

## Problem Set 2

### Instrucciones Generales

En esta segunda tarea, se espera que los/as estudiantes puedan desarrollar los conceptos revisados en la segunda parte del curso. Para esto, analizaremos una base de datos histórica que contiene indicadores de desempeño de plantas de manufactura, en términos de su productividad. Cada alumno/a deberá subir a la plataforma de U-Cursos dos archivos:

- Un informe con respuestas, resultados, gráficos y tablas que sostengan su análisis, tomando en cuenta las preguntas de este enunciado. Este informe **debe** estar en formato PDF, independientemente de si fue desarrollado en Word o en L<sup>A</sup>T<sub>E</sub>X(u otra herramienta). Sea **lo más sintético posible**.
- Un script de R que ejecute el análisis empírico asociado al informe. Este debe correr íntegramente, sin errores.

El trabajo **es personal**. No se tolerará el plagio, siguiendo los estándares detallados por la Universidad de Chile en sus códigos de conducta. El 7 estará en 180 puntos. Fecha de entrega: Viernes 24 de Mayo de 2024 a las 23:59.

### Sobre los datos

Usted trabajará con datos de la Encuesta Nacional Industrial Anual (ENIA) que contiene información a nivel de *plantas*, contando con observaciones anuales entre 1979 y 1998. En la base de datos que usted analizará, cada fila corresponde a una planta manufacturera indexada por la variable **ID**, observada en un año específico (**year**). Cada planta, además, tiene asociado un código clasificador de sector industrial (**ciiu3**) de tres dígitos (a nivel de grupo)<sup>1</sup>.

## 1 Estadística Descriptiva

1. Grafique la cantidad de plantas activas por año. En otras palabras, realice un gráfico de barras que contenga la cantidad de observaciones por año. Describa qué observa. [5 puntos]
2. Genere una variable dicotómica que tome valor 1 si la planta está activa **todos los años**, y 0 en caso contrario. Denote dicha variable como **siempre activa**, y reconozca que dicha variable es *constante a través de las observaciones de cada planta*. ¿Para cuántas plantas contamos con toda la información disponible? [5 puntos]

---

<sup>1</sup>Puede encontrar [aquí](#) información relativa a la construcción de esta categorización estandarizada internacionalmente.

- De manera similar, reconozca que hay un fenómeno de *entrada y salida* de plantas. Para esto, genere una variable dicotómica (**salida**) que tome valor 1 si la planta *sale* de la base de datos en el año  $t$ , y 0 si la planta sigue activa. Para construir esta variable, considere el siguiente ejemplo. Suponga que usted observa datos para la planta  $i$  en los años 1979, 1980, 1981 y 1982. Luego desde 1983 en adelante no observa datos (deja de estar operativa). Para el período 1979 - 1981, la variable **salida** debe tomar valor 0. En el año 1982 la planta “cierra” y la planta sale de la base de datos en 1983, por lo que la variable **salida** debe tomar valor 1 en 1982. ¿Cuántas plantas cierran cada año? [10 puntos]
- Considere la siguiente métrica:

$$\text{Tasa de Salida}_t = y(t) = \frac{1}{N_t} \sum_{i=1}^{N_t} \text{Salida}_{it}$$

Donde  $i$  representa cada planta. Genere un gráfico en el plano  $(t, y(t))$  que muestre su evolución con una línea. [5 puntos]

- Considere las métricas de productividad dadas por **tfp\_ls**, **tfp\_fe**, **tfp\_lp**. Para cada una de estas, genere un gráfico que incorpore los histogramas de la productividad de las plantas para 1979 y 1997, a modo de comparación entre años.<sup>2</sup> [5 puntos]
- Usted reconoce que comparar la productividad de plantas *entre* industrias pareciera no ser la mejor idea, dado que hay industrias que son estructuralmente más productivas que otras. Para solucionar este problema, usted generará, utilizando cada medida de productividad, un indicador de *productividad relativa* ( $y_{ict}^*$ ), dado por:

$$y_{ict}^* = \frac{y_{ict} - \bar{y}_{ct}}{s_{ct}}$$

En donde  $i$  indexa a la planta,  $c$  a la industria a la cual pertenece y  $t$  el año en que es observada. En función de lo mismo,  $y_{ict}$  corresponde a la métrica de productividad *base*. Note que el segundo término en el numerador corresponde a la productividad media de la industria  $c$  en el año  $t$  y el término del denominador corresponde a la desviación estándar muestral de la industria  $c$  en  $t$ . Agréguele el sufijo **\_re** para nombrar a cada variable (e.g.: **tfp\_ls\_re**). [15 puntos]

- ¿Cuál es el valor esperado de cada una de las tres medidas de productividad relativa? Desarrolle matemáticamente, y confirme sus predicciones graficando las funciones de distribución de probabilidad de toda la muestra. [15 puntos]

## 2 En la búsqueda de la productividad

En esta sección, usted analizará los determinantes de la productividad de las empresas chilenas durante el período de datos disponibles. Para esto, considere siempre las métricas de productividad relativas que construyó en la sección anterior.

---

<sup>2</sup>Las medidas no son comparables *entre* sí, por lo que deberá generar exactamente tres gráficos, uno por medida, con dos histogramas superpuestos en cada uno.

1. Considere un modelo de regresión univariada en el que su variable dependiente sea productividad relativa, y su variable independiente el stock de capital de la empresa, dado por la variable **K<sub>lr</sub>**. Estime a través de Mínimos Cuadrados Ordinarios, para cada una de las tres métricas, el siguiente modelo:

$$\text{Productividad}_i = \beta_0 + \beta_1 \text{Stock de Capital}_i + \varepsilon_i \quad (1)$$

Reporte una tabla con las tres regresiones (una por cada métrica). Describa los resultados, poniendo especial detención en la dirección, la magnitud y la significancia de los coeficientes obtenidos. Muestre empíricamente que la estimación punto de  $\beta_1$  que obtiene a través del comando **lm** es equivalente al ratio entre la covarianza entre las variables dependiente e independiente, y la varianza de la variable independiente. [15 puntos]

2. Usted busca construir un modelo estadístico más ajustado a la realidad, por lo tanto decide estimar una regresión que incluya también el número de trabajadores de la planta (dado por **employment**), en la forma:

$$\text{Productividad}_i = \beta_0 + \beta_1 \text{Stock de Capital}_i + \beta_2 \# \text{Trabajadores}_i + \varepsilon_i \quad (2)$$

Muestre los resultados en una tabla similar a la de la pregunta anterior. ¿Cómo cambia la estimación punto de  $\beta_1$  en los distintos modelos? ¿Cómo afecta el stock de capital y el número de trabajadores a la productividad de las plantas? [15 puntos]

3. Asuma que el segundo modelo, incluyendo el número de trabajadores es el *verdadero*. Encuentre una expresión para el sesgo que tiene la primera estimación respecto de  $\beta_1$ . [10 puntos]
4. Considere el mismo modelo anterior, pero esta vez incluya en la regresión la cantidad de trabajadores calificados, dado por la variable **employees**. Reporte los resultados de correr las mismas tres regresiones para cada métrica de productividad, esta vez considerando dos muestras: (i.) todas las plantas, y (ii.) solo las que se mantienen siempre abiertas.<sup>3</sup> ¿Se observan dinámicas distintas para cada muestra? Plantee una explicación económica al respecto. [10 puntos]
5. Agregue como regresor al modelo anterior la variable de cantidad de trabajadores no calificados (**workers**). ¿Qué problema encuentra? ¿Cuál es el problema matemático que le impide correr el modelo? [10 puntos]
6. Considere la siguiente especificación:

$$\begin{aligned} \text{Productividad}_i = & \beta_0 + \beta_1 \text{Stock de Capital}_i + \beta_2 \# \text{Trabajadores} \\ & + \sum_{t=1980}^{1998} \gamma_t \mathbb{1}(\text{Año}_i = t) + \varepsilon_i \end{aligned} \quad (3)$$

Corra la regresión para toda la muestra, considerando los términos dados por la sumatoria como variables dicotómicas que toman valor 1 si la observación corresponde al año  $t$

---

<sup>3</sup>Note que obtendrá resultados para 6 regresiones.

(i.e.: 1980, 1981, ..., hasta 1998). Muestre los resultados de su regresión y analice el  $R^2$ , comparándolo con el de los resultados obtenidos en la especificación (2). ¿Cómo cambia? Describa mecánicamente, utilizando la fórmula vista en clases, qué es lo que ocurre que explica la dirección del cambio. [10 puntos]

7. Deténgase en la variable **tech.assis**, que indica la cantidad de recursos recibidos por la planta en cada año en términos monetarios para asistencia técnica. Genere una variable dicotómica que tome valor 1 si la planta recibe algún tipo de asistencia técnica (i.e.: que los recursos económicos sean más que 0). Nómbrala **tratamiento\_tecnico** Aproximadamente, ¿cuántas plantas al año reciben asistencia técnica? [5 puntos]
8. Corra tres regresiones del tipo:

$$\begin{aligned} \text{Productividad}_i = & \beta_0 + \beta_1 \text{Stock de Capital}_i + \beta_2 \# \text{Trabajadores} \\ & + \beta_3 \text{Tratamiento Asistencia Técnica} + \varepsilon_i \end{aligned} \quad (4)$$

En el que la variable de tratamiento asistencia técnica sea la dicotómica que generó anteriormente. Muestre los resultados de sus estimaciones, y discuta por qué el estimador  $\hat{\beta}_3$  no debe ser entendido como el impacto causal de recibir asistencia técnica sobre la productividad de la planta. [25 puntos]

9. Estime un modelo de probabilidad lineal para explicar la variable dicotómica de **salida**, utilizando la métrica de productividad dada por **tfp\_fe\_re**. Para esto, considere la siguiente especificación:

$$\begin{aligned} \text{Salida}_{i,t} = & \beta_0 + \beta_1 \text{Productividad}_{i,t} + \beta_2 \text{Productividad}_{i,t-1} + \beta_3 \text{Productividad}_{i,t-2} + \\ & \beta_4 \text{Dicotómica Asistencia Técnica}_{i,t} + \varepsilon_{i,t} \end{aligned}$$

En donde se consideran las métricas de productividad para el mismo año ( $t$ ), la del año pasado ( $t - 1$ ) y la del antepasado ( $t - 2$ ). Describa los resultados, deteniéndose particularmente en los efectos de las productividades de años anteriores sobre la probabilidad de cierre, además de distinguir si recibir asistencia técnica es útil para la sobrevivencia de las plantas. [10 puntos]

10. Testee las siguientes hipótesis:

- $\beta_1 < \beta_2$
- $\beta_2 = \beta_3$

Reporte sus resultados, incluyendo los p-values, e indique cuáles son las conclusiones económicas que entrega realizar dicho ejercicio. [10 puntos]