

ICG 028 PREPARACIÓN Y EVALUACIÓN DE PROYECTOS

CLASE 04: PROYECCIÓN DE LA DEMANDA

Año: 2021
Profesor: Sebastián Egaña

1. Introducción

En el video disponible para esta semana, dentro del aula virtual, se abordó lo relacionado con modelos de proyección o estimación de demanda de base matemática, como lo son los modelos de series de tiempo. Esto quiere decir, variables que poseen una relación que se da a través del tiempo.

En el caso del video se expresó la relación entre ventas y población.

Abordemos esto un poco más en detalle.

2. Estimación de la demanda y regresión lineal

2.1. Regresión Lineal

En términos simples, una regresión lineal corresponde al intento matemático y estadístico de analizar como un conjunto de datos se adapta o no a un comportamiento lineal, o sea a una relación lineal; ya sea esta relación directa o no.

Recuerde que en cursos matemáticos se le hablaba de la existencia de funciones, por ejemplo oferta en donde el precio depende de la cantidad demandada:

$$\text{Precio} = \alpha + \beta * \text{Cantidad} \quad (1)$$

En dónde solo se le decía, que la relación era la siguiente:

$$\text{Precio} = 2 + 2 * \text{Cantidad} \quad (2)$$

La regresión lineal, corresponde al intento de estimar los valores del intercepto o constante y de la pendiente utilizando un método matemático, en base a una muestra para extrapolar el comportamiento de una población y con esto haciendo uso de principios estadísticos. Por ejemplo, pretendemos saber si la muestra utilizada nos permite confirmar el signo de la pendiente y el valor de la misma.

2.2. ¿Por qué utilizamos esto en proyectos?

Una de las dificultades más relevantes en proyectos, corresponde a la necesidad de poder estimar de manera fidedigna las futuras ventas de nuestro proyecto. Sobre lo mismo, existe una sobre confianza en los modelos matemáticos para dicho tema, pensando en que esto le da mayor sustento a la forma en que estimamos las ventas de nuestro proyecto.

2.3. Un ejemplo relacionado

Se tiene un proyecto para una empresa de exportación, en donde se pretende realizar una estimación del Tipo de cambio nominal (TCN de ahora en adelante).

- Pregunta ¿por qué podría ser relevante el TCN para un proyecto como este? ¿Existe otro tipo de proyectos en dónde también podría ser relevante?

Para esto, el especialista econométrico plantea que el TCN posee una relación con variables como el precio del cobre (Pcu) y el dolar index (DXY).

- ¿Cuál es la razón de pensar en que existe una relación entre el TCN y el PCU? ¿Para el DXY?

Esto corresponde a una parte importante de la modelación; plantear la relación teórica entre las variables.

Se plantea el siguiente modelo:

$$TCN = \alpha + \beta * Pcu \quad (3)$$

Por otra parte, también podría ser:

$$TCN = \alpha + \beta * DXY \quad (4)$$

¿Cuál debería ser la relación entre cada variable y el TCN. Esto corresponde a un paso previo de la estimación econométrica.

2.4. Rstudio

Veamos un ejemplo. No abordaremos en detalle los códigos ahora expuestos, pero se utilizan para ir introduciendo esto en el curso.

```
library(tidyverse)
library(readxl)

ejemplo <- read_excel("ejemplo.xlsx")

data <- ejemplo %>%
  mutate(year = substring(fecha, 1, 4),
         ln_tcn = log(TCN),
         ln_dxy = log(`DXY Index`),
         ln_pcu = log(Pcu))

data_subset <- data %>%
  filter(year >= 2014)

modelo_01 <- lm(TCN ~ `DXY Index`, data = data_subset)

summary(modelo_01)

##
## Call:
## lm(formula = TCN ~ `DXY Index`, data = data_subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -78.329 -24.316  -7.695  11.964 150.390
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -124.5077    92.0229  -1.353   0.18
```

```
## `DXY Index`      8.3562      0.9744      8.575 7.9e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.88 on 77 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.4885, Adjusted R-squared:  0.4819
## F-statistic: 73.54 on 1 and 77 DF, p-value: 7.904e-13
```

```
modelo_02 <- lm(TCN ~ Pcu, data = data_subset)
```

```
summary(modelo_02)
```

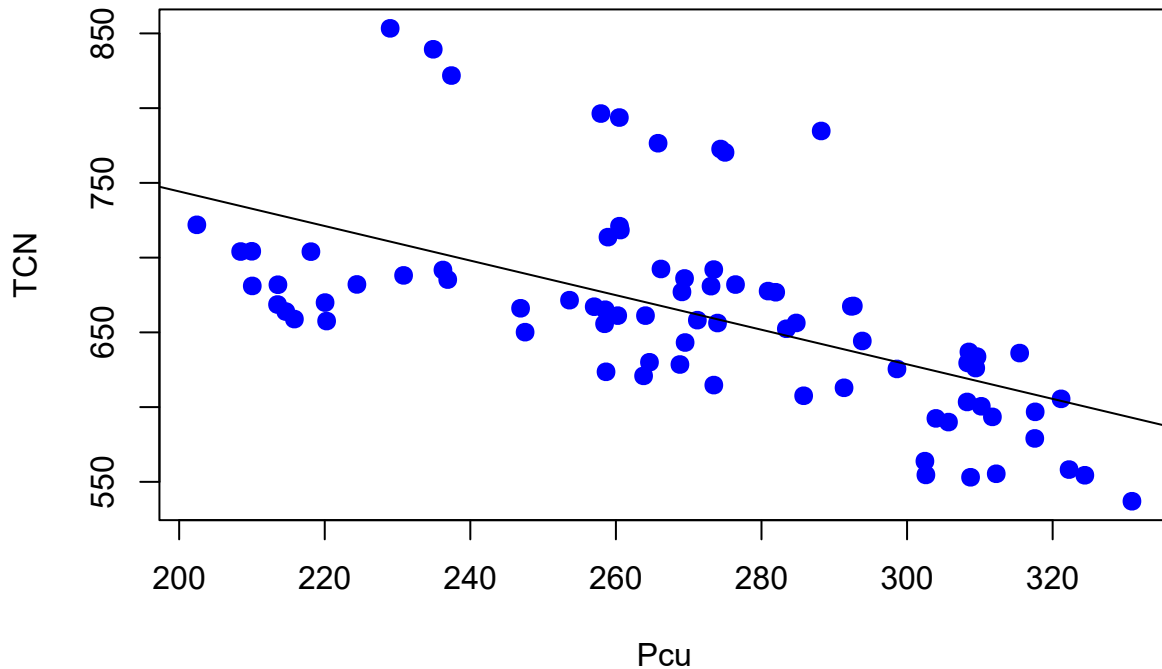
```
##
## Call:
## lm(formula = TCN ~ Pcu, data = data_subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -70.99 -36.83 -11.43  23.39 142.63
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  975.5807    49.1646  19.843  < 2e-16 ***
## Pcu          -1.1565     0.1807  -6.402  1.1e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.08 on 77 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.3473, Adjusted R-squared:  0.3389
## F-statistic: 40.98 on 1 and 77 DF, p-value: 1.103e-08
```

Veamos un gráfico relacionado:

```
plot(data_subset$Pcu, data_subset$TCN,
     pch = 16, cex = 1.3, col = "blue",
     main = "Relación TCN y Pcu",
     xlab = "Pcu", ylab = "TCN")

abline(lm(TCN ~ Pcu, data = data_subset))
```

Relación TCN y Pcu



- ¿Cómo interpretaría usted estos valores?

Pensemos en el ejemplo del video, ¿existe alguna particularidad en los valores de la estimación? Los valores asociados a los parámetros estimados son particularmente grandes (especialmente el valor de la constante).

Una solución a esto, muy utilizada en la práctica, es la utilización de los logaritmos de las variables. Esto tiene dos ventajas:

1. Soluciona problemas de escala.
2. Permite la interpretación en porcentajes.

Un último ejemplo con la regresión logaritmos:

```
modelo_03 <- lm(ln_tcn ~ ln_pcu, data = data_subset)
```

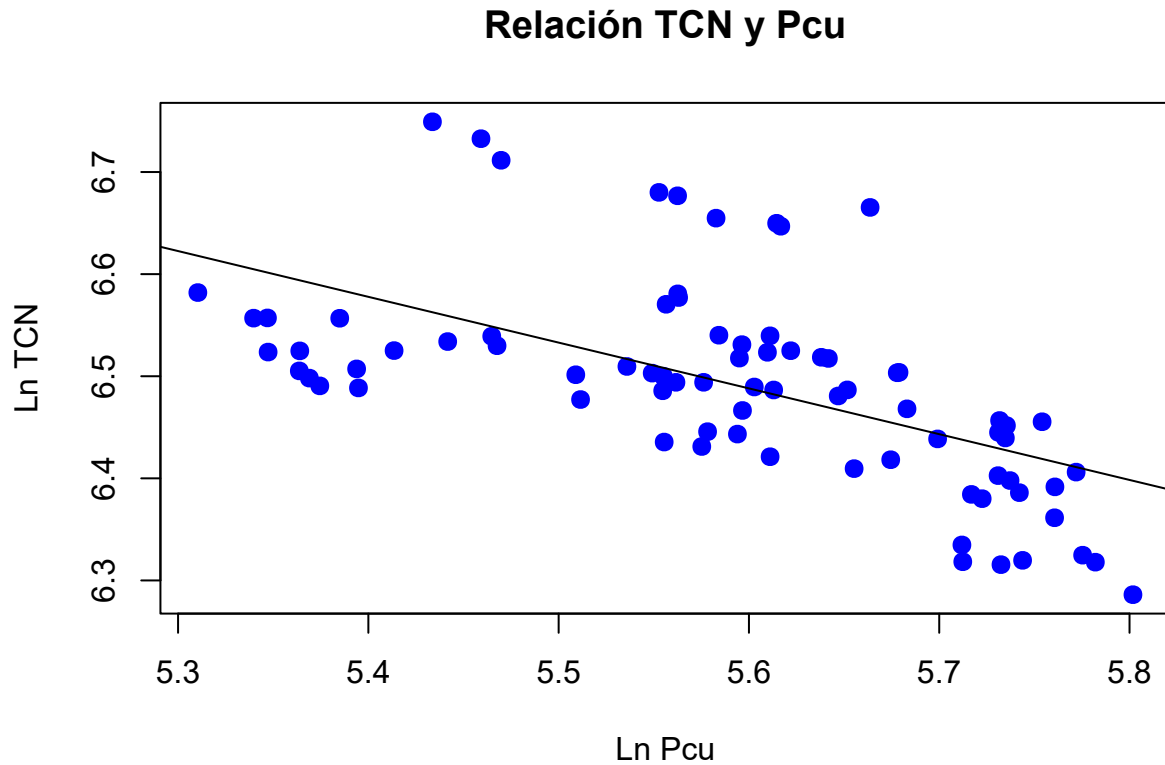
```
summary(modelo_03)
```

```
##
## Call:
## lm(formula = ln_tcn ~ ln_pcu, data = data_subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11939 -0.05259 -0.01125  0.04046  0.20579
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.99887    0.39263  22.919 < 2e-16 ***
## ln_pcu       -0.44835    0.07021  -6.386 1.18e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.08006 on 77 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared: 0.3462, Adjusted R-squared: 0.3377
## F-statistic: 40.78 on 1 and 77 DF, p-value: 1.181e-08
```

```
plot(data_subset$ln_pcu, data_subset$ln_tcn,
     pch = 16, cex = 1.3, col = "blue",
     main = "Relación TCN y Pcu",
     xlab = "Ln Pcu", ylab = "Ln TCN")

abline(lm(ln_tcn ~ ln_pcu, data = data_subset))
```



2.5. ¿Cómo analizar una regresión?

¿Cómo analizar una regresión?

- Significancia individual

Por ejemplo para la constante, se evalúa la siguiente hipótesis:

Hipótesis Nula o H_0

$$\beta_1 = 0 \quad (5)$$

Hipótesis Alternativa o H_1

$$\beta_1 \neq 0 \quad (6)$$

Se busca el poder rechazar la hipótesis nula, ¿pero por qué?

Este proceso se puede realizar de dos maneras (hay más):

- Estadístico calculado

Por ejemplo para el caso de la constante, este valor corresponde a 2,081 según la siguiente fórmula:

$$t_{calculado} = \frac{Coficiente}{error\ estándar\ del\ coeficiente} \quad (7)$$

lo que debe contrastarse con valores de tabla, según el siguiente detalle:

$$t_{tabla} = t - student_{10\%,n-1} \quad (8)$$

Este valor de probabilidad se evalúa generalmente al 10 %, al 5 % y al 1 %, y se le denomina valor significativo.

- P-Valor

Corresponde a la probabilidad acumulada por el valor t calculado, considerando el espacio desde el estadístico hacia el final o inicio de la distribución.

Se contrasta contra el valor de significancia. En caso de P-valor ser menor que el nivel de significancia, ya sea (1 %, 5 % o 10 %), se rechaza la hipótesis nula.

- Significancia global

Se aplica lo mismo, pero considerando el siguiente test:

Hipótesis Nula o H_0

$$\beta_2 = \beta_3 \dots \beta_n = 0 \quad (9)$$

Hipótesis Alternativa o H_1

$$\beta_2 = \beta_3 \dots \beta_n \neq 0 \quad (10)$$

No se considera la constante, y se considera lo mismo para la significancia individual pero con una distribución F.

- Bondad del Ajuste

Se relaciona con el valor del R cuadrado y R cuadrado Ajustado. Nos habla del ajuste del modelo un modelo de tipo lineal. El segundo considera el ajuste por el número de variables independientes.

Veamos esto en relación al último modelo calculado:

```
modelo_03 <- lm(ln_tcn ~ ln_pcu, data = data_subset)
```

```
summary(modelo_03)
```

```
##
## Call:
## lm(formula = ln_tcn ~ ln_pcu, data = data_subset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11939 -0.05259 -0.01125  0.04046  0.20579
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.99887     0.39263   22.919 < 2e-16 ***
## ln_pcu        -0.44835     0.07021   -6.386 1.18e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.08006 on 77 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared: 0.3462, Adjusted R-squared: 0.3377
## F-statistic: 40.78 on 1 and 77 DF, p-value: 1.181e-08
```

2.6. ¿Cómo aplicamos esto al proyecto?

Teniendo las estimaciones del tipo de cambio, deberíamos poder generar las estimaciones de las cantidades a vender. Conociendo el precio, podríamos multiplicar todos los factores y obtenemos nuestro valor inicial.

Por ejemplo:

$$Ventas_{Año1} = TCN_{Año1} * Cantidad_{Año1} = 720 * 1,000 = 720,000 \quad (11)$$

- ¿Ve algún problema en esto? Lo contestaremos más adelante.

Desde este punto, solo queda poder aplicar un crecimiento a estas ventas para los otros años a estimar.

- TIP: Truco clásico, es aplicar la proyección del PIB del país más o menos algo. Pero, ¿de qué depende esto realmente? ¿quién implica basarnos en el PIB?

2.7. ¿Cómo se aplica esto en la práctica

Esto podemos explicarlo partiendo desde una pregunta:

- Tiene dos modelos de estimación de sus ventas: el primero, en base a un modelo econométrico y el segundo en base a una pregunta realizada a sus propios vendedores, sobre cuánto esperan vender ellos. ¿En cuál estimación o proyección confía más?
- Por lo demás, existe un problema práctico sobre dichos modelos, si lo vemos en su especificación estimamos el siguiente modelo:

$$TCN = \alpha + \beta * Pcu \quad (12)$$

El problema corresponde a algo que se deja entrever en la clase de los videos. ¿Cuál es la temporalidad de la estimación? ¿Qué implica esto?

3. Un ejemplo práctico de la estimación de las ventas

Lo primero que debemos considerar acá es la existencia de información histórica o no. En el caso de tenerla, todo se simplifica. La principal problemática siempre, corresponde a poder obtener un valor inicial desde donde partir nuestra proyección.

Se presenta a continuación un ejemplo en el contexto de contar con información histórica, como también un caso en dónde no.

3.1. Supuestos de la estimación

En relación al ejemplo, algo que debe considerarse necesario para esto es la validación de los supuestos de su estimación.

- Por ejemplo, los vendedores son un buen punto de partida por su conocimiento en ventas y el carácter motivacional que tiene el cumplimiento de una meta (se debe considerar acá la relación entre sueldo variable y fijo).
- Por otra parte, un modelo econométrico para el TCN es bueno en la medida en que existe una relación real con las variables utilizadas y se cumplan los requerimientos estadísticos de significancia.
- A la vez, una encuesta realizada a los futuros clientes, cumpliendo con todos los requerimientos propios de la construcción y validación de encuestas, también es un buen punto de partida.

Pero cabe preguntarse esto, ¿cómo se relaciona esto con la etapa del proyecto que estamos generando? ¿todos los modelos aplican para una evaluación a nivel de perfil o de prefactibilidad? En este sentido, no solo los supuestos influyen, sino el tipo la etapa del proyecto y su influencia sobre la evaluación a realizar.

4. Fechas relevantes

Unidad	Evaluación	Ponderación	Fecha
Unidad I	Cuestionario - Semana 2 - No presencial	(5 %)	31/03/2021
	Estudio de Caso - Semana 4 - No presencial	(5 %)	18/04/2021
	Ejercicio práctico - Semana 7 - No presencial	(5 %)	09/05/2021
	Prueba Escrita - Semana 9 - Presencial I	(35 %)	19/05/2021
Unidad II	Ejercicio práctico - Semana 11 - No presencial	(5 %)	06/06/2021
	Estudio de Caso - Semana 13 - No presencial	(5 %)	20/06/2021
	Ejercicio práctico - Semana 15 - No presencial	(5 %)	04/07/2021
	Prueba Escrita - Semana 17 - Presencial	(35 %)	07/07/2021