



DEPARTAMENTO DE INGENIERÍA CIVIL INDUSTRIAL
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE
IN3242-1 ESTADÍSTICA

PROBLEM SET 2

TAREA N°2

Integrantes: Adolfo Rojas
Profesor: Matías R. Labbé
Auxiliar: Diego Guevara
Fabiana Torres
Ayudantes: Antonia González
Javiera Donoso T.

Fecha de entrega: 31 de Mayo
Santiago de Chile

Índice de Contenidos

1. Estadística Descriptiva	1
2. En la búsqueda de la productividad	5

Índice de Figuras

1. Manufactureras activas por año	1
2. Gráfico de la tasa de salida anual	2
3. Histogramas de productividad entre industrias	3
4. Funciones de densidad probabilística de productividad relativa por industria	4

Índice de Tablas

1. Regresiones de Productividad Relativa	5
2. Regresiones de Productividad Relativa con Stock de Capital y Número de Trabajadores	6
3. Regresiones para todas las plantas	8
4. Regresiones para las plantas siempre activas	9
5. Regresiones de Productividad, Stock de Capital, Número de Trabajadores y Asistencia Técnica	11
6. Regresión lineal probabilística de salida para la industria tfp fe	11
7. Test de hipótesis para $\beta_1 < \beta_2$ y $\beta_1 = \beta_3$	12

Índice de Códigos

1. Snippet con modificación del data frame para agregar productividad relativa por industria	3
2. Snippet con calculo de coeficientes de regresión lineal	5

1. Estadística Descriptiva

Ítem 1.1

Del gráfico de barras de la figura 1 vemos como en un inicio de la encuesta se alcanza el máximo de las fabricas activas para luego dar paso a una baja de las mismas hasta 1983, luego para el periodo 1987 - 1996 incrementa el número de fabricas activas hasta abruptamente tener que cerrar una gran cantidad de estas para los dos años siguientes que es hasta donde nos proporciona la encuesta

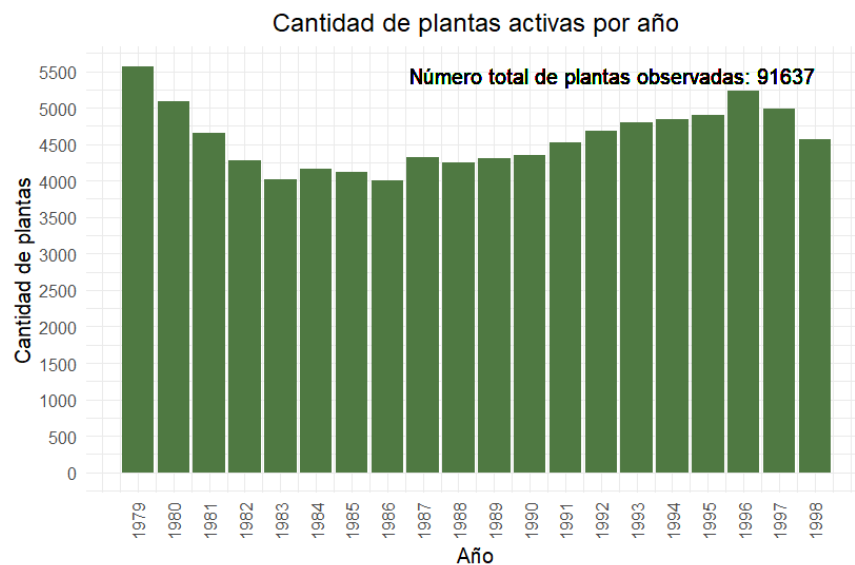


Figura 1: Manufactureras activas por año

Ítem 1.2

De las 91637 observaciones/plantas que nos presenta la base de datos, mediante la suma de la variable dicotómica sabemos que solo 1340 de estas se encuentran operativas hasta la fecha de la encuesta (1998)

Ítem 1.3

Ignorando el último año de la encuesta presente la siguiente tabla muestra la cantidad de fábricas que se dieron de baja con su respectivo año de salida

Año	Fábricas Salientes
1979	571
1980	492
1981	441
1982	424
1983	324
1984	197
1985	278
1986	231
1987	239
1988	228
1989	219
1990	169
1991	210
1992	259
1993	277
1994	363
1995	426
1996	639
1997	741

Ítem 1.4

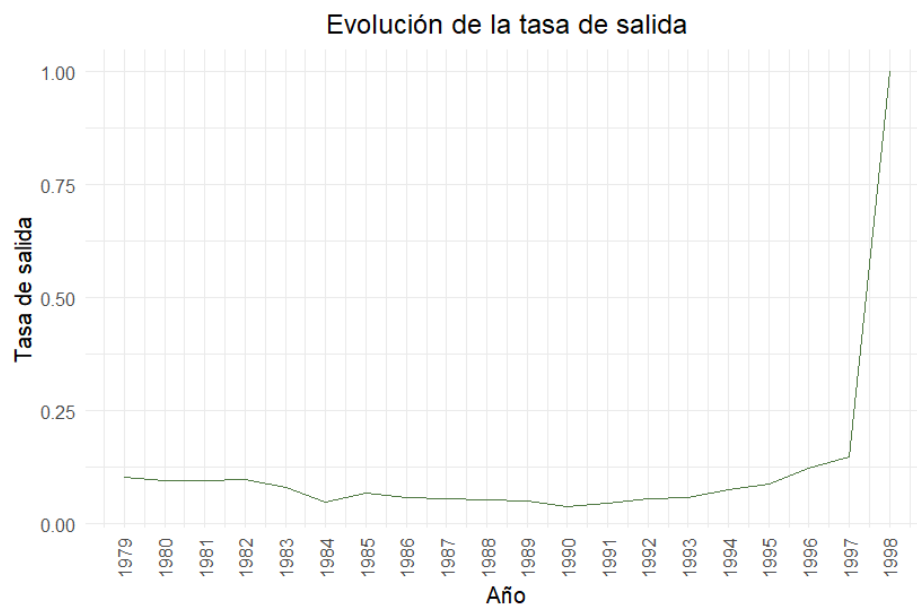


Figura 2: Gráfico de la tasa de salida anual

Dicho gráfico fue armado con el uso de la variable dicotómica del ítem anterior y la métrica $Tasa\ de\ Salida(t) = \frac{1}{N_t} \sum_{i=1}^{N_t} Salida_{i,t}$

Ítem 1.5

A continuación los gráficos de frecuencia que comparan la productividad por sector entre 1979 - 1997

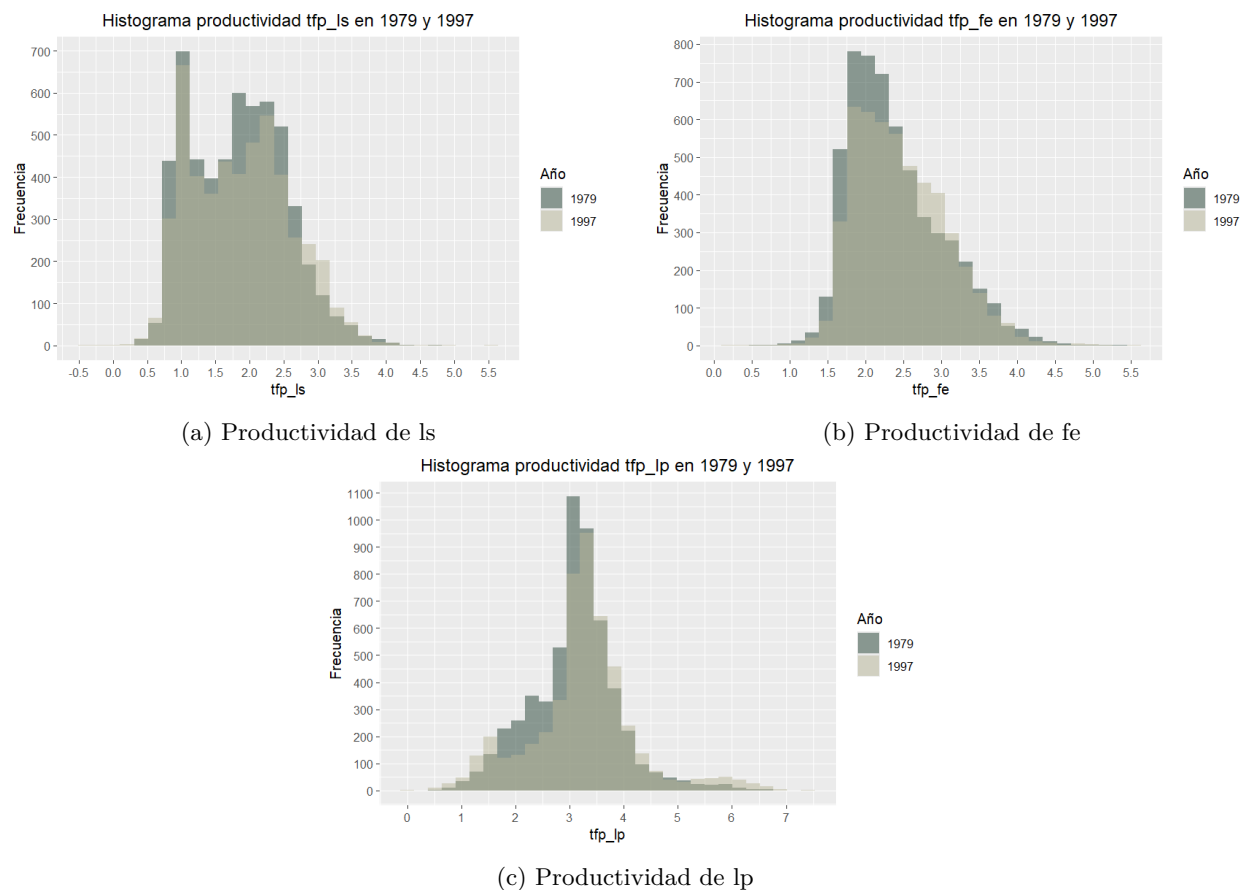


Figura 3: Histogramas de productividad entre industrias

Ítem 1.6

Código 1: Snippet con modificación del data frame para agregar productividad relativa por industria

```
1 fab <- fab %>%
2   group_by(ciiu3, year) %>%
3   mutate(
4     mean_tfp_ls = mean(tfp_ls, na.rm = TRUE),
5     sd_tfp_ls = sd(tfp_ls, na.rm = TRUE),
6     tfp_ls_re = (tfp_ls - mean_tfp_ls) / sd_tfp_ls,
7   )
```

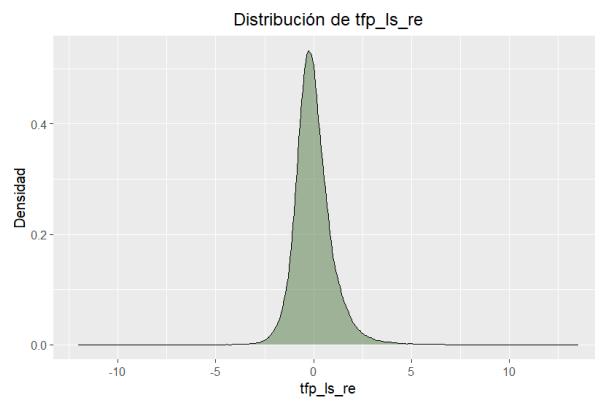
```

8 mean_tfp_fe = mean(tfp_fe, na.rm = TRUE),
9 sd_tfp_fe = sd(tfp_fe, na.rm = TRUE),
10 tfp_fe_re = (tfp_fe - mean_tfp_fe) / sd_tfp_fe,
11
12 mean_tfp_lp = mean(tfp_lp, na.rm = TRUE),
13 sd_tfp_lp = sd(tfp_lp, na.rm = TRUE),
14 tfp_lp_re = (tfp_lp - mean_tfp_lp) / sd_tfp_lp
15 ) %>%
16 ungroup() %>%
17 # Eliminar las columnas temporales de medias y desviaciones estándar
18 select(-mean_tfp_ls, -sd_tfp_ls, -mean_tfp_fe, -sd_tfp_fe, -mean_tfp_lp, -sd_tfp_lp)
19
20 View(fab)

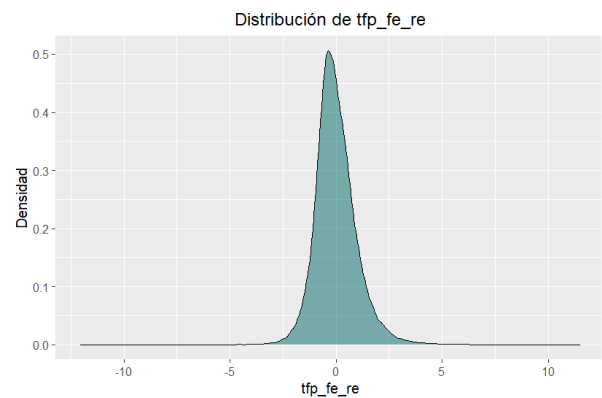
```

Ítem 1.7

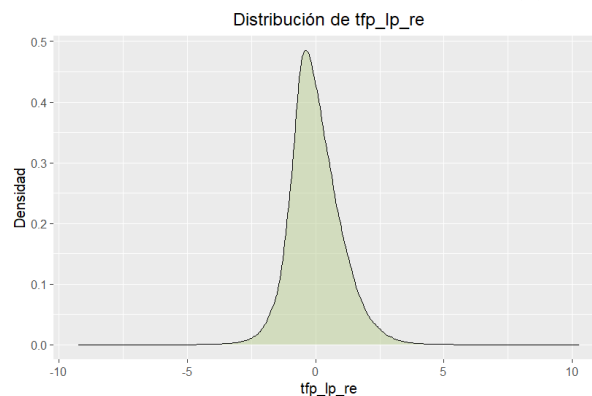
Tomamos el valor esperado de la variable aleatoria $E[y_{ict}^*] = \frac{E[y_{ict}] - \bar{y}_{ct}}{s_{ct}}$ Luego como $E[y_{ict}] = \bar{y}_{ct}$ se tiene que $\frac{\bar{y}_{ct} - \bar{y}_{ct}}{s_{ct}} = 0$ que es justamente lo que se puede visualizar en cada una de las PDF



(a) PDF de ls relativa



(b) PDF de fe relativa



(c) PDF de lp relativa

Figura 4: Funciones de densidad probabilística de productividad relativa por industria

2. En la búsqueda de la productividad

Ítem 2.1

tfp_ls_re: El intercepto es -0.0004618976 , no significativo, lo que indica que el valor de la productividad relativa tfp_ls_re cuando el stock de capital es cero es aproximadamente -0.00046 . El coeficiente de K_r , 0.0000000027 , no es significativo, lo que sugiere que un aumento en el stock de capital está asociado con un ligero aumento en tfp_ls_re . Sin embargo, no es estadísticamente significativo. R^2 : 0.0000073198 es muy bajo, lo que indica que el stock de capital es insignificante en la variabilidad en tfp_ls_re .

tfp_fe_re: El intercepto es significativo al 1 %, lo que indica que el valor de tfp_fe_re cuando el stock de capital es cero es aproximadamente -0.0156 . El coeficiente de K_r es significativo al 1 %, lo que indica que un aumento en el stock de capital está asociado con un aumento en tfp_fe_re . R^2 es bajo, con lo que el stock de capital explica aproximadamente el 0.83 % de la variabilidad en tfp_fe_re .

tfp_lp_re: Al igual que tfp_fe_re el intercepto es significativo al 1 %. Indica que el valor de tfp_lp_re cuando el stock de capital es cero es aproximadamente -0.0186 . El coeficiente de K_r también es significativo al 1 % y positivo con lo que un aumento en el stock de capital está asociado con un aumento en tfp_lp_re . R^2 es bajo, lo que muestra que el stock de capital explica aproximadamente el 1.19 % de la variabilidad en tfp_lp_re .

	Dependent variable		
	tfp_ls_re	tfp_fe_re	tfp_lp_re
Constant	-0.0004618976 (0.0033409250)	-0.0155998600^{***} (0.0033269610)	-0.0186081900^{***} (0.0033210330)
K_r	0.0000000027 (0.0000000033)	0.0000000911^{***} (0.0000000033)	0.0000001087^{***} (0.0000000033)
Observations	91,637	91,637	91,637
R2	0.0000073198	0.0083493130	0.0118800400
Adjusted R2	-0.0000035930	0.0083384920	0.0118692500
Residual Std. Error (df = 91635)	0.9968375000	0.9926710000	0.9909022000
F Statistic (df = 1; 91635)	0.6707574000	771.5311000000^{***}	$1,101.7150000000^{***}$

Note: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Tabla 1: Regresiones de Productividad Relativa

Ahora para verificar empíricamente que nuestra estimación de β_1 y la dada por el comando es la misma hacemos uso del siguiente Snippet

Código 2: Snippet con calculo de coeficientes de regresión lineal

```
1 reg_tfp_ls <- lm(tfp_ls_re ~ K_r, data = datos)
2 reg_tfp_fe <- lm(tfp_fe_re ~ K_r, data = datos)
3 reg_tfp_lp <- lm(tfp_lp_re ~ K_r, data = datos)
```

```

5 # Tabla con los resultados de las regresiones
6 stargazer(reg_tfp_ls, reg_tfp_fe, reg_tfp_lp, type = "text",
7           title = "Regresiones de Productividad Relativa vs. Stock de Capital",
8           header = FALSE, intercept.bottom = FALSE,
9           digits = 10)
10
11 # Calculamos covarianza entre la productividad relativa con K_r y la varianza de K_r
12 cov_tfp_ls_kr <- cov(datos$tfp_ls_re, datos$K_r, use = "complete.obs")
13 var_kr <- var(datos$K_r, use = "complete.obs")
14 beta1_ls <- cov_tfp_ls_kr / var_kr
15
16 cov_tfp_fe_kr <- cov(datos$tfp_fe_re, datos$K_r, use = "complete.obs")
17 beta1_fe <- cov_tfp_fe_kr / var_kr
18
19 cov_tfp_lp_kr <- cov(datos$tfp_lp_re, datos$K_r, use = "complete.obs")
20 beta1_lp <- cov_tfp_lp_kr / var_kr
21
22 # Printeamos la resta de nuestra estimación y los resultados de lm()
23 cat("diferencia de beta1 para tfp_ls_re:", beta1_ls - coef(reg_tfp_ls)["K_r"], "\n")
24 cat("diferencia de beta1 para tfp_fe_re:", beta1_fe - coef(reg_tfp_fe)["K_r"], "\n")
25 cat("diferencia de beta1 para tfp_lp_re:", beta1_lp - coef(reg_tfp_lp)["K_r"], "\n")

```

el cual nos da una diferencia insignificante entre estimaciones de β_1 : 0.0000000000000000000034328, 0.000000000000000000001323489 y $-0.000000000000000000006352747$ respectivamente, con lo que podemos concluir que en efecto son iguales

Ítem 2.2

	<i>Dependent variable:</i>		
	tfp_ls_re	tfp_fe_re	tfp_lp_re
	(1)	(2)	(3)
Constant	0.0181720700*** (0.0036555690)	-0.0413921500*** (0.0036373930)	-0.1110547000*** (0.0035588940)
K_r	0.0000000220*** (0.0000000036)	0.0000000644*** (0.0000000036)	0.0000000130*** (0.0000000035)
employment	-0.0003201047*** (0.0000255972)	0.0004430743*** (0.0000254699)	0.0015880970*** (0.0000249202)
Observations	91,637	91,637	91,637
R2	0.0017110460	0.0116134500	0.0538143500
Adjusted R2	0.0016892570	0.0115918800	0.0537937000
Residual Std. Error (df = 91634)	0.9959934000	0.9910413000	0.9696534000
F Statistic (df = 2; 91634)	78.5293600000***	538.3456000000***	2,605.8440000000***

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

Tabla 2: Regresiones de Productividad Relativa con Stock de Capital y Número de Trabajadores

tfp_ls: La estimación de β_1 aumentó ligeramente al incluir el número de trabajadores y se volvió significativa. Esto indica que el stock de capital tiene una relación positiva con la productividad relativa, pero su efecto es pequeño y se atenúa al incluir el número de trabajadores.

tfp_fe: La estimación de β_1 disminuyó pero se mantuvo significativa. Aunque el stock de capital sigue teniendo un impacto positivo, este es menor en el modelo multivariado en comparación con el modelo univariado.

tfp_lp: La estimación de β_1 disminuyó considerablemente pero se mantuvo significativa, sugiriendo que el efecto positivo del stock de capital en la productividad relativa es pequeño y disminuye al incluir el número de trabajadores.

Impacto del stock de capital: El impacto del stock de capital K_r es positivo y significativo en todos los modelos, contribuyendo positivamente a la productividad relativa. Sin embargo, la magnitud del impacto es pequeña en todos los casos.

Número de trabajadores: Para tfp_ls, el número de trabajadores tiene un impacto negativo y significativo sobre la productividad relativa, indicando que más trabajadores no necesariamente aumentan la productividad por trabajador. En contraste, para tfp_fe y tfp_lp, el número de trabajadores tiene un impacto positivo y significativo.

Ítem 2.3

Utilizando la fórmula del sesgo por variable omitida para sacar los sesgos de las estimaciones de β_1 en el modelo univariado con respecto al modelo multivariado se obtiene:

$$\text{Sesgo}_i = \beta_2 \cdot \frac{\text{Cov}(\text{Stock de Capital}_i, \text{Número de Trabajadores}_i)}{\text{Var}(\text{Stock de Capital}_i)}$$

Luego haciendo el calculo en R obtenemos dichos sesgos para cada industria

	Sesgo
tfp_ls_re	-0.0000000192911
tfp_fe_re	0.00000002670185
tfp_lp_re	0.00000009570657

Ítem 2.4

	<i>Dependent variable:</i>		
	tfp_ls_re	tfp_fe_re	tfp_lp_re
Constant	0.0210768400*** (0.0036618360)	-0.0401507800*** (0.0036457430)	-0.1087405000*** (0.0035658330)
K_r	0.0000000121*** (0.0000000037)	0.0000000602*** (0.0000000037)	0.0000000051 (0.0000000036)
employment	-0.0006688614*** (0.0000398377)	0.0002940310*** (0.0000396626)	0.0013102560*** (0.0000387932)
employees	0.0012260780*** (0.0001073687)	0.0005239720*** (0.0001068968)	0.0009767684*** (0.0001045538)
Observations	91,637	91,637	91,637
R^2	0.0031296710	0.0118725400	0.0547147000
Adjusted R^2	0.0030970350	0.0118401900	0.0546837600
Residual Std. Error (df = 91633)	0.9952909000	0.9909168000	0.9691972000
F Statistic (df = 3; 91633)	95.8938500000***	366.9960000000***	1,767.9570000000***

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

Tabla 3: Regresiones para todas las plantas

Una diferencia que podemos notar es que R^2 es más grande para la muestra de plantas siempre activas (excepto por la 2da industria que son relativamente iguales), o sea que el modelo se ajusta mejor y lo hace más predecible a este set de datos "homogéneo" lo cual hace sentido pues presentan menos variabilidad, una sorpresa es el hecho de que el estimador del regresor de la variable **employees** es más significativa en el primer modelo, o sea que en las empresas que se mantienen en pie no hacen un gran impacto el hecho de tener obreros más calificados, también notamos la diferencia entre interceptos, para el 2do modelo se tiene una mayor productividad base en comparación, finalmente pasando a K_r vemos que en el 2do modelo tiene más relevancia (sobretudo para la 3era industria), con lo que el capital tiene un impacto más significativo y consistente en la productividad para las plantas siempre abiertas.

	<i>Dependent variable:</i>		
	tfp_ls_re	tfp_fe_re	tfp_lp_re
Constant	0.1228350000*** (0.0060676460)	0.1380382000*** (0.0061708600)	0.1164579000*** (0.0061946490)
K_r	0.0000000162** (0.0000000067)	0.0000000872*** (0.0000000068)	0.0000000508*** (0.0000000068)
employment	-0.0006287453*** (0.0000484332)	-0.0000144943 (0.0000492571)	0.0010511430*** (0.0000494469)
employees	0.0007123417*** (0.0001329076)	-0.0000791355 (0.0001351685)	0.0001022168 (0.0001356895)
Observations	26,800	26,800	26,800
R2	0.0083537150	0.0096030500	0.0769471800
Adjusted R2	0.0082426930	0.0094921690	0.0768438400
Residual Std. Error (df = 26796)	0.8813688000	0.8963614000	0.8998168000
F Statistic (df = 3; 26796)	75.2439500000***	86.6061300000***	744.5860000000***

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

Tabla 4: Regresiones para las plantas siempre activas

Ítem 2.5

Obtenemos el mismo set de coeficientes expuestos en la tabla 3 a diferencia de que se agrega una nueva fila la cual es vacía, esto debido a que esta nueva variable es una combinación lineal de las otras variables endógenas del modelo *employment* (fuerza laboral total) y *employees* (trabajadores calificados) con lo que habría multicolinealidad en el mismo

Ítem 2.6

Ítem 2.7

Año	Plantas con asistencia técnica
1979	235
1980	199
1981	172
1982	181
1983	157
1984	194
1985	199
1986	189
1987	188
1988	283
1989	203
1990	210
1991	244
1992	241
1993	265
1994	247
1995	266
1996	293
1997	255
1998	262

Ítem 2.8

Como se vió en clases, es difícil demostrar una relación directa de implicancia (causalidad), pues si bien de la tabla podemos ver como hay una correlación fuerte con el modelo de productividad, no es sencillo mostrar que este efecto es único de la variable puesto que puede darse el caso (y lo más probable es que haya sido así) que alguna variable omitida le de más peso, produciendo así un sesgo asociado a la misma

	<i>Dependent variable:</i>		
	tfp_ls_re	tfp_fe_re	tfp_lp_re
Constant	0.0118316500*** (0.0036736470)	-0.0497987700*** (0.0036516620)	-0.1200724000*** (0.0035711710)
K_r	0.0000000174*** (0.0000000036)	0.0000000583*** (0.0000000036)	0.0000000064* (0.0000000035)
employment	-0.0003896308*** (0.0000259543)	0.0003508912*** (0.0000257990)	0.0014892130*** (0.0000252303)
tratamiento_tecnico	0.2432041000*** (0.0156893300)	0.3224588000*** (0.0155954400)	0.3458984000*** (0.0152516800)
Observations	91,637	91,637	91,637
R2	0.0043220050	0.0162033900	0.0590958300
Adjusted R^2	0.0042894070	0.0161711800	0.0590650200
Residual Std. Error (df = 91633)	0.9946955000	0.9887429000	0.9669487000
F Statistic (df = 3; 91633)	132.5858000000***	503.0733000000***	1,918.4130000000***

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

Tabla 5: Regresiones de Productividad, Stock de Capital, Número de Trabajadores y Asistencia Técnica

Ítem 2.9

	coefficients	Std. Error	t value	Pr(> t)
Constant	0.120983	0.001255	96.420	< 0.0000000000000002***
Productividad	-0.017251	0.001710	-10.086	< 0.0000000000000002***
Productividad _{t-1}	-0.004052	0.001882	-2.154	0.0313*
Productividad _{t-2}	0.001422	0.001719	0.828	0.4079
tratamiento técnico	-0.015534	0.005520	-2.814	0.0049**

Tabla 6: Regresión lineal probabilística de salida para la industria tfp fe

Recordamos que correlación no implica causalidad, luego de la tabla es directo que la probabilidad de cierre base es alta, como podríamos haber supuesto, la probabilidad de salida será menor si la productividad de la planta en el año actual es alta (la productividad del año pasado produce un efecto similar pero no tan significativo), una sorpresa es ver como el regresor asociado a la productividad del año ante pasado es positiva, o sea que si hace un par de años la planta era muy productiva su probabilidad de cierre aumentaba (lo cual es razonable pues si la planta se mantiene productiva constantemente entonces es poco probable que cierre pero si luego con el paso de los años su productividad va bajando es porque algo no va del todo bien y se pasa a considerar su cierre), ahora viendo el último regresor que sería el tratamiento técnico podemos ver que si dicha planta cuenta con soporte su probabilidad de cierre será menor

Ítem 2.10

Como podemos ver en la siguiente tabla, asumiendo un nivel de significancia = 0.5 tenemos que ambas hipótesis son rechazadas, del primer test podemos concluir que el peso que tiene para determinar la salida de una planta la producción actual es mayor que la del año pasado mientras que de la otra hipótesis solo podemos concluir que las productividades del año pasado y antepasado no tienen el mismo peso a la hora de determinar la probabilidad de cierre de una fábrica

	Diferencia	Valor t	p-value
$\beta_1 < \beta_2$	-0.01319868	-5.190816	0.0000001049765
$\beta_1 = \beta_3$	-0.005474408	-2.148237	0.0316983

Tabla 7: Test de hipótesis para $\beta_1 < \beta_2$ y $\beta_1 = \beta_3$