

# trabajo\_practico\_carbone.pdf

## Pregunta 1: Cobb-Duglas

Usted está interesado en estimar la función de producción de una empresa condicional al factor trabajo (L) y al factor capital (K). Para eso, se sabe que la función Cobb-Douglas es:

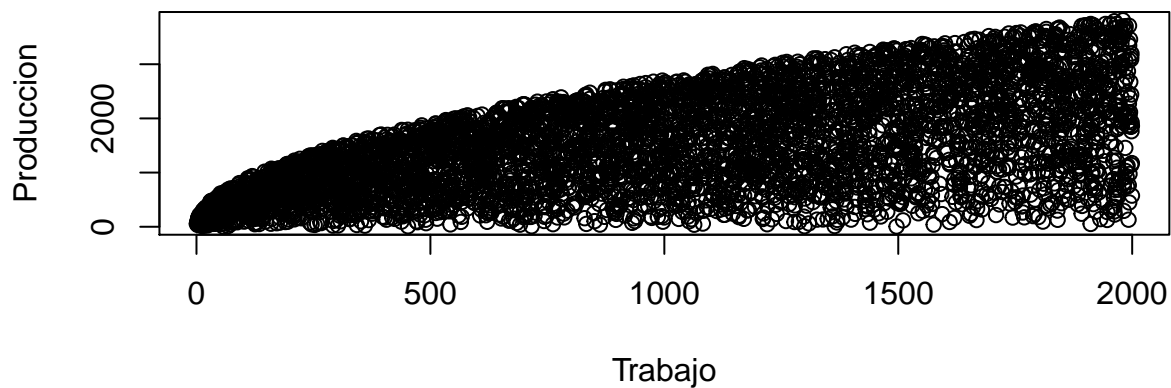
$$Y = AK^{\alpha}L^{\beta} :$$

Donde Y es la producción, A es la tecnología, L es el factor trabajo y K es el factor capital.

- Utilice la base de datos llamada "Data.xlsx", la hoja Pregunta 1.
- Grafique la relación entre trabajo y producción. Comente.
- Grafique la relación entre capital y producción. Comente.
- Estime los coeficientes  $\alpha$  y  $\beta$ . Interprete cada uno de los coeficientes.
- Realice un test para corroborar si  $\alpha$  es estadísticamente distinto a 0.5.
- Realice un test para corroborar si  $(\alpha + \beta)$  es estadísticamente distinto a 1.

punto a y b

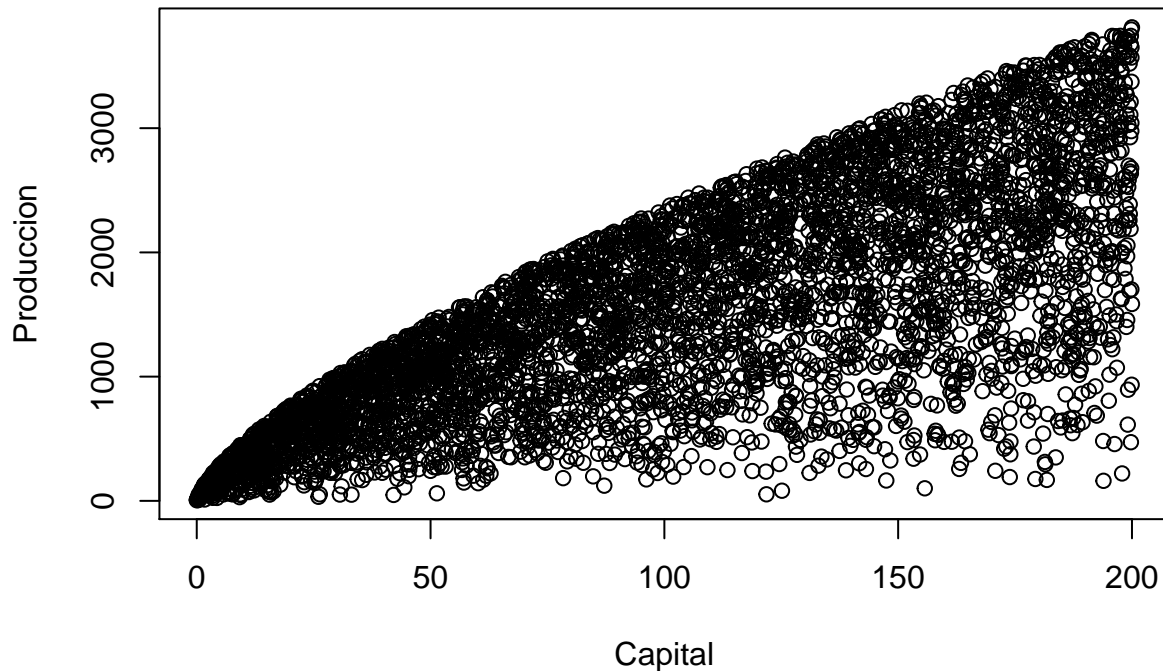
```
library(readxl)
pregunta_1=read_excel("Data.xlsx",sheet="Pregunta 1")
plot(x=pregunta_1$L,y=pregunta_1$Y,xlab = "Trabajo",ylab = "Produccion")
```



En la figura anterior se puede ver que a medida que se incrementa el trabajo, se incrementa la producción, pero no linealmente sino más bien con una forma de función de raíz. Esto se debe a la ley de rendimientos decrecientes.

### Punto c

```
plot(x=pregunta_1$k,y=pregunta_1$Y,xlab = "Capital",ylab = "Produccion")
```



En la figura anterior se puede ver que a medida que se incrementa el capital, se incrementa la producción, pero no linealmente sino mas bien con una forma de funcion de raíz, aunque mas empinada que con el trabajo, es decir, mas parecida a una lineal.

### Punto d

Como lo único que sabemos hacer son regresiones lineales lo primero que hay que hacer es aplicarle el logaritmo para linealizar la función y “bajar” los  $\alpha$  y  $\beta$  así:

$$Y = AK^{\alpha}L^{\beta} \implies \ln(Y) = \ln(AK^{\alpha}L^{\beta})$$

Aplicando propiedades de logaritmos tenemos:

$$\ln(Y) = \ln(A) + \alpha * \ln(K) + \beta * \ln(L)$$

Esta ecuación se corresponde a una ecuación de la forma log-log para un modelo de regresion lineal. Ya sabemos que esta regresión representa que un aumento de 1% en K implica una variación en Y de  $\alpha\%$ . Esto es básicamente la elasticidad de la producción respecto del capital (K)

Ahora corremos la regresion lineal:

```

pregunta_1$ln_y<-log(pregunta_1$Y)
pregunta_1$ln_l<-log(pregunta_1$L)
pregunta_1$ln_k<-log(pregunta_1$K)

regresor_log_log<-lm(pregunta_1$ln_y~ pregunta_1$ln_l+pregunta_1$ln_k,data = pregunta_1)
summary(regresor_log_log)

```

```

##
## Call:
## lm(formula = pregunta_1$ln_y ~ pregunta_1$ln_l + pregunta_1$ln_k,
##     data = pregunta_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.073864 -0.000501 -0.000011  0.000495  0.054299
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.419e-01  3.608e-04   2056  <2e-16 ***
## pregunta_1$ln_l 5.000e-01  4.562e-05  10960  <2e-16 ***
## pregunta_1$ln_k 7.000e-01  4.517e-05  15499  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.003165 on 4997 degrees of freedom
## Multiple R-squared:      1, Adjusted R-squared:      1
## F-statistic: 1.807e+08 on 2 and 4997 DF, p-value: < 2.2e-16

```

Recordamos la ecuación:  $\ln(Y) = \ln(A) + \alpha * \ln(K) + \beta * \ln(L)$ .

Como resultado observamos que  $\alpha$  acompaña a K y  $\beta$  a L. Entonces vemos que  $\beta = 1/2$  y  $\alpha = 0.7$ . Ambos son significativos al 10%, 5% y 1%.

El intercepto A que es la tecnología también es significativo y su valor es .741. En este caso no parece tener mucho sentido, puesto que si capital ni trabajo no veo como puede haber producción.

El R ajustado es 1 es decir que marca una correlacion perfecta. El estadístico F es significativo, eso significa que todas las variables conjuntamente son significativas.

En términos del problema la elasticidad produccion-trabajo es 0.5 y la elasticidad produccion-capital es 0.741. Eso significa que cuando el trabajo aumento en 1% la produccion lo hace en 0.5% y cuando el capital aumenta 1% la produccion lo hace en 0.741%. Esto se condice con el grafico donde veíamos que la pendiente era mas empinada con el capital que con el trabajo.

## Punto e

Tenemos que ver si  $\hat{\alpha}$  que es el regresor del capital (K) es estadísticamente diferente de 0.5.

Eso significa plantear como hipotesis nula que  $H_0 = \hat{\alpha} - 0.5 = 0$  hipotesis alternativa  $H_a = \hat{\alpha} - 0.5 \neq 0$ .

Rechazar la hipotesis nula implica que  $\hat{\alpha}$  es estadísticamente distinto a 0.5.

entonces para  $\hat{\alpha}=0.7$  calculamos el estadístico t:

$$\frac{\hat{\alpha} - 0.5}{\text{se}(\hat{\alpha})}$$

```
k<- coef(summary(regresor_log_log))[3,1]
stErrorK<-coef(summary(regresor_log_log))[3,2]

tvalue<-((k-0.5)/stErrorK)
tvalue
```

```
## [1] 4428.413
```

Este valor de t es mucho mas alta que cualquier que pueda obtener en tabla de distribucion t. Por lo tanto RHN por lo tanto  $\alpha=0.7$  es estadisticamente distinto de a 0.5

Pero también podemos calcular el p-value. El p value nos indica el nivel de significancia por debajo del cual no se rechaza la hipotesis nula:

$$p\text{-value} = 1 - \phi(t), \text{ siendo } t = \frac{\hat{\beta} - \beta}{ee(\hat{\beta})}$$

Asi que volviendo a hacer lo calculos obtenemos el p-value:

```
## Aca puedo usar lower.tail=FALSE, pero para respetar la ecuacion no lo uso.
## tambien se podría usar pt, es decir una distribucion t-student, pero para seguir la teoría
# que indica que para una muestra grande ya se toman los valores de una normal se elige esta misma.
p_value<-1-pnorm(tvalue)
sprintf("p-value: %f, t-value: %f",p_value,tvalue)
```

```
## [1] "p-value: 0.000000, t-value: 4428.413402"
```

Por cuestiones de calculo de coma flotante el p-value es cero, pero de todas formas el significado es el mismo, es muy pequeño por lo tanto es significativo al 10%, al 5% y al 1% y se rechaza la hipotesis nula.

## Punto f

Aca hay dos formas, una es con una prueba F y la otra que es mas sencilla pero menos general. Optamos por la segunda dado que tiene un procedimiento identico al primero.

entonces  $H_0 : \alpha + \beta = 1$  y  $H_1 : \alpha + \beta \neq 1$ .

Estimamos el nuevo t.

$$t = \frac{(\alpha + \beta) - 1}{ee(\alpha + \beta)}$$

Esto quedaria:

$$t = \frac{(\alpha + \beta) - 1}{\sqrt{var(\alpha) + var(\beta) + 2 * cov(\alpha, \beta)}}$$

```
l<- coef(summary(regresor_log_log))[2,1]
stErrorL<-coef(summary(regresor_log_log))[2,2]

vcov(regresor_log_log)[3,2]
```

```
## [1] -6.542509e-12
```

```
# t_no_restringido<-((k+l)-1)/sqrt(((stErrorK^2+stErrorL^2)+2*(-6.542509^-12)))

t_no_restringido<-((k+l)-1)/sqrt(((stErrorK^2+stErrorL^2)+2*(vcov(regresor_log_log)[3,2])))

p_value_no_restringido<-1-pnorm(t_no_restringido)
sprintf("p-value: %f y el t_no_restringido: %f",p_value_no_restringido,t_no_restringido)

## [1] "p-value: 0.000000 y el t_no_restringido: 3120.564231"
```

Nuevamente puede verse como en el caso anterior, el p-value <0.01 por lo tanto RHN por lo tanto es estadísticamente distinto de 1.

## Pregunta 2: Regresión múltiple con variables dicotómicas.

Usted está interesado en estudiar la ecuación de Mincer (impacto del nivel educativo en los ingresos). Para eso, utilice la hoja Pregunta 2 de la base de datos llamada "Data.xlsx". Esta base de datos posee la siguiente información: - Sexo (variable categórica) que toma el valor 1 cuando el individuo es varón y 2 cuando es mujer. - Edad (variable continua). - Nivel educativo (variable categórica) que toma el valor 1 para primaria incompleta, 2 primaria completa, 3 secundaria incompleta, 4 secundaria completa 5 universitario incompleto, 6 universitario completo. - Ingreso (variable continua). Para eso usted intenta estimar los siguiente modelos:

$$\text{ingreso}_i = \beta_0 + \beta_1 * \text{edad} + u_i$$

$$\text{ingreso}_i = \beta_0 + \beta_1 * \text{edad} + \delta * \text{Mujer} + u_i$$

$$\text{ingreso}_i = \beta_0 + \beta_1 * \text{edad} + \delta * \text{Mujer} + \gamma * \text{NivEd} + u_i$$

```
library(stargazer)
```

```
##
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
data_mincer=read_excel("Data.xlsx",sheet="Pregunta 2")
regresor_mincer_1<-lm(data_mincer$ingreso~data_mincer$edad)
regresor_mincer_2<-lm(data_mincer$ingreso~data_mincer$edad+factor(data_mincer$sexo))
regresor_mincer_3<-lm(data_mincer$ingreso~data_mincer$edad+factor(data_mincer$sexo)+factor(data_mincer$NivEd))
```

```
stargazer(regresor_mincer_1,regresor_mincer_2,regresor_mincer_3, type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               ingreso
##                               (1)          (2)          (3)
## -----
## edad                        383.730***      381.740***      393.819***
##                               (36.705)      (36.360)      (34.711)
##
## sexo)2                      -6,568.779***    -10,943.300***
##                               (908.501)      (863.372)
##
## nivel_ed)2                  5,440.980**
##                               (2,581.933)
##
## nivel_ed)3                  9,146.914***
##                               (2,522.591)
##
## nivel_ed)4                  13,960.820***
##                               (2,433.356)
##
## nivel_ed)5                  16,995.350***
##                               (2,569.602)
##
## nivel_ed)6                  31,329.630***
##                               (2,474.183)
##
## Constant                    12,087.350***      15,056.600***      357.215
##                               (1,567.460)      (1,606.077)      (2,802.751)
## -----
## Observations                2,684            2,684            2,684
## R2                          0.039            0.058            0.199
## Adjusted R2                 0.039            0.057            0.197
## Residual Std. Error  23,583.000 (df = 2682)  23,360.740 (df = 2681)  21,549.630 (df = 2676)
## F Statistic           109.294*** (df = 1; 2682) 81.831*** (df = 2; 2681) 95.272*** (df = 7; 2676)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

## Interpretaciones:

### Modelo 1

$R^2$  ajustado=0.039, lo que implica que el modelo en general es muy poco representativo esto seguro se debe a que la regresión que mas representaría el modelo es una polinómica y no una lineal (ocurre con todos los modelos que siguen abajo). F al 10%, 5% y 1% para todos los modelos lo que significa que es globalmente significativo

Significancia de la constante y el parametro regresor. Ambos son significativos al 10%, 5% y 1%.

Cuando la edad es 0 los ingresos son 12,087.350 (esto claramente no tiene sentido y se debe a que los datos parten de una edad legal para trabajar)

Por cada año que aumenta la edad el ingreso se incrementa en 383.73.

### Modelo 2

Significancia global  $R^2$  ajustado=0.057. Lo que implica que el modelo en general es muy poco significativo.

Significancia de la constante y los parametro regresores. Todos son significativos al 10%, 5% y 1%.

Para el caso en que se es varón y la edad es cero el ingreso es de 15,056.60 (nuevamente esto no tiene sentido pero es una extrapolación. Podríamos interpretar como que el salario sin antigüedad o al ingresar a un trabajo con edad legal es el mencionado.)

Para el caso en que se es mujer y la edad es cero (mismos comentarios sobre la extrapolación que en los casos anteriores) el salario se decrementa en -6,568.779 o sea que el ingreso es  $15,056.60 - 6,568.779$ .

En todos los casos por cada año de antigüedad el ingreso se incrementa en 381.740.

### Modelo 3

Significancia global  $R^2$  ajustado=0.197. Lo que implica que el modelo en general es muy poco significativo.

Significancia de la constante y los parametro regresores. Todos son significativos al 10%, 5% y 1% con salvedad del nivel educativo que lo es al 10% y al 5% y el intercepto que no es significativo.

Para el caso en que se es varón, la edad es cero y tiene primaria incompleta el ingreso es de 357.21 (nuevamente esto no tiene sentido pero es una extrapolación.)

En el caso de tener primaria completa el ingreso se incrementa en 5,440.98, para la secundaria incompleta en 9,146.914, para la secundaria completa 13,960.820, universitario incompleto 16,995.350 y universitario completo 31,329.630.

Para el caso en que se es mujer el ingreso se decrementa en \$10,943.300.

En todos los casos por cada año de antigüedad el ingreso se incrementa en \$393.82.

### Pregunta 3: Ecuación de Mincer con logaritmos en el ingreso

Realice el punto anterior con la particularidad que los modelos ahora son:

$$\ln(\text{ingreso}_i) = \beta_0 + \beta_1 * \text{edad} + \delta * \text{Mujer} + \gamma * \text{NivEd} + u_i$$

Corremos la regresión log-lin

```
regresor_mincer_log<-lm(log(data_mincer$ingreso)~data_mincer$edad+factor(data_mincer$sexo)+factor(data_mincer$nivel_ed))
stargazer(regresor_mincer_log,type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               ingreso)
## -----
## edad                        0.015***
##                               (0.001)
##
## sexo)2                      -0.461***
##                               (0.032)
##
## nivel_ed)2                  0.341***
##                               (0.095)
##
## nivel_ed)3                  0.431***
##                               (0.092)
##
## nivel_ed)4                  0.698***
##                               (0.089)
##
## nivel_ed)5                  0.745***
##                               (0.094)
##
## nivel_ed)6                  1.317***
##                               (0.091)
##
## Constant                    8.763***
##                               (0.103)
##
## -----
## Observations                2,684
## R2                          0.223
## Adjusted R2                 0.221
## Residual Std. Error        0.789 (df = 2676)
## F Statistic                109.976*** (df = 7; 2676)
## =====
## Note:                       *p<0.1; **p<0.05; ***p<0.01
```

Aquí podemos ver que todos los regresores son significativos al 10%, 5% y 1%, pero el modelo general R<sup>2</sup> explica muy poco (el 22%). Sin embargo como F=109.976 es significativo el modelo en forma global, es decir, todos sus regresores juntos.



Acá el análisis es el mismo que el anterior, pero se modifica el valor de los regresores y el intercepto, cambiando la interpretación por una variación en el parámetro, edad por ejemplo, en un año genera un incremento del ingreso en  $(0.015 * 100) \%$  lo que es igual a 1,5%. Es decir, por cada año que se incrementa el aumento del ingreso se incrementa en  $(0.015 * 100)\%$  o sea 1,5%. Misma interpretación vale para el resto de los regresores.