

# Modelo de regresión lineal simple

Joel Alejandro Zavala Prieto

## Contents

<b>Informacion de contacto</b>	<b>2</b>
<b>1) Explicación de los distintos modelos</b>	<b>3</b>
Modelo lineal . . . . .	3
Modelo log-log o doble logarítmico . . . . .	4
Modelo semi-logarítmico: lineal-logarítmico . . . . .	6
Modelo semi-logarítmico: logarítmico-lineal . . . . .	8
Modelo logarítmico recíproco . . . . .	10
<b>2) ¿Qué son las pruebas de hipótesis y para qué sirven?</b>	<b>12</b>
<b>3) Datos de consumo privado nacional y el ingreso disponible total</b>	<b>13</b>
Modelo lineal asociado . . . . .	13
Calculando las medidas requeridas . . . . .	13
Estimando modelo lineal por MCO . . . . .	15
Parametros obtenidos y modelo estimado . . . . .	15
Valores ajustados y residuales . . . . .	16
<b>4) Grafica de dispersion y linea de tendencia de los datos de consumo privado con ingreso disponible</b>	<b>17</b>

## Informacion de contacto

mail: [alejandro.zavala1001@gmail.com](mailto:alejandro.zavala1001@gmail.com)

Facebook: <https://www.facebook.com/AlejandroZavala1001>

# 1) Explicación de los distintos modelos

## Modelo lineal

El modelo lineal dado por:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Cuya función ajustada es:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Dando valores arbitrarios a:

$$\hat{\beta}_0 \quad , \quad \hat{\beta}_1$$

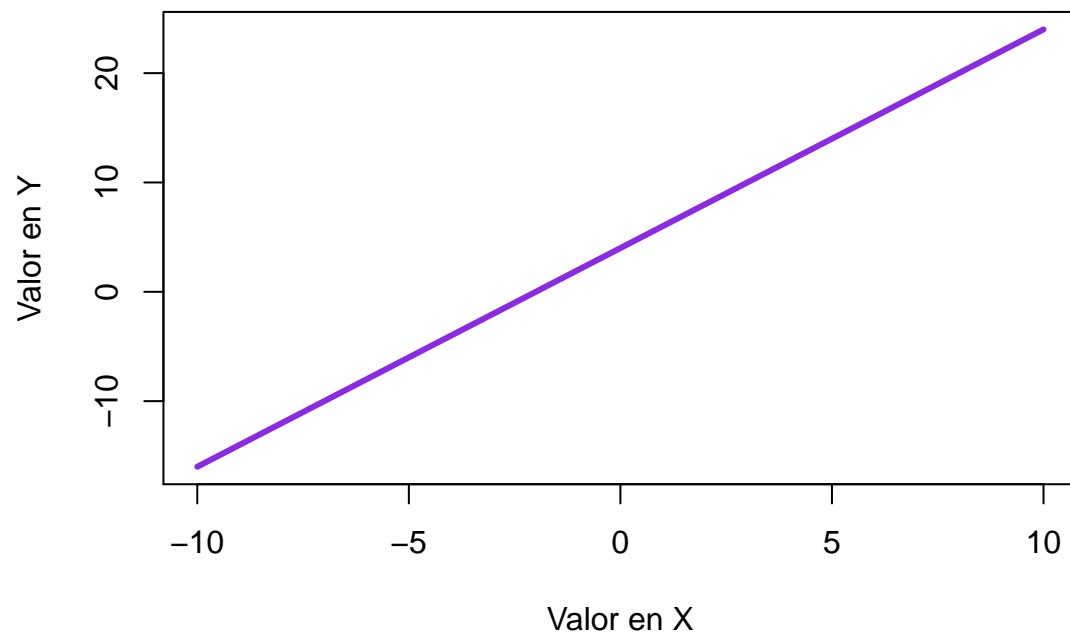
Se obtienen distintas rectas

$$\begin{aligned} \hat{\beta}_0 &= \text{ordenada al origen} \\ \hat{\beta}_1 &= \text{pendiente} \end{aligned}$$

Veamos el modelo

$$\hat{y}_i = 2 + 4x_i$$

## Modelos lineales



## Modelo log-log o doble logarítmico

El modelo lineal dado por:

$$\ln y_i = \beta_0 + \beta_1 \ln x_i + u_i$$

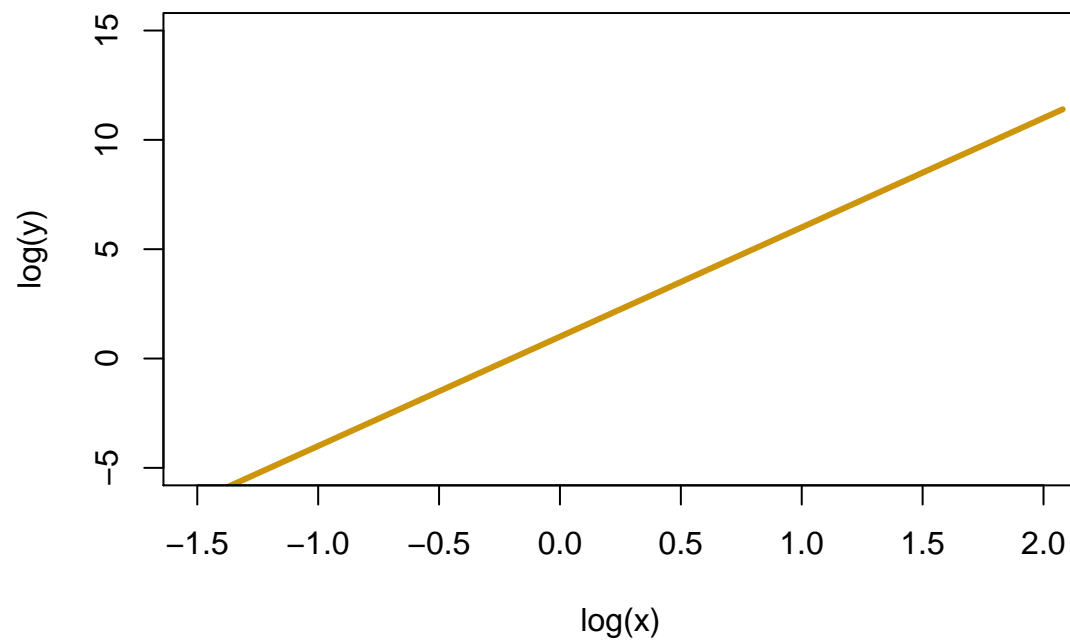
Cuya función ajustada es:

$$\begin{aligned}\ln \hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 \ln x_i \\ x_i &> 0\end{aligned}$$

Viendo el modelo:

$$\begin{aligned}\ln \hat{y}_i &= 1 + 5 \ln x_i \\ x_i &> 0\end{aligned}$$

### Modelos log-log



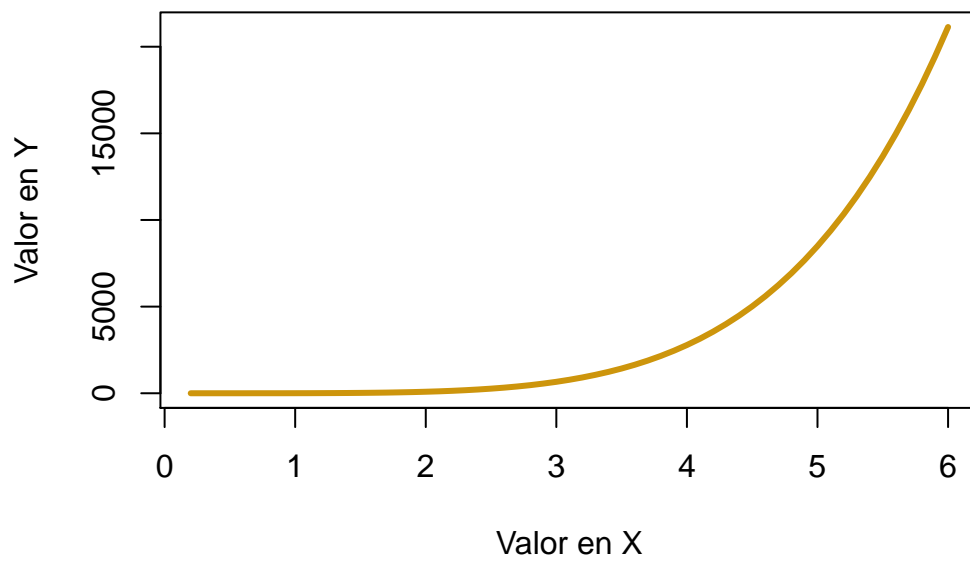
Notemos que el mismo modelo:

$$\begin{aligned}\ln \hat{y}_i &= 1 + 5 \ln x_i \\ x_i &> 0\end{aligned}$$

En otra forma tambien es:

$$\begin{aligned}\hat{y}_i &= e^{1+5 \ln x_i} \\ x_i &> 0\end{aligned}$$

### Modelos log-log



## Modelo semi-logarítmico: lineal-logarítmico

El modelo lineal logarítmico dado por:

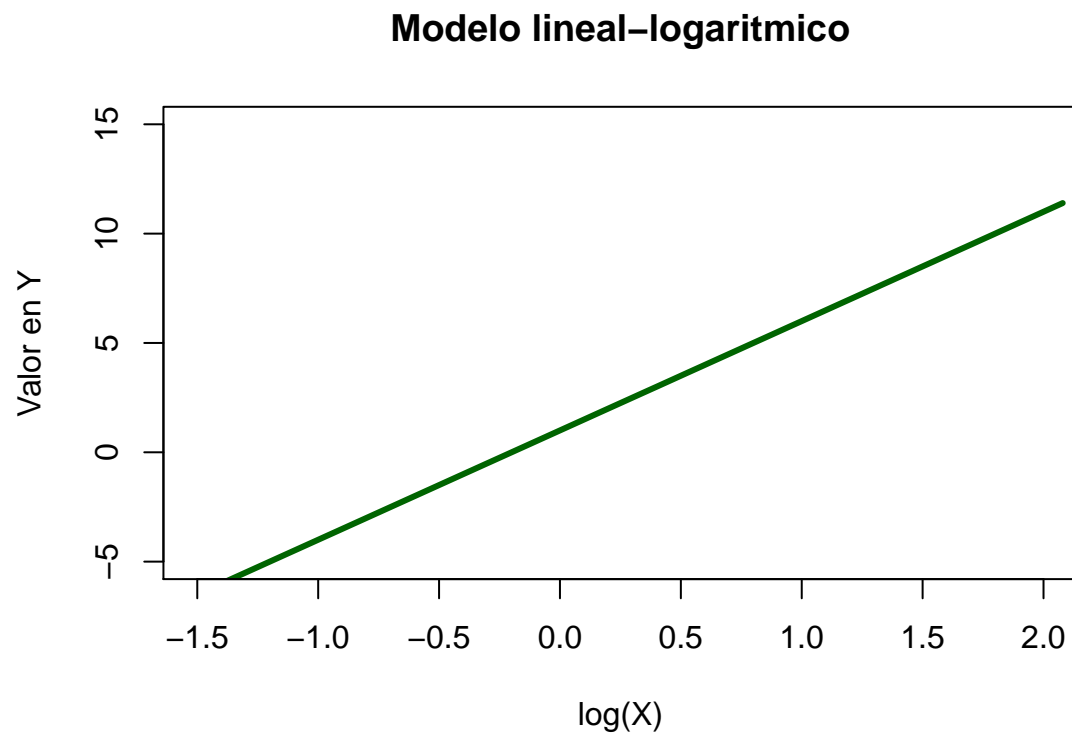
$$y_i = \beta_0 + \beta_1 \ln x_i + u_i$$
$$x_i > 0$$

Cuya función ajustada es:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \ln x_i$$
$$x_i > 0$$

Analizando el modelo:

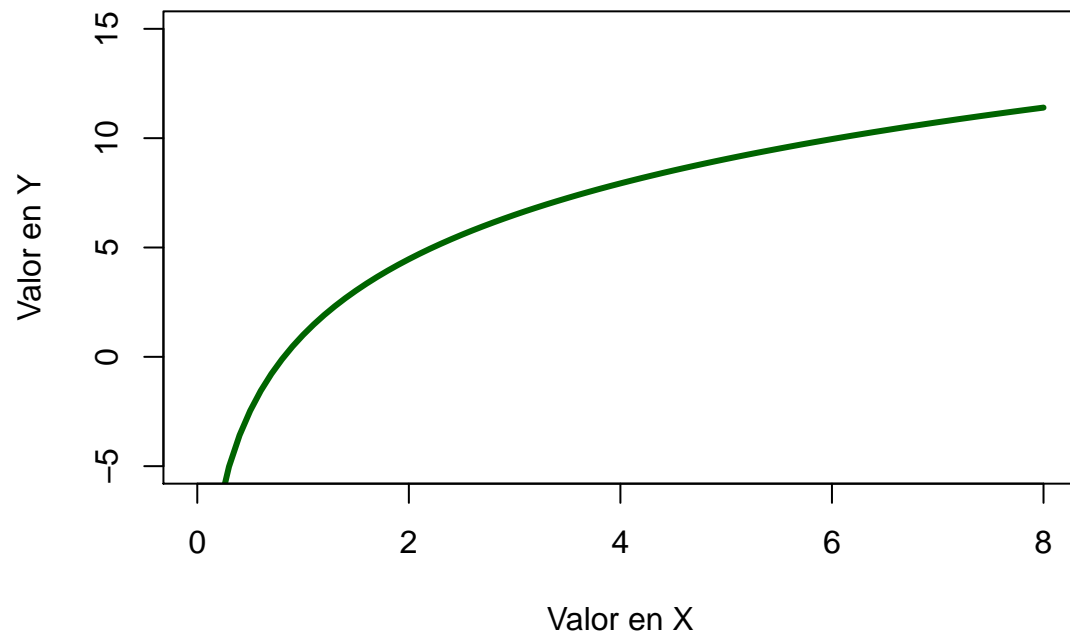
$$\hat{y}_i = 1 + 5 \ln x_i$$
$$x_i > 0$$



Cuya función ajustada también es:

$$\hat{y}_i = 1 + 5 \ln x_i$$
$$x_i > 0$$

### Modelos lineal-logarítmico



## Modelo semi-logarítmico: logarítmico-lineal

$$\ln y_i = \beta_0 + \beta_1 x_i + u_i$$
$$x_i > 0$$

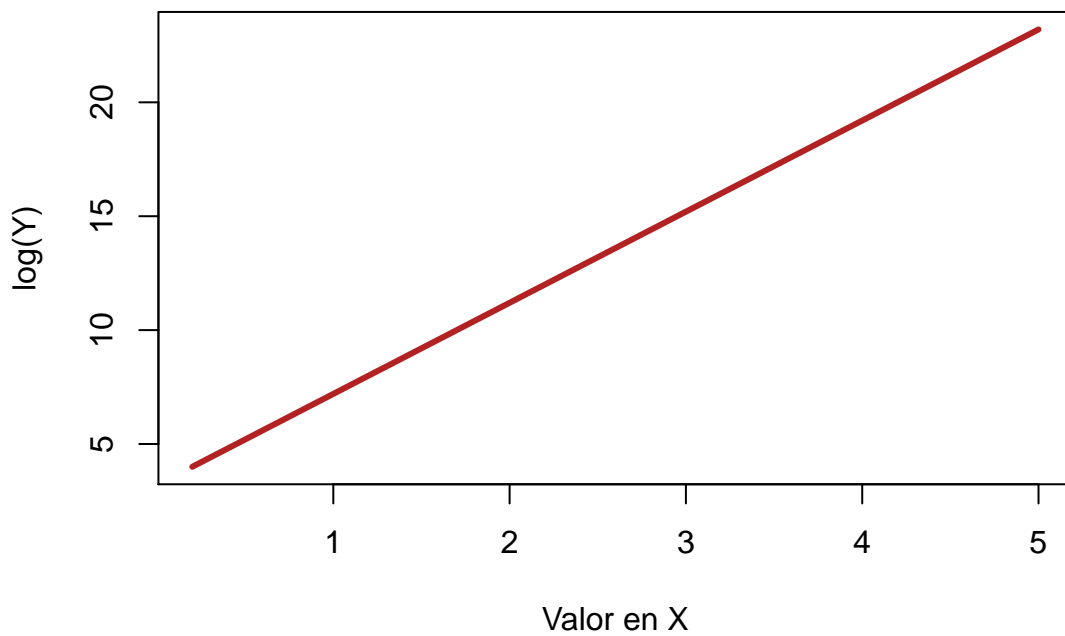
Cuya función ajustada es:

$$\ln \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$
$$x_i > 0$$

Analizando

$$\ln \hat{y}_i = 3.2 + 4x_i$$
$$x_i > 0$$

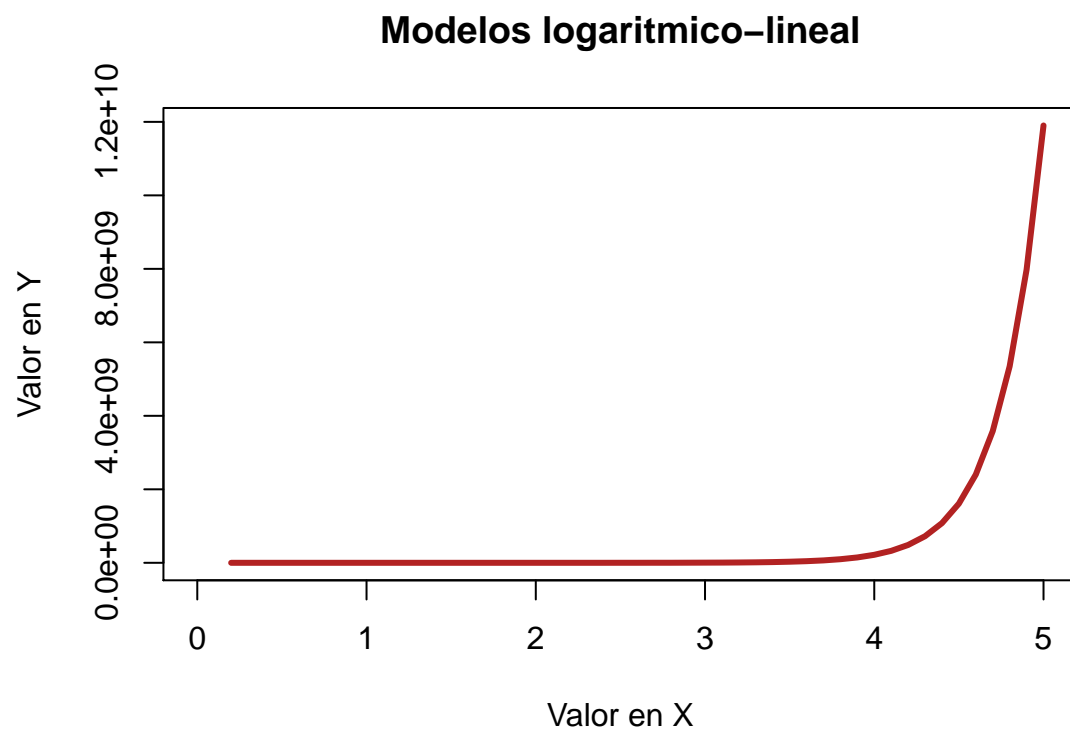
## Modelos logarítmico-lineal





Cuya representación también es

$$\hat{y}_i = e^{3.2+4x_i}$$
$$x_i > 0$$



## Modelo logarítmico recíproco

El modelo log-recíproco esta dado por:

$$\ln y = \beta_0 - \beta_1 \frac{1}{x} + u_i$$
$$x_i > 0$$

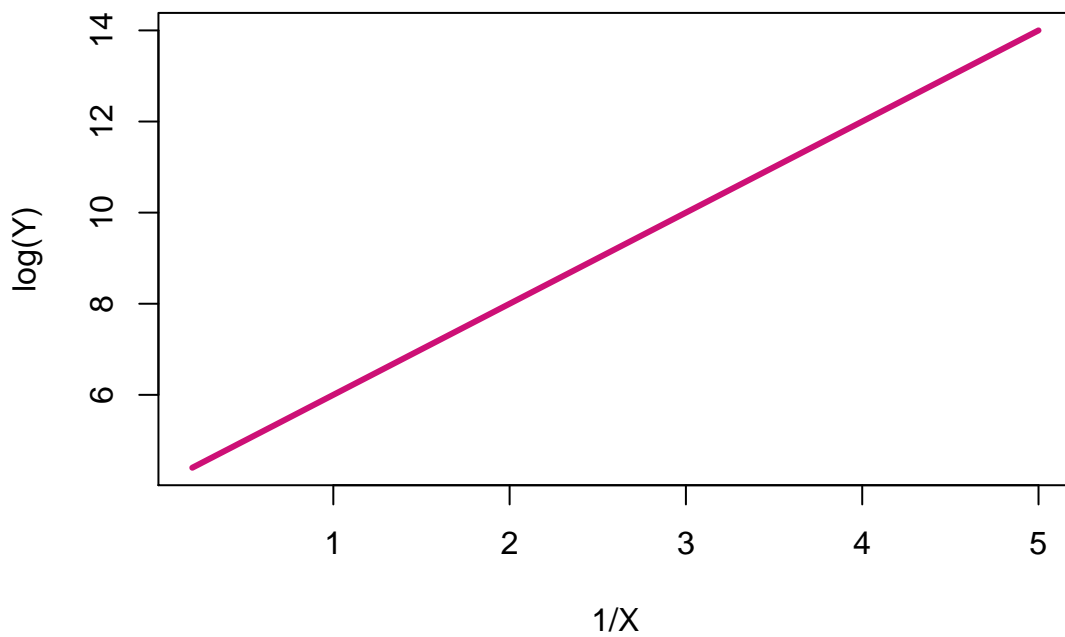
Cuya función ajustada es:

$$\ln \hat{y}_i = \hat{\beta}_0 - \hat{\beta}_1 \frac{1}{x}$$
$$x_i \neq 0$$

Analizando el siguiente caso:

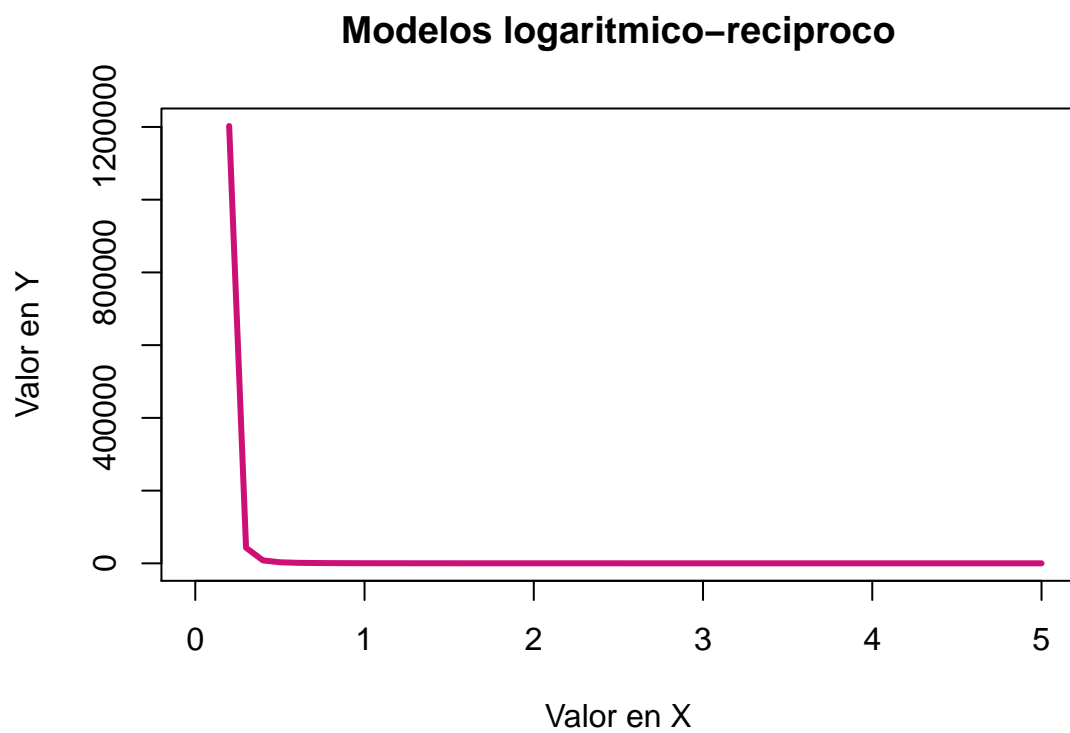
$$\ln \hat{y}_i = 4 - \frac{2}{x}$$
$$x_i \neq 0$$

## Modelos logaritmico-reciproco



Cuya representación también es:

$$\hat{y}_i = e^{4 - \frac{2}{x}}$$
$$x_i \neq 0$$



## 2) ¿Qué son las pruebas de hipótesis y para qué sirven?

Una prueba de hipótesis es una regla que especifica si se puede aceptar o rechazar una afirmación acerca de una población dependiendo de la evidencia proporcionada por una muestra de datos. Se hacen inferencias estadísticas en:

1. Medias
2. Proporciones
3. Diferencias entre medias
4. Diferencias entre proporciones
5. Medias con tamaño muestral menor a 30
6. Diferencias de media con tamaño muestral menor a 30
7. Varianzas

Una prueba de hipótesis examina dos hipótesis opuestas sobre una población: la hipótesis nula (si la media es igual a algo, si la proporción será cero o la diferencia nula) y la hipótesis alternativa. La hipótesis nula es el enunciado que se probará. Por lo general, la hipótesis nula es un enunciado de que “no hay efecto” o “no hay diferencia”.

Estas pruebas se realizan bajo el supuesto del teorema de límite central y dependiendo la distribución de sus variables por funciones conocidas, con apoyo de tablas de probabilidad se acepta o se rechaza la hipótesis a un nivel  $\alpha$  es decir:

$$1 - \alpha = P\left(\frac{x - \mu}{\sigma} > \frac{\tau - \mu}{\sigma}\right)$$

Se compara el estadístico (del lado izquierdo) y se rechaza la hipótesis nula si la probabilidad

$$1 - \alpha$$

es mayor. En caso contrario se acepta la hipótesis.

Usualmente se suele ocupar el **p value** que es otro estadístico el cual nos dice que si obtenemos el **p value** y este valor es menor a la probabilidad al nivel establecido (si fuera al 5%) no se rechaza la hipótesis pero si es mayor se rechaza la hipótesis nula

### 3) Datos de consumo privado nacional y el ingreso disponible total

Tiempo	Consumo	Ingreso
2003	5526.04	5575.87
2004	6031.74	6145.48
2005	6562.65	6674.08
2006	7150.74	7424.31
2007	7726.77	7958.09
2008	8268.75	8609.55
2009	8006.57	8959.36
2010	8734.84	9684.60
2011	9504.28	10803.50
2012	10237.97	11333.43
2013	10819.25	11798.73
2014	11509.61	12698.24
2015	12162.87	13752.80
2016	13188.67	14636.40
2017	14301.25	15486.60
2018	15215.29	16301.38

#### Modelo lineal asociado

El modelo asociado es:

$$y_t = \alpha_0 + \alpha_1 x_t + u_i$$

$y_t$ =consumo privado nacional

$x_t$ =ingreso nacional disponible

#### Calculando las medidas requeridas

La cantidad de datos de consumo e ingreso son:

```
tam_consumo=length(consumo)
tam_consumo
```

```
## [1] 16
```

```
tam_ingreso=length(ingreso)
tam_ingreso
```

```
## [1] 16
```

La media poblacional cuya fórmula es:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

La media poblacional de consumo e ingreso son:

```
prom_consumo=mean(consumo)
prom_consumo
```

```
## [1] 9684.206
```

```
prom_ingreso=mean(ingreso)
prom_ingreso
```

```
## [1] 10490.15
```

La varianza poblacional cuya fórmula es:

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

La varianza poblacional de consumo e ingreso son:

```
varianza_consumo=sum((consumo-prom_consumo)^2)/tam_consumo
varianza_consumo
```

```
## [1] 8227420
```

```
varianza_ingreso=sum((ingreso-prom_ingreso)^2)/tam_ingreso
varianza_ingreso
```

```
## [1] 10868551
```

La covarianza poblacional entre dos variables es:

$$S_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

La covarianza poblacional entre consumo e ingreso es

```
covarianza_con_ing=sum((consumo-prom_consumo)*(ingreso-prom_ingreso))/tam_ingreso
# tam_ingreso=tam_consumo
covarianza_con_ing
```

```
## [1] 9418046
```

## Estimando modelo lineal por MCO

```
mco_con_ingre <- lm(consumo ~ ingreso)
mco_con_ingre
```

```
##
## Call:
## lm(formula = consumo ~ ingreso)
##
## Coefficients:
## (Intercept)      ingreso
##    594.0590      0.8665
```

## Parametros obtenidos y modelo estimado

Por lo tanto el modelo lineal queda como:

$$\hat{y}_t = 594.059 + 0.8665x_t$$

Es decir tiene una pendiente positiva de 0.8665 que es el valor promedio cuando  $x_t = 0$

## Valores ajustados y residuales

A manera de resumen se tiene que

Reales_consumo	Ajustados_consumo	Residuales
5526.04	5425.550	100.489645
6031.74	5919.117	112.622580
6562.65	6377.149	185.500680
7150.74	7027.224	123.516385
7726.77	7489.744	237.026015
8268.75	8054.234	214.515925
8006.57	8357.344	-350.774440
8734.84	8985.765	-250.924900
9504.28	9955.292	-451.011750
10237.97	10414.476	-176.506095
10819.25	10817.659	1.591455
11509.61	11597.084	-87.473960
12162.87	12510.860	-347.990200
13188.67	13276.500	-87.829600
14301.25	14013.198	288.052100
15215.29	14719.205	496.085230



#### 4) Grafica de dispersion y linea de tendencia de los datos de consumo privado con ingreso disponible

El modelo lineal queda como:

$$\hat{consumo}_t = 594.059 + 0.8665(ingreso_t)$$

