 4). En estos casos, la corrección se realizó en un solo paso y se puede encontrar en cualquiera de estos.

Por otro lado, habían marcas con nombres diferentes pero pertenecientes a la misma empresa en donde no era muy claro si debían tomarse como una sola o no. Por ejemplo, las marcas Ice Draft Light y Bud Ice son propiedad al igual que Budweiser de la multinacional Anheuser-Busch InBev. A pesar de que las dos primeras compartían elementos de identificación y construcción de marca con Budweiser (el vocablo Bud y la referencia a Budweiser en la presentación del producto) se decidió dejarlas como marcas diferentes. Para cada caso similar se tomó la decisión de agrupar marcas o no con base en las estrategias de marketing de las empresas y las diferencias en los productos mismos.

Finalmente, la base *beer* contiene la nueva variable *marca* con todas las correcciones mencionadas anteriormente, donde ahora se tienen 753 observaciones y 150 marcas diferentes de cerveza.

```
ls()
remove(list = ls())

# Cargando paquetes
library(readr)
library(rmarkdown)
library(tidyverse)
library(stringr)
library(tibble)
library(dplyr)
library(tokenizers)
library(writexl)
library(expss)
library(ggplot2)
library(stargazer)

# Definiendo directorio de trabajo
setwd("~/Documents/Maestria/Primer semestre/Econometría Avanzada/Parcial 1")

# Cargando la base de datos
beer <- read_csv("beer.csv")

# Primera aproximación de clasificación utilizando la primera palabra de "descrip"

prim_palabra_descrip <- word(beer$descrip, 1)
frec_prim_palabra <- sort(table(prim_palabra_descrip), decreasing=T)
head(frec_prim_palabra, 10)
```

```

#-----#
# Creando la variable *marca* #
#-----#

# 1. Extraer la primera palabra de la variable "descrip"
beer$marca <- word(beer$descrip, 1)

# Correcciones:

# 2. Marcas distintas que coinciden en la primera palabra
beer$marca[str_detect(beer$descrip, "^OLD STYLE")]<-"OLD STYLE"
beer$marca[str_detect(beer$descrip, "^OLD MILWAUKEE")]<-"OLD MILWAUKEE"
beer$marca[str_detect(beer$descrip, "^OLD ENGLISH")]<-"OLD ENGLISH"
beer$marca[str_detect(beer$descrip, "^SAMUEL SMITH")]<-"SAMUEL SMITH"
beer$marca[str_detect(beer$descrip, "^SAMUEL ADAMS")]<-"SAMUEL ADAMS"
beer$marca[str_detect(beer$descrip, "^SAM SMITH")]<-"SAMUEL SMITH"
beer$marca[str_detect(beer$descrip, "^SAM ADAMS")]<-"SAMUEL ADAMS"
beer$marca[str_detect(beer$descrip, "^ST REGIS")]<-"ST REGIS"
beer$marca[str_detect(beer$descrip, "^ST PAULI")]<-"ST PAULI GIRL"
beer$marca[str_detect(beer$descrip, "^BLUE MOON")]<-"BLUE MOON"
beer$marca[str_detect(beer$descrip, "^BLUE RIDGE")]<-"BLUE RIDGE"
beer$marca[str_detect(beer$descrip, "^RED WOLF")]<-"RED WOLF"
beer$marca[str_detect(beer$descrip, "^RED DOG")]<-"RED DOG"
beer$marca[str_detect(beer$descrip, "^RED STRIPE")]<-"RED STRIPE"
beer$marca[str_detect(beer$descrip, "^RED HOOK")]<-"RED HOOK"
beer$marca[str_detect(beer$descrip, "^RED RIVER")]<-"RED RIVER"
beer$marca[str_detect(beer$descrip, "^BEER LIMIT")]<-"BEER LIMIT"
beer$marca[str_detect(beer$descrip, "^BEER OF AMERICA")]<-"BEER OF AMERICA"
beer$marca[str_detect(beer$descrip, "^BEERS OF THE")]<-"BEERS OF THE WORLD"
beer$marca[str_detect(beer$descrip, "^BIG BEAR")]<-"BIG BEAR"
beer$marca[str_detect(beer$descrip, "^BIG SHOULDERS")]<-"BIG SHOULDERS"

# 3. Misma marca con nombres diferentes
beer$marca[str_detect(beer$marca, "^PETES$")]<-"PETE'S WICKED"
beer$marca[str_detect(beer$marca, "^PETE'S$")]<-"PETE'S WICKED"
beer$marca[str_detect(beer$marca, "^MURPHYS$")]<-"MURPHY'S IRISH"
beer$marca[str_detect(beer$marca, "^MURPHY'S$")]<-"MURPHY'S IRISH"
beer$marca[str_detect(beer$marca, "^FOSTERS$")]<-"FOSTER'S LAGER"
beer$marca[str_detect(beer$marca, "^FOSTER'S$")]<-"FOSTER'S LAGER"
beer$marca[str_detect(beer$marca, "^GUINNESS$")]<-"GUINNESS STOUT"
beer$marca[str_detect(beer$marca, "^GUINNESS$")]<-"GUINNESS STOUT"
beer$marca[str_detect(beer$marca, "^BECKS$")]<-"BECK'S"
beer$marca[str_detect(beer$marca, "^HACKER-PSCHORR$")]<-"HACKER PSCHORR"
beer$marca[str_detect(beer$marca, "^PSCHORR$")]<-"HACKER PSCHORR"
beer$marca[str_detect(beer$marca, "^STROHS$")]<-"STROH'S"
beer$marca[str_detect(beer$marca, "^FOSTERS$")]<-"FOSTER'S"
beer$marca[str_detect(beer$marca, "^LEINNKGL$")]<-"LEINENKUGEL"
beer$marca[str_detect(beer$marca, "^LABATT$")]<-"LABATTS"
beer$marca[str_detect(beer$marca, "^MICHEAL$")]<-"MICHAEL SHEA'S"
beer$marca[str_detect(beer$marca, "^MICHAEL$")]<-"MICHAEL SHEA'S"
beer$marca[str_detect(beer$marca, "^SHARPS$")]<-"MILLER"
beer$marca[str_detect(beer$marca, "^CORONITA$")]<-"CORONA"
beer$marca[str_detect(beer$marca, "^CHICAGO'S$")]<-"LEGACY"

```

```

# 4. Marcas con nombres de más de una palabra
beer$marca[str_detect(beer$marca, "^SPECIAL$")]<-"SPECIAL EXPORT"
beer$marca[str_detect(beer$marca, "^GOOSE$")]<-"GOOSE ISLAND"
beer$marca[str_detect(beer$marca, "^MILWAUKEE'S$")]<-"MILWAUKEE'S BEST"
beer$marca[str_detect(beer$marca, "^CARLINGS$")]<-"CARLING BLACK LABEL"
beer$marca[str_detect(beer$marca, "^DOS$")]<-"DOS EQUIS"
beer$marca[str_detect(beer$marca, "^NEW$")]<-"NEW AMSTERDAM"
beer$marca[str_detect(beer$marca, "^ROLLING$")]<-"ROLLING ROCK"
beer$marca[str_detect(beer$marca, "^BLACK$")]<-"BLACK DOG"
beer$marca[str_detect(beer$marca, "^CARTA$")]<-"CARTA BLANCA"
beer$marca[str_detect(beer$marca, "^MEISTER$")]<-"MEISTER BRAU"
beer$marca[str_detect(beer$marca, "^NAKED$")]<-"NAKED ASPEN"
beer$marca[str_detect(beer$marca, "^RHINO$")]<-"RHINO CHASERS"
beer$marca[str_detect(beer$marca, "^SIERRA$")]<-"SIERRA NEVADA"
beer$marca[str_detect(beer$marca, "^STATE$")]<-"STATE STREET"
beer$marca[str_detect(beer$marca, "^COLD$")]<-"COLD SPRING"
beer$marca[str_detect(beer$marca, "^COLT$")]<-"COLT 45"
beer$marca[str_detect(beer$marca, "^ELK$")]<-"ELK MOUNTAIN"
beer$marca[str_detect(beer$marca, "^HACKER$")]<-"HACKER PSCHORR"
beer$marca[str_detect(beer$marca, "^HARD$")]<-"HARD CORE"
beer$marca[str_detect(beer$marca, "^LITTLE$")]<-"LITTLE KINGS"
beer$marca[str_detect(beer$marca, "^MCEWANS$")]<-"MCEWANS SCOTCH"
beer$marca[str_detect(beer$marca, "^OS$")]<-"OS CLASSIC"
beer$marca[str_detect(beer$marca, "^PILSNER$")]<-"PILSNER URQUELL"
beer$marca[str_detect(beer$marca, "^WHEAT$")]<-"WHEAT HOOK"
beer$marca[str_detect(beer$marca, "^ICE$")]<-"ICE DRAFT LIGHT"
beer$marca[str_detect(beer$marca, "^O$")]<-"ROYAL GUARD"
beer$marca[str_detect(beer$marca, "^JOHN$")]<-"JOHN COURAGE"
beer$marca[str_detect(beer$marca, "^JW$")]<-"JW DUNDEE'S"

# 5. Marcas de vinos y sidra
beer$num <- c(1:787)
#se crea esta variable para identificar las observaciones de las marcas que se van a eliminar

beer <- beer[-c(1, 22, 23, 74, 122, 123, 124, 125, 250, 319, 383, 384, 385, 386,
               476, 477, 557, 558, 604, 605, 606, 626, 627, 628, 629, 630, 631,
               678, 679, 703, 704, 705, 706, 750), ]

# Identificando cuántas marcas de cervezas quedan
brands <- sort(table(beer$marca), decreasing = TRUE)
brands

```

b. Reporte estadísticas descriptivas del valor de las ventas, el número de unidades vendidas y el precio promedio por marca en una tabla. Comente lo que encuentra.

c. Haga una búsqueda sobre las marcas incluidas en la lista y construya las siguientes variables dummy: 1) “importada”: variable igual a 1 si la cerveza es originaria de un país distinto a los Estados Unidos y cero en otro caso; y 2) “artesanal”: variable igual a 1 si la cerveza es producida artesanalmente y cero en otro caso. Reporte estadísticas descriptivas, por característica, en una tabla.

En primer lugar, con respecto a la variable *importada*, es fácil distinguir si una cerveza es nacional o importada cuando proviene de una empresa pequeña o mediana con una localización definida. Sin embargo, la industria de la cerveza ha crecido ampliamente durante las últimas décadas, y para grandes multinaciones tal vez

no es claro distinguir esta característica. El aumento de la demanda a nivel mundial ha desencadenado en una serie de decisiones logísticas por parte de las multinacionales encaminadas a reducir costos a través de la producción local, razón por la cual una misma marca puede tener fabricas en muchos países. Por ejemplo, la marca de cerveza mexicana Corona decidió en el 2019 expandir su producción por fuera de su país estableciéndose en otros lugares como China; Corona está aumentando la disponibilidad en otros mercados y al mismo tiempo está reduciendo sus costos.

Ahora bien, al analizar un país con una demanda tan significativa como lo es Estados Unidos, resulta natural pensar que muchas marcas extranjeras prefieran elaborar su producción destinada al público norteamericano en Estados Unidos. Para estos casos se decidió que se tomará una marca como importada siempre y cuando su casa matriz se encuentre fuera de Estados Unidos, incluso si la cerveza para este país es fabricada ahí mismo. Por ejemplo, la marca de cerveza St. Pauli Girl es de origen alemán y su casa matriz se encuentra en la ciudad de Bremen, Alemania. Aun cuando esta cerveza en Estados Unidos es fabricada en la ciudad de San Luis, en Misuri, se tomó como importada. Además, resulta lógico definir esta variable así puesto que la estrategia de marketing de muchas de estas empresas con origen extranjero que tienen sedes en EE. UU. consiste en resaltar su origen.

En segundo lugar, el criterio para clasificar una cerveza como artesanal o de producción masiva está mucho menos claro. La Asociación americana de cerveceros define a las cervecerías artesanales como aquellas que cumplen con 3 características: son pequeñas, independientes y tradicionales. El límite para el tamaño de una cervecería está comunmente definido por el número anual de barriles producidos o la participación en el mercado nacional (6 millones de barriles o 3% de la producción nacional). Por su parte, la independencia hace referencia al porcentaje máximo de la empresa de la que puede ser dueño un actor que no pertenezca a la industria en cuestión (máximo 25%), y la tradición se refiere a los ingredientes y recetas originales (Brewers Association, s.f.).

Desde luego, la definición de cerveza artesanal no es única y existen más aproximaciones a este concepto que suponen más o menos restricciones sobre el producto. Además, no cabe duda que esta clasificación se hace más imprecisa con los grandes cambios y la expansión que ha sufrido la industria cervecera. Para ilustrar esto, marcas como Guinness y Löwenbräu conservan su receta original prácticamente sin variaciones desde su creación pero fueron incorporadas a grupos multinacionales más grandes, por lo que ya no es claro si pierden o no su calidad de artesanal. Además, también se suma la inexactitud de buscar esta información, la cual muchas veces es escasa, 20 años después del momento en el que fueron recolectados los datos, pues muchas cosas pudieron suceder en este periodo que cambiaran la percepción de estas marcas.

En general, teniendo en cuenta que la construcción de esta variable está sesgada por problemas de definición y medición, se utilizó la información a nuestro alcance para identificar elementos en las cervezas entre los anteriormente mencionados y poder clasificar las 150 marcas que se encuentran en la base de datos.

```
#-----#
# Creando las variables *importada* y *artesanal* #
#-----#

# Variable *importada*
beer$importada <- ifelse( beer$marca == "BECK'S" | beer$marca == "MOLSON" |
  beer$marca == "LOWENBRAU" | beer$marca == "TECATE" |
  beer$marca == "FOSTER'S LAGER" | beer$marca == "LABATTS" |
  beer$marca == "MURPHY'S IRISH" | beer$marca == "ST PAULI GIRL" |
  beer$marca == "CARLING BLACK LABEL" | beer$marca == "CORONA" |
  beer$marca == "DOS EQUIS" | beer$marca == "GUINNESS STOUT" |
  beer$marca == "HEINEKEN" | beer$marca == "WARSTEINER" |
  beer$marca == "AMSTEL" | beer$marca == "MOOSEHEAD" |
  beer$marca == "BASS" | beer$marca == "CARTA BLANCA" |
  beer$marca == "HACKER PSCHORR" | beer$marca == "KIRIN" |
  beer$marca == "MODELO" | beer$marca == "MORETTI" |
  beer$marca == "SAMUEL SMITH" | beer$marca == "SAPPORO" |
```

```

beer$marca == "SPATEN" | beer$marca == "ASAHI" |
beer$marca == "CLAUSTHALER" | beer$marca == "ELEPHANT" |
beer$marca == "GROLSCH" | beer$marca == "HARP" |
beer$marca == "MCEWANS SCOTCH" | beer$marca == "PILSNER URQUELL" |
beer$marca == "STEINLAGERS" | beer$marca == "TENNENTS" |
beer$marca == "TOSELLI" | beer$marca == "BOHEMIA" |
beer$marca == "BUCKLER" | beer$marca == "CARLSBERG" |
beer$marca == "CASTLEMAINE" | beer$marca == "COOPERS" |
beer$marca == "COUNTRY" | beer$marca == "DOM" |
beer$marca == "DORTMUNDER" | beer$marca == "DUVEL" |
beer$marca == "FRANZISKANER" | beer$marca == "FULLERS" |
beer$marca == "HAACK-BECK" | beer$marca == "JOHN COURAGE" |
beer$marca == "LEZAJSK" | beer$marca == "MACKESON" |
beer$marca == "MOUSSY" | beer$marca == "NEWCASTLE" |
beer$marca == "O.B." | beer$marca == "OKOCIM" |
beer$marca == "PACIFICO" | beer$marca == "PERONI" |
beer$marca == "RED STRIPE" | beer$marca == "SUNTORY" |
beer$marca == "TSINGTAO" | beer$marca == "WHITBREAD" |
beer$marca == "YOUNGS" | beer$marca == "ZYWIEC" |
beer$marca == "ROYAL GUARD" , 1, 0)

# Variable *artesanal*
beer$artesanal <- ifelse(beer$marca == "SAMUEL ADAMS" | beer$marca == "STROH'S" |
beer$marca == "AUGSBURGER" | beer$marca == "PETE'S WICKED" |
beer$marca == "LEINENKUGEL" | beer$marca == "GOOSE ISLAND" |
beer$marca == "LOWENBRAU" | beer$marca == "OREGON" |
beer$marca == "BLUE MOON" | beer$marca == "KILLIAN'S" |
beer$marca == "ROGUE" | beer$marca == "BERGHOFF" |
beer$marca == "GUINNESS STOUT" | beer$marca == "SHIPYARD" |
beer$marca == "NEW AMSTERDAM" | beer$marca == "CELIS" |
beer$marca == "PYRAMID" | beer$marca == "ROLLING ROCK" |
beer$marca == "ANCHOR" | beer$marca == "BADERBRAU" |
beer$marca == "BLACK DOG" | beer$marca == "BOULDER" |
beer$marca == "DEVIL" | beer$marca == "DUSSELDORFER" |
beer$marca == "HACKER PSCHORR" | beer$marca == "NAKED ASPEN" |
beer$marca == "POINT" | beer$marca == "SAMUEL SMITH" |
beer$marca == "SIERRA NEVADA" | beer$marca == "STATE STREET" |
beer$marca == "COLD SPRING" | beer$marca == "ELK MOUNTAIN" |
beer$marca == "LEGACY" | beer$marca == "RED HOOK" |
beer$marca == "SUMMIT" | beer$marca == "TOSELLI" |
beer$marca == "COOPERS" | beer$marca == "DOM" |
beer$marca == "JW DUNDEE'S" | beer$marca == "RAZOR'S" |
beer$marca == "RED RIVER" | beer$marca == "WATNEY'S" , 1, 0)

# Exportando la base de datos
beer2 <- subset(beer, select = c(1:8, 10, 11))
write.csv(beer2 , file = "beer2.csv", row.names = FALSE)

```

e. Ejecute la misma regresión del punto anterior utilizando el comando correspondiente de R. Reporte los estimados en una tabla en la que incluya los resultados del literal anterior, interprete los coeficientes y compare. ¿Hay alguna diferencia entre el procedimiento manual y el automático?

Procedimiento automático

```
reg_auto <- lm(q ~ price + importada + artesanal , data = beer2)
summary(reg_auto)
```

Table 1: Resultados regresión MCO

	<i>Variable dependiente:</i>	
	q	
	(1)	(2)
price	-701,311.40*** (141,464.00)	-701,311.40*** (141,464.00)
importada	72,514.27 (105,876.90)	72,514.27 (105,876.90)
artesanal	86,144.85 (91,731.34)	86,144.85 (91,731.34)
Constante	730,440.50*** (97,565.98)	730,440.50*** (97,565.98)
Observaciones	753	753
R ²	0.04	0.04
R ² Ajustado	0.04	0.04
Residual Std. Error (df = 749)	993,428.80	993,428.80
Estadístico F (df = 3; 749)	10.77***	10.77***

Nota: *p<0.1; **p<0.05; ***p<0.01

h. Con base en la especificación del literal (d), haga una partición de la matriz de variables regresoras $X = [X_1 X_2]$ donde la matriz X_1 esté compuesta por la variable *price* y la matriz $X_{\{2\}}$ esté compuesta por las demás variables regresoras. Muestre que, en este caso, el efecto parcial de la variable *price* sobre $E[q|X_1, X_2]$ se puede recuperar siguiendo el Teorema de Frisch-Waugh. Describa esquemáticamente los pasos que usted sigue al aplicar el Teorema y explique la intuición asociada a cada uno de dichos pasos.

Teniendo en cuenta la partición del enunciado, nuestro modelo se puede reescribir como:

$$q = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

donde X_1 es una matriz de dimensión $n \times 1$ que contine únicamente la variable *price* y X_2 es una matriz de dimensión $n \times 3$ que está compuesta por una columna de unos (de tal forma que el modelo tenga una constante) y las variables *importada* y *artesanal*.

En este caso, lo que se busca a través del Teorema de Frisch-Waugh es recuperar el efecto parcial de la variable *price* a través de 3 pasos que se enlistan a continuación:

- Paso 1: Regresar a q sobre X_2 . La regresión sería

$$q = X_2\gamma + \epsilon$$

y los residuales de esta regresión se guardan como e_2 . En este paso se está calculando qué parte de q no es explicada por X_2 . En otras palabras, se está limpiando a q del efecto de X_2 .

- Paso 2: Regresar a X_1 sobre X_2 . La regresión sería

$$X_1 = X_2\omega + \epsilon$$

y los residuales de esta regresión se guardan como e_{21} . Ahora, se está calculando qué parte de X_1 no es explicada por X_2 . Es decir, se está limpiando a X_1 del efecto de X_2 .

- Paso 3: Regresar los residuales de la primera regresión (e_2) contra los residuales de la segunda (e_{21}). La regresión será

$$e_2 = e_{21}\beta_1 + \epsilon$$

Finalmente, se obtiene que $\hat{\beta}_1$ es el mismo que si se hubiera hecho la regresión de q contra X_1 y X_2 , específicamente, $\hat{\beta}_1 = -701311.4$.

Lo razón de lo anterior es que, al buscar estimar solamente el coeficiente de la variable *price*, se debe primero remover el efecto de todas las demás variables regresoras, y de esta forma sí se podrá obtener el efecto parcial de esta variable sobre q . Precisamente, los residuales que se incluyen en la regresión del tercer paso son los que obtuvieron el efecto limpio de X_2 .

Para terminar de evidenciar esto, escribamos el estimador $\hat{\beta}_1$ como

$$\hat{\beta}_1 = [e_{21}'e_{21}]^{-1}e_{21}'e_2$$

Adicionalmente sabemos que $e_2 = M_{x2}q$ y $e_{21} = M_{x2}X_1$, donde M_{x2} es la matriz creadora de residuales generada por las columnas de X_2 . Remplazando esto en el estimador de β_1 :

$$\hat{\beta}_1 = [(M_{x2}X_1)'M_{x2}X_1]^{-1}(M_{x2}X_1)'M_{x2}q$$

Distribuyendo la transpuesta:

$$\hat{\beta}_1 = [X_1'M_{x2}M_{x2}X_1]^{-1}X_1'M_{x2}M_{x2}q$$

Como M_{x2} es simétrica ($X_1' = X_1$), entonces:

$$\hat{\beta}_1 = [X_1'M_{x2}M_{x2}X_1]^{-1}X_1'M_{x2}M_{x2}q$$

Como M_{x2} es idempotente ($M_{x2}M_{x2} = M_{x2}$), entonces:

$$\hat{\beta}_1 = [X_1'M_{x2}X_1]^{-1}[X_1'M_{x2}q]$$

donde el primer término está indicando la parte de X_1 que no explica X_2 y el segundo término es la parte de q que no explica X_2 . Es decir, se está calculando $\hat{\beta}_1$ con unas matrices a las que se les quitó el efecto de X_2 .

```
beer2 <- read_csv("beer2.csv")

#-----#
# Aplicación Teorema de Frisch-Waugh #
#-----#

# Paso 1
p1 <- lm(q ~ importada + artesanal, data = beer2)
```



```
beer2$e2 <- residuals(p1)
# Se incluye el intercepto en la regresión porque X2 tiene un vector de unos

# Paso 2
p2 <- lm(price ~ importada + artesanal, data = beer2)
beer2$e21 <- residuals(p2)
#Se incluye el intercepto en la regresión porque X2 tiene un vector de unos

# Paso 3
p3 <- lm(e2 ~ e21 -1, data = beer2)
summary(p3)
# No se incluye el intercepto porque X1 solo lo constituye la variable *price*
```

Estimación de la elasticidad-precio de la demanda de cerveza

a. Haga un gráfico de dispersión entre el logaritmo de las cantidades vendidas y el logaritmo del precio e incluya una línea de ajuste lineal, escriba el modelo de regresión correspondiente a esta línea de ajuste y comente. ¿Qué tan bueno es el ajuste de este modelo? ¿Cómo se puede mejorar el ajuste?

```
beer2$log_price <- log(beer2$price)
beer2$log_q <- log(beer2$q)

reg_1 <- lm(log_q ~ log_price, data = beer2)
summary(reg_1)

library(ggplot2)
ggplot(beer2, aes(x=log_price, y=log_q)) +
  geom_point(shape=20) + # Use hollow circles
  geom_smooth(method=lm, col = "red", se=FALSE) + labs(y = "Log(q)", x = "Log(price)") + theme_bw()
```

b. Estime la elasticidad precio de la demanda de cerveza mediante una regresión lineal que incluya las siguientes variables de control: una constante y las características de la cerveza del literal (c) del punto 1.1.

```
reg_2 <- lm(log_q ~ log_price + importada + artesanal, data = beer2)
summary(reg_2)
```

c. Estime de nuevo la elasticidad precio de la demanda de cerveza incluyendo, esta vez, constantes específicas por marca. Reporte los resultados de esta regresión junto con los del literal anterior, compárelos y comente: ¿cuál modelo genera la mejor estimación de la elasticidad y por qué?

```
reg_3 <- lm(log_q ~ log_price + importada + artesanal + marca, data = beer2)
summary(reg_3)
```

Table 2:

	<i>Variable dependiente:</i>	
	Log(q)	
	(1)	(2)
Log(price)	-2.506*** (0.287)	-1.925*** (0.470)
Importada	0.089 (0.306)	1.499 (3.128)
Artesanal	-0.416 (0.265)	0.284 (3.960)
Constante	8.222*** (0.222)	7.677*** (2.798)
Controles de marca	No	Sí
Observaciones	753	753
R ²	0.137	0.401
R ² Ajustado	0.134	0.252
Residual Std. Error	3.010 (gl = 749)	2.797 (gl = 602)
Estadístico F	39.677*** (gl = 3; 749)	2.687*** (gl = 150; 602)

Nota:

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

2. Simulaciones de Monte Carlo

a. Simule 150 observaciones de una variable pseudoaleatoria x_1 que se distribuye uniforme en el intervalo $[0, 10]$ y 150 observaciones de una variable pseudoaleatoria v que se distribuye normal con media 2 y varianza 9.

```
remove(list = ls())
set.seed(123)

n <- 150 # n: número de observaciones

x1 <- runif(n, 0, 10)
v <- rnorm(n, 2, 3)
```

b. Cree la variable $x_2 = x_1 + v$.

```
x2 <- x1 + v
```

c. Simule 150 observaciones de una variable aleatoria ε que viene de una distribución normal con media 0 y varianza igual a 25. Cree la variable $y = 10 - x_1 + 2x_2 + \varepsilon$.

```
epsilon <- rnorm(n, 0, 5)

y = 10 - x1 + (2*x2) + epsilon
```

d. Lleve a cabo una regresión de MCO de y sobre x_1 y x_2 , incluyendo el intercepto. ¿Qué tan bueno es el ajuste?

Table 3:

<i>Dependent variable:</i>	
	y
x1	-0.667*** (0.204)
x2	1.670*** (0.143)
Constant	11.035*** (0.867)
Observations	150
R ²	0.564
Adjusted R ²	0.558
Residual Std. Error	4.991 (df = 147)
F Statistic	95.029*** (df = 2; 147)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

```
reg <- lm(y ~ x1 + x2)
summary(reg)
```

e. Repita los pasos de los literales (c) y (d) de este experimento 1000 veces manteniendo x_1 y x_2 constantes y calcule las medias muestrales de los coeficientes estimados. ¿Cómo es la media de cada coeficiente con respecto a su respectivo parámetro poblacional?

Table 4: Estadísticas descriptivas experimento 1000 repeticiones

Statistic	Media	Desv. Est.	Mín.	Máx.
$\hat{\beta}_0$	10.021	0.887	7.140	12.863
$\hat{\beta}_1$	-1.001	0.199	-1.620	-0.324
$\hat{\beta}_2$	1.999	0.140	1.525	2.426

```
r <- 1000 # r: número de repeticiones

# Creamos vectores para guardar las estimaciones en cada repetición
betax0 <- rep(0, r)
betax1 <- rep(0, r)
betax2 <- rep(0, r)

for (i in 1:r){

  epsilon_i <- rnorm(n, 0, 5)
  y_i = 10 - x1 + (2*x2) + epsilon_i

  datos_i <- data.frame(Y = y_i , X1 = x1, X2 = x2, E = epsilon_i)

  reg_i <- lm( Y ~ X1 + X2, data = datos_i)
```

```

betax0[i] <- reg_i$coefficients[1]
betax1[i] <- reg_i$coefficients[2]
betax2[i] <- reg_i$coefficients[3]
}

estimaciones <- data.frame(BETA0 = betax0, BETA1 = betax1, BETA2 = betax2)

mean(estimaciones$BETA0)
mean(estimaciones$BETA1)
mean(estimaciones$BETA2)

```

f. Repita los pasos de los literales (a), (b) y (e) para $n = 450, 1500, 3000$. Grafique la distribución de los coeficientes para cada tamaño de muestra incluyendo $n = 150$. ¿Qué sucede con la distribución de los coeficientes a medida que el tamaño de muestra crece? ¿Qué nos sugiere este ejercicio con respecto a la normalidad asintótica del estimador de MCO?

```

# Creando lista para guardar los resultados para cada tamaño de muestra
result <- vector("list",3)
result

for (j in c(450, 1500, 3000)){

  # Hallando las realizaciones de *x1* y *v* y calculando *x2* para cada tamaño de muestra
  x1_j <- runif(j, 0, 10)
  v_j <- rnorm(j, 2, 3)
  x2_j <- x1_j + v_j

  # Definiendo el número de repeticiones y creando vectores para guardar la
  r <- 1000

  # Creando vectores para guardar la información de cada repetición
  betax0_j <- rep(0, r)
  betax1_j <- rep(0, r)
  betax2_j <- rep(0, r)

  for (i in 1:r){

    # En cada repetición se halla una realización de *epsilon* y se calcula *y*
    epsilon_i <- rnorm(j, 0, 5)
    y_i = 10 - x1_j + (2*x2_j) + epsilon_i

    # Creando un data frame que contenga las realizaciones de cada repetición
    datos_i <- data.frame(Y = y_i, X1 = x1_j, X2 = x2_j, E = epsilon_i)

    # Realizando la regresión
    reg_i <- lm( Y ~ X1 + X2, data = datos_i)

    # Guardando las estimaciones de los coeficientes de la regresión de cada repetición
    betax0_j[i] <- reg_i$coefficients[1]
    betax1_j[i] <- reg_i$coefficients[2]
    betax2_j[i] <- reg_i$coefficients[3]
  }
}

```

```

# Creando data frame que contenga las estimaciones para cada tamaño de muestra
estimaciones_j <- data.frame(BETA0 = betax0_j, BETA1 = betax1_j, BETA2 = betax2_j)

# Guardando los resultados para cada tamaño de muestra
result[[j]] <- estimaciones_j

}

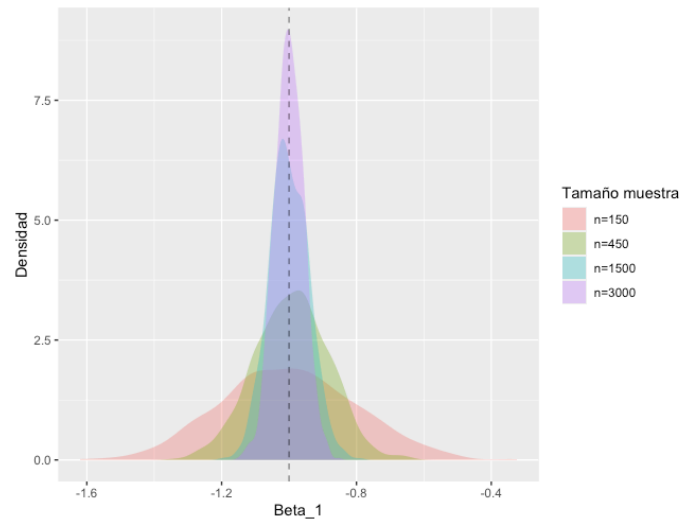
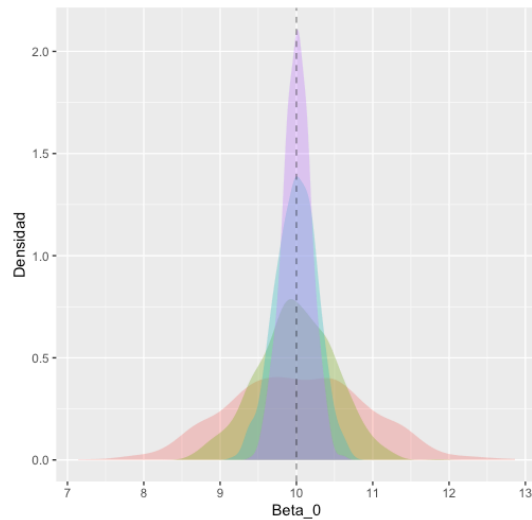
# Compilando en una sola base las estimaciones de los coeficientes de todas las
# repeticiones y tamaños de muestra

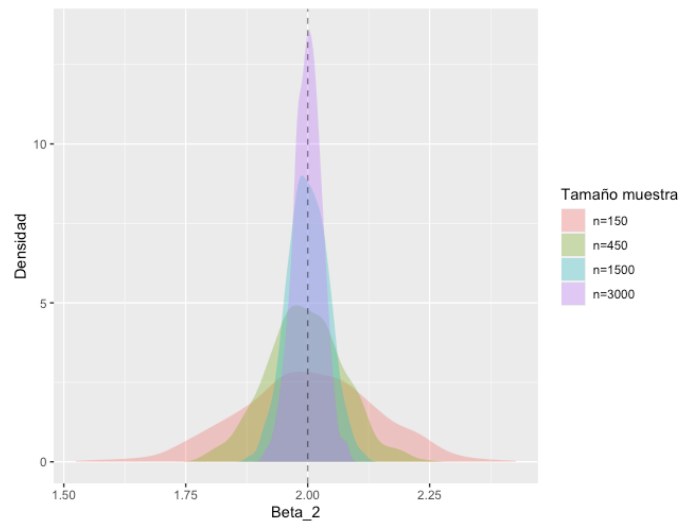
Com <- data.frame(estimaciones, result[[450]], result[[1500]], result[[3000]])

```

Table 5:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
BETA0 n=150	1,000	10.021	0.887	7.140	9.396	10.630	12.863
BETA1 n=150	1,000	-1.001	0.199	-1.620	-1.137	-0.870	-0.324
BETA2 n=150	1,000	1.999	0.140	1.525	1.907	2.091	2.426
BETA0.1 n=450	1,000	9.998	0.506	8.607	9.666	10.352	11.861
BETA1.1 n=450	1,000	-0.992	0.110	-1.345	-1.066	-0.917	-0.641
BETA2.1 n=450	1,000	1.995	0.079	1.782	1.943	2.046	2.248
BETA0.2 n=1500	1,000	9.994	0.271	9.160	9.812	10.186	10.756
BETA1.2 n=1500	1,000	-1.000	0.059	-1.198	-1.039	-0.959	-0.787
BETA2.2 n=1500	1,000	2.000	0.041	1.877	1.972	2.030	2.118
BETA0.3 n=3000	1,000	10.003	0.190	9.395	9.878	10.133	10.644
BETA1.3 n=3000	1,000	-1.001	0.043	-1.146	-1.029	-0.971	-0.852
BETA2.3 n=3000	1,000	2.000	0.029	1.909	1.980	2.020	2.082





Referencias

Craft Brewer Definition. (s.f.). Brewers Association. Recuperado el día 9 marzo 2021 de <https://www.brewersassociation.org/statistics-and-data/craft-brewer-definition/>