Equipo 6
Álvaro Acedo 174052
Diana Santiago 175325
Jimena Reyes 173361
Paulina Mazariegos 171929
Pamela Goya 171789

ESTADISTICA APLICADA II
Tarea No. 3

Dr. Víctor M. Guerrero
Ago-Dic, 2021

1. Una compañía que se dedica a la venta y reparación de PC's quiere conocer el número de ingenieros de servicio que requerirá en los próximos años.

   Un elemento a considerar para determinar dicho número es el tiempo que se lleva cada servicio, lo cual depende del total de componentes de cada PC que deben ser reparadas o reemplazadas.

   Para establecer esta relación se eligió dentro del registro de la compañía, una muestra aleatoria de los servicios realizados.

   Los datos corresponden al Tiempo de Servicio (en minutos) y al Número de Componentes reparadas.

Componentes	1	2	3	4	4	5	6	6	7	8	9	9	10	10
Tiempo	23	29	49	64	74	87	96	97	109	119	149	145	154	166

a) **Haga** un diagrama de dispersión e **indique** por qué un modelo de regresión lineal simple podría ser útil para explicar el Tiempo de Servicio en función del Número de Componentes reparadas.

b) **Realice** la estimación de los parámetros involucrados en el modelo, **interprete** los valores e **indique** el porcentaje de variabilidad que se haya logrado explicar.

c) **Obtenga** un intervalo de 95% de confianza para la pendiente de la recta de regresión.

d) **Calcule** los valores estimados por el modelo y los residuos. **Verifique** que la suma de los residuos sea igual a cero.

e) **Pronostique** el Tiempo de Servicio que se llevaría reparar una PC con 4 componentes en mal estado y **calcule** el error estándar del pronóstico.

f) El gerente de la compañía esperaría que el incremento en tiempo, para cada componente adicional que requiere reparación, fuera de 12 minutos. ¿Es razonable este valor?, **explique** por qué sí o por qué no lo es.

2. El problema de seleccionar una determinada forma funcional en el modelo de regresión es crucial para interpretar la relación que existe entre las variables del modelo.

   Dentro de los modelos lineales (en los parámetros) se encuentra la familia de modelos dictada por la siguiente ecuación, que involucra variables transformadas y que puede considerarse básica

$$Y^{(\lambda_1)} = \beta_0 + \beta_1 X^{(\lambda_2)} + \varepsilon .$$

   En esta ecuación aparecen variables transformadas del tipo genérico

$$Z^{(\lambda)} = \begin{cases} \dfrac{Z^{\lambda} - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \log(Z) & \text{si } \lambda = 0 \end{cases}$$

   en donde la variable debe ser real y positiva, o sea, $Z > 0$.

   Con los siguientes valores de λ_1 y λ_2, **realice** las gráficas de Y *vs*. X para las funciones descritas por la ecuación básica, en el **supuesto de que el error es 0,** NOTE QUE AMBAS VARIABLES DEBEN SER POSITIVAS**.**

   Haga explícito el **papel que juegan** los parámetros β_0 y β_1 en tales funciones.

   a) $\lambda_1 = \lambda_2 = 1$;

   b) $\lambda_1 = \lambda_2 = 0$;

   c) $\lambda_1 = 1,\ \lambda_2 = 0$;

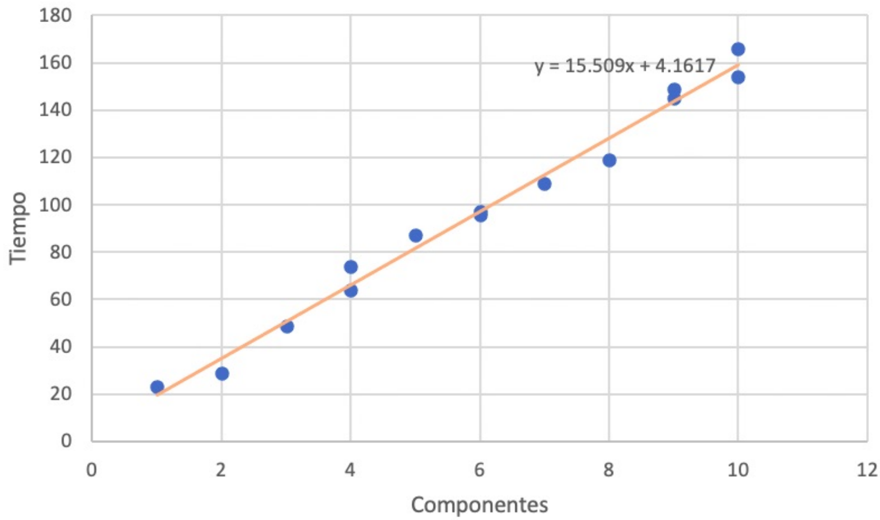
   d) $\lambda_1 = 1,\ \lambda_2 = -1$;

   e) $\lambda_1 = 0,\ \lambda_2 = 1$;

   f) $\lambda_1 = 0,\ \lambda_2 = -1$

a) **Haga** un diagrama de dispersión e **indique** por qué un modelo de regresión lineal simple podría ser útil para explicar el Tiempo de Servicio en función del Número de Componentes reparadas.

Sean X: componentes
Y: tiempo.



El modelo de regresión lineal simple resulta útil en este caso porque se cumplen todas las condiciones:

1. Existe una única variable explicativa (el número de componentes)
2. La gráfica nos muestra una tendencia del tipo de línea recta, por lo que se puede hacer referencia a una relación lineal en las variables.
3. Asumimos linealidad en los parámetros.

Bien

Es por eso que se propone un modelo del tipo:

$$Y = \beta_0 + \beta_1 X$$

b) **Realice** la estimación de los parámetros involucrados en el modelo, **interprete** los valores e **indique** el porcentaje de variabilidad que se haya logrado explicar.

Tenemos que: (cálculos en excel)

$$\sum_{i=1}^{14} X_i = 84 \Rightarrow \bar{X} = \frac{84}{14} = 6 \qquad \sum_{i=1}^{14} X_i^2 = 618$$

$$\sum_{i=1}^{14} Y_i = 1361 \Rightarrow \bar{Y} = \frac{1361}{14} = 97.2142 \qquad \sum_{i=1}^{14} y_i^2 = 160077$$

$$\sum_{i=1}^{14} x_i y_i = 9934$$

Usando las siguientes fórmulas:

$$S_{XY} = \frac{\sum X_i Y_i - \sum X_i \sum Y_i / n}{n-1} = \text{Covarianza muestral de X y Y.}$$

$$S_X^2 = \frac{\sum X_i^2 - (\sum X_i)^2 / n}{n-1} = \text{Varianza muestral de X.}$$

$$\Rightarrow S_{xy} = 136 \qquad \Rightarrow b_1 = \frac{S_{xy}}{S_x^2} = 15.5087$$

$$S_x^2 = 8.7692$$

Bien

$$\Rightarrow b_0 = \bar{Y} - b_1 \bar{X} = 4.1616$$

$\therefore$ La recta de regresión es: $Y = 4.1616 + 15.5087X$

b0 es el valor autónomo de Y:
Con un valor de X=0 (es decir, 0 componentes), la variable Y tomaría el valor b0 (es decir, el tiempo sería 15.5087)
Podemos interpretar este valor como el tiempo mínimo requerido para iniciar la producción de componentes.

No se trata de producir, sino reparar y en es

b1 es la pendiente de la recta de regresión:
Podemos interpretar este valor como el tiempo marginal que toma la producción de cada componente adicional.

Notar que:

$$SCE = \sum(\hat{Y}_i - \bar{Y})^2 = 27419.509$$

$$SCR = \sum(Y_i - \hat{Y}_i)^2 = 348.8483$$

$$SCT = \sum(Y_i - \bar{Y})^2 = 27768.3571$$

$\Rightarrow \dfrac{SCE}{SCT} = 98.71\%$ es el porcentaje de variabilidad que se logró explicar con el modelo

$\dfrac{SCR}{SCT} = 1.29\%$

Bien

Lo ideal es que:

$$\frac{SCE}{SCT} \approx 1 \qquad \frac{SCR}{SCT} \approx 0$$

$\therefore$ Nuestro modelo resultó ser bastante bueno para explicar el fenómeno.

c) **Obtenga** un intervalo de 95% de confianza para la pendiente de la recta de regresión.

Bajo el supuesto de que los errores del modelo tienen una distribución normal, podemos calcular un IC al $(1-\alpha)\%$ para β_1 de la siguiente forma:

$$\beta_1 \in \left\{ b_1 \pm t_{(n-2),\alpha/2}\, \hat{e}e(b_1) \right\}$$

Sean: $n = 14$

$\alpha = 0.05 \implies \beta_1 \in \{ b_1 \pm t_{(12),0.025} \cdot \hat{e}e(b_1) \}$

C.aux:

$t_{(12),0.025} = 2.1788$

$$\hat{e}e(b_1) = S \frac{1}{\sqrt{\sum x_i^2}}$$

Sea $S = \sqrt{\frac{SCR}{n-2}} = \sqrt{\frac{348.8483}{12}} = 5.3917$

$$\sum_{i=1}^{14} x_i^2 = 114 \implies \hat{e}e(b_1) = \frac{5.3917}{\sqrt{114}} = 0.5049$$

$\therefore L = b_1 - t_{(12),0.025} \cdot \hat{e}e(b_1) = 15.5087 - 2.1788(0.5049) =$
$= 14.4085$

$U = b_1 + t_{(12),0.025} \cdot \hat{e}e(b_1) = 15.5087 + 2.1788(0.5049) =$
$= 16.6090$

Bien

$\therefore \beta_1 \in [14.4085, 16.6090]$ con un 95% de confianza

d) **Calcule** los valores estimados por el modelo y los residuos. **Verifique** que la suma de los residuos sea igual a cero.

X_i	Y_i	Estimación $\hat{Y}_i$	Residuo $Y_i - \hat{Y}_i$
1	23		
2	29		
3	49		
4	64	66.19674185	-2.19674185
4	74	66.19674185	7.803258145
5	87	81.70551378	5.294486216
6	96	97.21428571	-1.21428571
6	97	97.21428571	-0.21428571
7	109	112.7230576	-3.72305764
8	119	128.2318296	-9.23182957
9	149	143.7406015	5.259398496
9	145	143.7406015	1.259398496
10	154	159.2493734	-5.24937343
10	166	159.2493734	6.750626566
		Suma	**0.00**

No veo los valores, ni la verificación. -8

e) **Pronostique** el Tiempo de Servicio que se llevaría reparar una PC con 4 componentes en mal estado y **calcule** el error estándar del pronóstico.

Definimos $X_0 = 4$

El pronóstico para Y_0 está dado por:

$$\hat{Y}_0 = \bar{Y} + b_1 (X_0 - \bar{X})$$

$$\Rightarrow \hat{Y}_0 = 97.2142 + 15.5087(4-6) = 66.1967$$

El error éstandar del pronóstico es:

$$\hat{ee}(\hat{y}_0) = \sqrt{\hat{Var}(\hat{y}_0)} \quad \text{donde } \hat{Var}(\hat{y}_0) = \frac{\sigma^2}{n} + (X_0 - \bar{X})^2 \left[\frac{\sigma^2}{\sum (X_i - \bar{X})^2} \right]$$

Sustituyendo $\sigma^2 = S^2$

$$\Rightarrow \hat{Var}(\hat{y}_0) = S^2 \left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right]$$

$$= 5.3917 \left(\frac{1}{14} + \frac{(4-6)^2}{114} \right) = 3.0965$$

$\Rightarrow \hat{ee}(\hat{y}_0) = 1.7596$ Está mal, debe ser 5.6716 -3

f) El gerente de la compañía esperaría que el incremento en tiempo, para cada componente adicional que requiere reparación, fuera de 12 minutos. ¿Es razonable este valor?, **explique** por qué sí o por qué no lo es.

Esperar que el incremento en tiempo , para cada componente adicional que requiere reparación, es de 12 minutos no es razonable.

En el inciso anterior, obtuvimos que un intervalo del 95% de confianza para Beta 1(el cambio del tiempo debido a un cambio unitario en el componente) está dado por:

(14.4085, 16.6090)

Bien

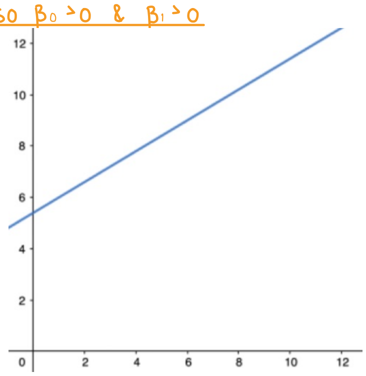
Por lo que el 95% de todos los posibles valores estimados para el parámetro tomarán un valor dentro de este intervalo. Nótese que 12 no pertenece al intervalo. Por lo tanto, los 12 minutos en los que el gerente espera que se repare un componente adicional no es razonable.

② Nota. para las siguientes gráficas solo se consideran los casos en los que X & $Y > 0$

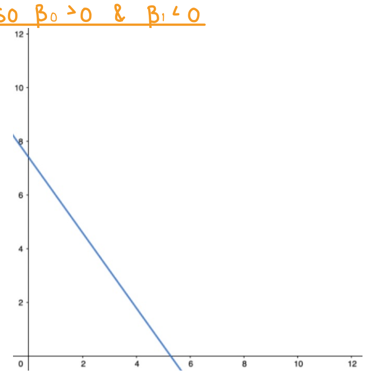
a) $\lambda_0 = \lambda_1 = 1$

$$Y = \beta_0 + 1 + \beta_1 (X - 1)$$

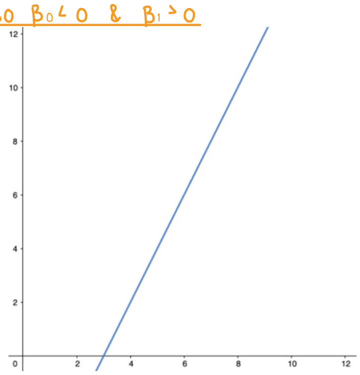
caso $\beta_0 > 0$ & $\beta_1 > 0$



caso $\beta_0 > 0$ & $\beta_1 < 0$



caso $\beta_0 < 0$ & $\beta_1 > 0$



caso $\beta_0 < 0$ & $\beta_1 < 0$

Nota: cuando β_0 y $\beta_1 < 0$, X & $Y < 0$

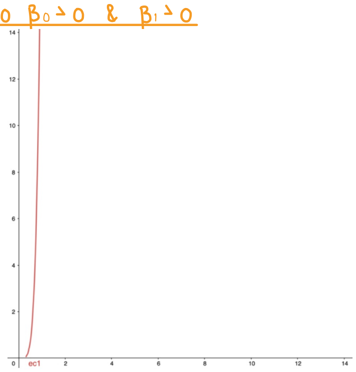
Bien

Papel que juegan los parámetros $\begin{cases} \beta_0 + 1 - \beta_1 : \text{ordenada al origen} \\ \beta_1 : \text{pendiente} \end{cases}$

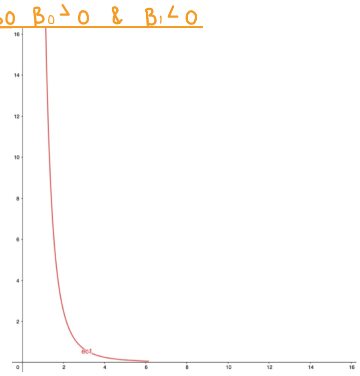
b) $\lambda_1 = \lambda_2 = 0$

$$\ln(Y) = \beta_0 + \beta_1 \ln(X)$$

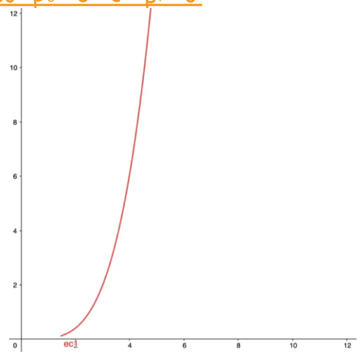
caso $\beta_0 > 0$ & $\beta_1 > 0$



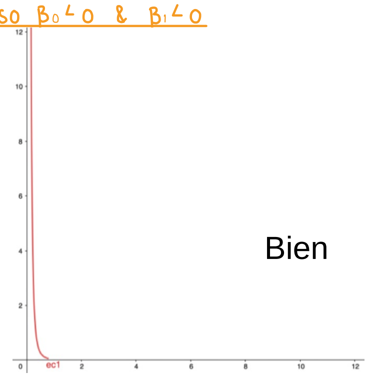
caso $\beta_0 > 0$ & $\beta_1 < 0$



caso $\beta_0 < 0$ & $\beta_1 > 0$



caso $\beta_0 < 0$ & $\beta_1 < 0$



Bien

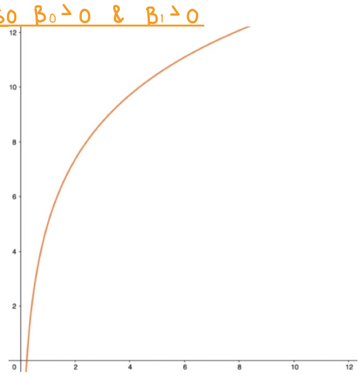
Papel que juegan los parámetros

- β_0 : qué tan separada está la gráfica del eje de las ordenadas
  entre más grande sea, más lejos del eje de las ordenadas
- β_1 : tasa a la que cambia Y cuando cambia X (no es una relación lineal)

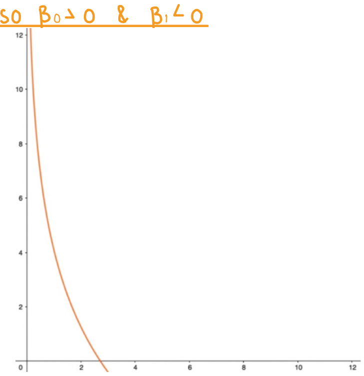
c) $\lambda_1 = 1$ & $\lambda_2 = 0$

$$Y = 1 + \beta_0 + \beta_1 \ln(X)$$

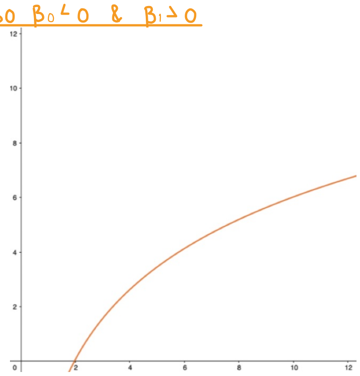
caso $\beta_0 > 0$ & $\beta_1 > 0$



caso $\beta_0 > 0$ & $\beta_1 < 0$

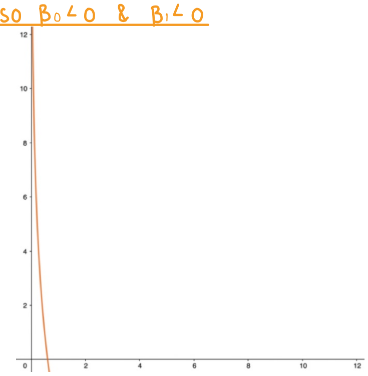


caso $\beta_0 < 0$ & $\beta_1 > 0$



Bien

caso $\beta_0 < 0$ & $\beta_1 < 0$



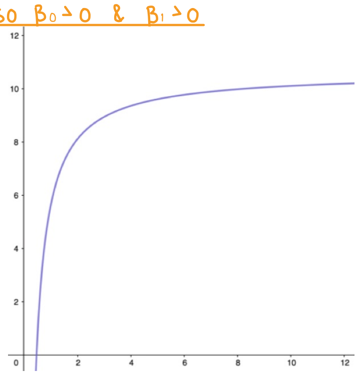
Papel que juegan los parámetros

- β_0 : qué tan separada está la gráfica del eje de las ordenadas y más cerca del de las abscisas
  entre más grande sea, más lejos del eje de las ordenadas
- β_1 : cambio no lineal en Y cuando cambia X

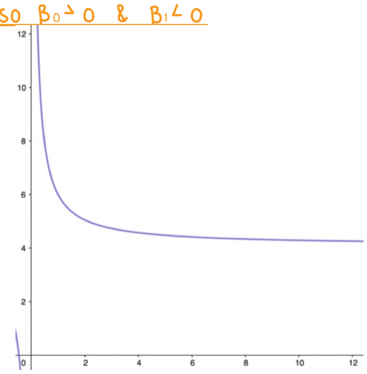
d) $\lambda_1 = 1$ & $\lambda_2 = -1$

$Y = \beta_0 + 1 - \beta_1 (X^{-1} - 1)$

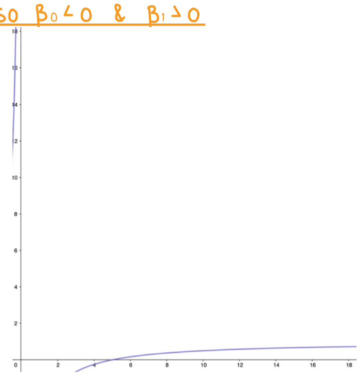
caso $\beta_0 > 0$ & $\beta_1 > 0$



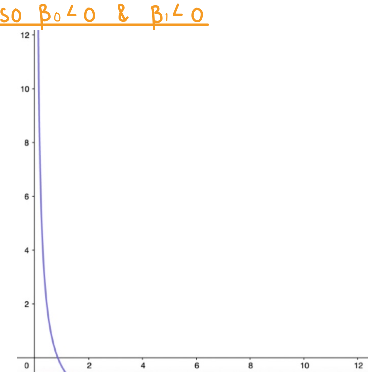
caso $\beta_0 > 0$ & $\beta_1 < 0$



caso $\beta_0 < 0$ & $\beta_1 > 0$



caso $\beta_0 < 0$ & $\beta_1 < 0$



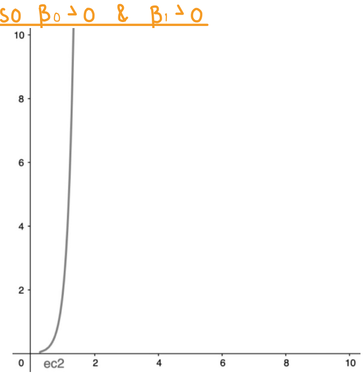
Bien

Papel que juegan los parámetros $\left\{ \begin{array}{l} \beta_0 : \text{qué tan pegada está el origen de la hipérbola a } 0 \\ \beta_1 : \text{si } \beta_1 > 0 \Longrightarrow \quad , \text{ si } \beta_1 < 0 \Longrightarrow \\ \quad \text{i.e: define si es cóncava o convexa} \end{array} \right.$

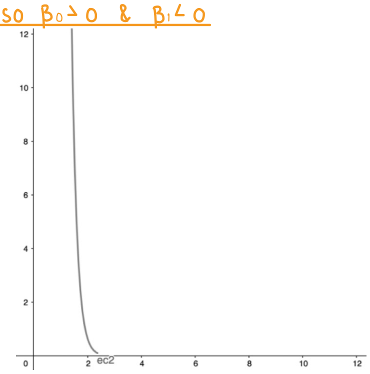
e) $\lambda_1 = 0$ & $\lambda_2 = 1$

$\ln(Y) = \beta_0 + \beta_1 (X - 1)$

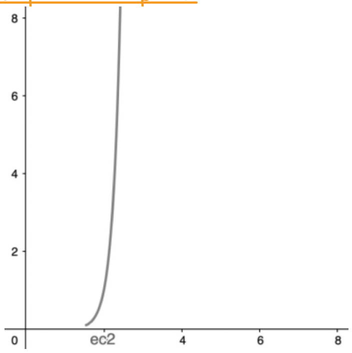
caso $\beta_0 > 0$ & $\beta_1 > 0$



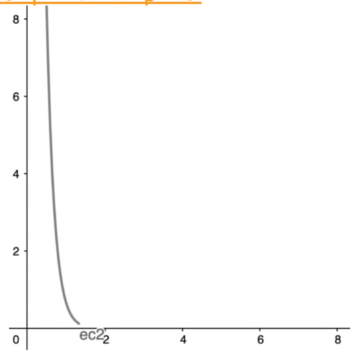
caso $\beta_0 > 0$ & $\beta_1 < 0$



caso $\beta_0 < 0$ & $\beta_1 > 0$



caso $\beta_0 < 0$ & $\beta_1 < 0$



Bien

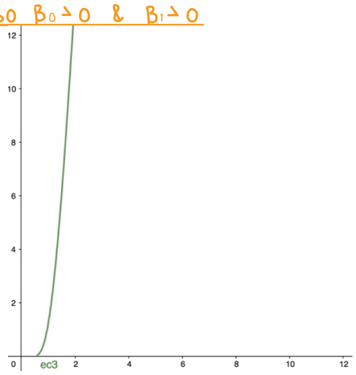
Papel que juegan los parámetros

β_0: qué tan pegada está la gráfica del eje de las ordenadas
para $\beta_0 \geq 0$, entre más grande sea, más lejos del eje de las ordenadas
β_1: cambio no lineal en Y cuando cambia X

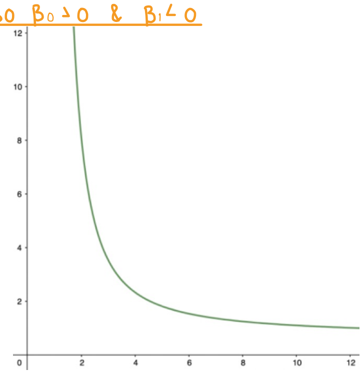
f) $\lambda_1 = 0$ & $\lambda_2 = -1$

$$\ln(Y) = \beta_0 - \beta_1 (X^{-1} - 1)$$

caso $\beta_0 > 0$ & $\beta_1 > 0$

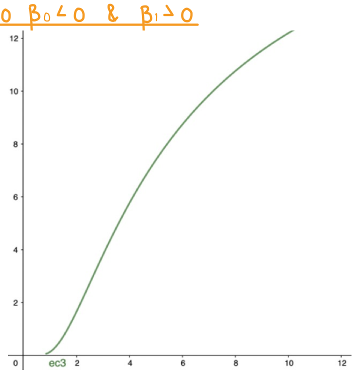


caso $\beta_0 > 0$ & $\beta_1 < 0$

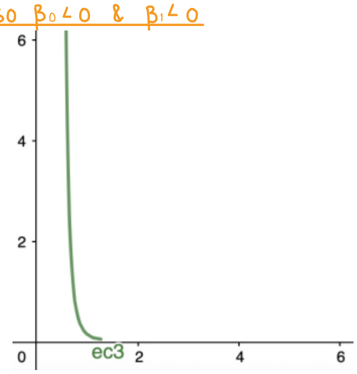


Bien

caso $\beta_0 < 0$ & $\beta_1 > 0$



caso $\beta_0 < 0$ & $\beta_1 < 0$



Papel que juegan los parámetros

β_0: qué tan pegada está la gráfica del eje de las ordenadas
β_1: cambio no lineal en Y cuando cambia X