

# Econometría Avanzada

## Primer examen parcial

Profesor: Jorge Florez  
Profesor asistente: Mateo Cardona  
Facultad de Economía–Universidad del Rosario

**Fecha de entrega:** viernes 12 de marzo

Este cuestionario debe ser llevado a cabo usando el *software* econométrico **R**. Por favor sea muy breve y concreto en sus respuestas y análisis.

### 1 Estimación de la demanda de cerveza

Se requiere estimar la demanda de cerveza con base en datos de ventas de cerveza provenientes de *Dominick's Finer Foods* (DFF), la segunda cadena de supermercados más grande en el área metropolitana de Chicago (Estados Unidos) en los años 90. Estos datos fueron recolectados por el *Kilts Center for Marketing* de la Escuela de Negocios *Booth* de la Universidad de Chicago y contienen información sobre el valor total de las ventas, el volumen total de ventas, entre otras variables, de un conjunto de marcas de cerveza a nivel de código de barras (UPC). Una unidad de observación en esta base de datos es una UPC de una marca de cerveza vendida por la cadena DFF.

Las variables de la base de datos son las siguientes:

- *upc*: número correspondiente al código de barras del producto.
- *descrip*: descripción del producto (incluye marca).
- *price*: precio promedio por una unidad (botella o lata) de 355 ml de cerveza.
- *q*: número de unidades de 355 ml de cerveza vendidas por UPC.
- *volsales*: ventas totales en volumen (medido mililitros).
- *sales*: ventas totales en dólares.

En la base de datos encontrará que una misma marca de cerveza (por ejemplo, Heineken) aparece con diferentes códigos UPC. Esto se debe a que los códigos UPC varían con la presentación del producto. Así, un *six pack* de cerveza Heineken tiene un código UPC distinto del de un *twelve pack* de Heineken.

#### 1.1 Análisis de regresión:

Lleve a cabo las siguientes operaciones:

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

- 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.
- d. Lleve a cabo una regresión del número de cervezas vendidas por semana ( $q$ ) contra una constante, el precio por unidad y las características encontradas en el literal anterior. Programe esta regresión manualmente, es decir, construya la expresión para el estimador, así como su error estándar y el coeficiente R-cuadrado. Así mismo, calcule los valores ajustados de la variable dependiente y los residuales y guarde estas dos nuevas variables en su base de datos.
- 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?
- f. Ejecute la misma regresión del punto anterior pero esta vez incluya constantes específicas por marca. Reporte los estimados en una tabla omitiendo los estimados de las constantes de marca y en la que incluya los resultados obtenidos en el literal (d). Interprete los coeficientes de la nueva regresión y compárelos con los de la primera regresión. ¿Cuál es el mejor modelo? ¿Qué pasa con los coeficientes de las características en la regresión con constantes por marca?
- g. Estime de nuevo el modelo del literal (d), pero excluya la constante. Prediga los valores ajustados de la cantidad y los residuales de este modelo; úse estas variables junto con las variables análogas que calculó en el literal (d) y calcule:

- $\frac{1}{n} \sum_{i=1}^n e_i$ ,
- $\frac{1}{n} \sum_{i=1}^n price_i * e_i$ ,
- $\bar{\hat{q}}_i, \bar{q}_i$ .

Compare los cálculos para ambas especificaciones. ¿Son estos cálculos, para ambas especificaciones, consistentes con las características algebraicas de MCO predichas por la teoría?

- h. Con base en la especificación del literal (d), haga una partición de la matriz de variables regresoras  $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2]$  donde la matriz  $\mathbf{X}_1$  esté compuesta por la variable *price* y la matriz  $\mathbf{X}_2$  esté compuesta por las demás variables regresoras. Muestre que, en este caso, el efecto parcial de la variable *price* sobre  $E[q|\mathbf{X}_1, \mathbf{X}_2]$  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.

## 1.2 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?
- 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.
- 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?

- d. **Falso o verdadero:** “no se puede rechazar la hipótesis de que la demanda de cerveza es perfectamente inelástica”. (Para justificar su respuesta, plantee la hipótesis que se quiere probar y la hipótesis alternativa, escriba el test correspondiente y su distribución, escriba el valor crítico y concluya.)
- e. ¿Por qué la demanda de cerveza es elástica?

## 2 Simulaciones Monte Carlo:

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