

# Ajuste de curvas

## TAREA 7

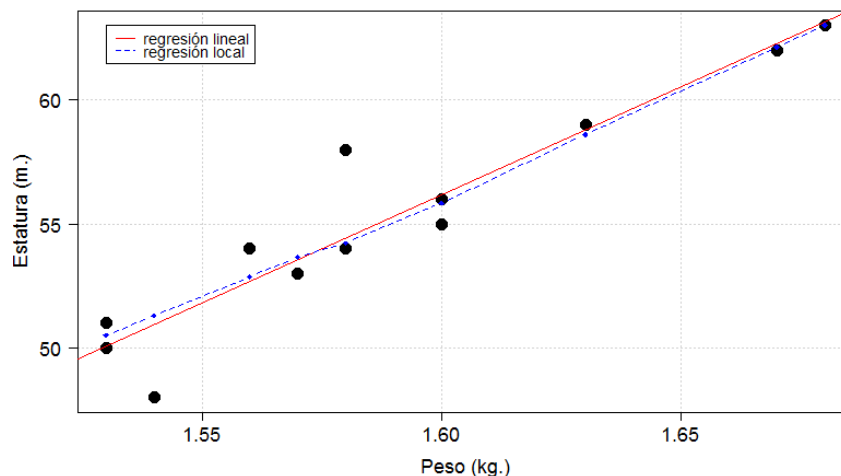
Alumno:  
Joaquín Arturo Velarde Moreno

### 1. Introducción

El objetivo del siguiente reporte es describir cómo se obtiene la correlación lineal a través de varios ejemplos de funciones matemáticas; ver su representación gráfica y la forma de su curva. Para cumplir con este objetivo, usaremos el programa R 4.0.2 [2] y de este modo, haremos cálculos con una serie de datos. Usaremos como apoyo el material de la Dra. Elisa Schaefer [1].

### 2. Definición correlación lineal

La correlación es una medida de la presencia de una relación lineal de un conjunto de datos que provienen de dos variables medidas al mismo tiempo sobre una serie de individuos, también se le conoce como datos bivariados. Estos datos se pueden representar por medio de una gráfica, por ejemplo, en la (Figura 2.1) se muestra gráficamente los datos de una encuesta que se hizo a alumnas de una escuela para ver la relación entre su peso y estatura.



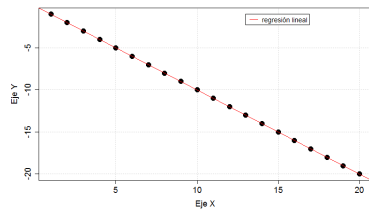
**Figura 2.1:** Relación entre peso(kg.) y altura(m.) de alumnas.

En el presente reporte usaremos la correlación de Pearson, la cual denotaremos por  $r$ ; el rango de la correlación  $r$  va de  $1$  a  $-1$ , lo cual puede generar tres casos (Figura 4.1):

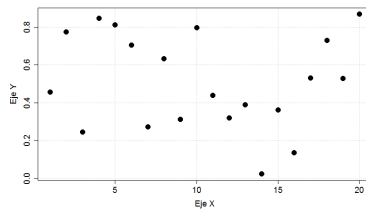
- $r$  igual a  $1$ : positiva,

- $r$  igual a  $-1$ : negativa,
- $r$  igual a  $0$ : sin correlación.

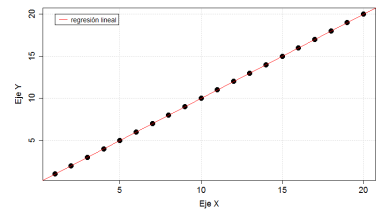
La mayoría de las veces no obtendremos mediciones exactas para cada caso, sino serán aproximadas y de estas, podremos decidir si existe una relación entre los datos o si son independientes entre sí.



(a) Correlación negativa  $Y$  disminuye cuando  $X$  crece.



(b) No hay relación alguna en los datos.



(c) Correlación positiva  $Y$  aumenta cuando  $X$  crece.

**Figura 2.2:** Correlaciones en sus valores negativo, positivo y 0.

### 3. Fórmula

La correlación de Pearson puede ser obtenida por la siguiente expresión:

$$\frac{\Sigma(x \times y) - \frac{1}{n}(\Sigma x \times \Sigma y)}{\sqrt{\left(\Sigma x^2 - \frac{1}{2}(\Sigma x)^2\right) \times \left(\Sigma y^2 - \frac{1}{2}(\Sigma y)^2\right)}}$$

donde:

- $x$  y  $y$  son nuestros datos bivariados,
- $n$  es el número de datos.

Esta ecuación puede ser expresada en R de la siguiente manera:

```
n          <- length(VectorX)
SumatoriaX <- sum(VectorX)
SumatoriaY <- sum(VectorY)

numerador  <- sum(VectorX * VectorY) - (SumatoriaX * SumatoriaY) / n
denominadorX <- sum(VectorX **2) - (SumatoriaX**2) / n
denominadorY <- sum(VectorY **2) - (SumatoriaY**2) / n
denominador <- sqrt(denominadorX * denominadorY)
correlacion <- numerador / denominador
```

Existe una manera más sencilla de hacer este cálculo obteniendo los promedios de cada vector, a los cuales denotaremos como  $\bar{X}$  y  $\bar{Y}$ , pudiéndose expresar de la siguiente manera:

$$\frac{\Sigma(x \times y)}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$$

donde:

- $y = Y - \bar{Y}$ ,
- $x = X - \bar{X}$ .

Esta ecuación puede ser expresada en R de la siguiente forma:

```

x      <- VectorX - mean(VectorX)
y      <- VectorY - mean(VectorY)

correlacion <- sum(x * y) / sqrt(sum(x**2) * sum(y**2))

```

## 4. Transformadas

No en todos los casos los conjuntos de datos aparecen de manera lineal; algunos se presentan muy dispersos y otros forman curvas, como es el caso del ejemplo de la función  $f(x) = x^2$  (Figura 4.3). En estos casos lo mejor es trabajar con transformadas, que son manipulaciones que se hacen a un vector de los datos para poder acomodarlos en una relación lineal. Una opción sencilla es la escalera de transformaciones de Tukey:

**Cuadro 1:** Escalera de Tukey con un rango de -2 a 2.

$\lambda$	-2	-1	$\frac{1}{2}$	0	$\frac{1}{2}$	1	2
	$\frac{1}{x^2}$	$\frac{1}{x}$	$\frac{1}{\sqrt{x}}$	$\log x$	$\sqrt{x}$	$x$	$x^2$

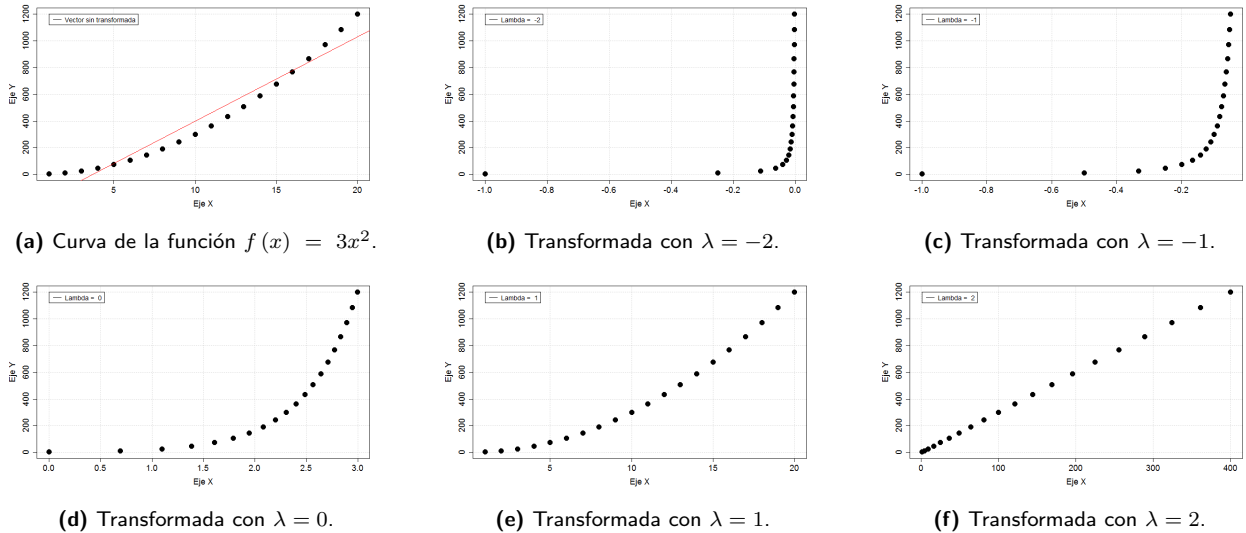
Podemos calcular y comparar cada transformada en R del siguiente modo:

```

for(i in seq(-2,2,1))
{
  if(i > 0)
  { z <- x^i}
  else if(i < 0)
  { z <- -1*x^i}
  else
  { z <- log(x)}
  cor(y,z)
}

```

Dada la función  $f(x) = 3x^2$  podemos buscar una transformada por la escalera de Tukey que nos encuentre el  $\lambda$  que mejore una relación lineal en nuestros conjuntos.



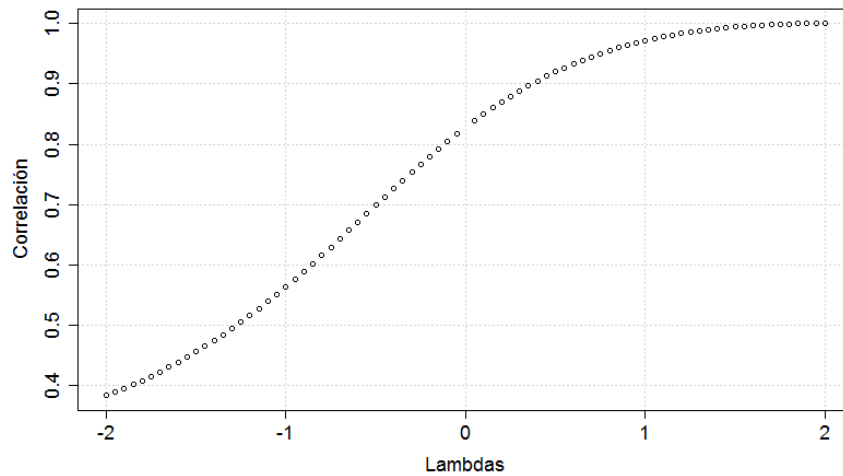
**Figura 4.1:** Transformadas de la función  $f(x) = 3x^2$ .

Una alternativa optimizada es la transformada Box-Scott, en esta expresión  $X$  se transforma a  $\frac{(x^\lambda - 1)}{\lambda}$ . Podemos calcular y comparar cada transformada en R de la siguiente manera:

```

for(i in seq(-2,2,0.05))
{
  z <- (abs(x)^i-1)/i
  w <- c(w, cor(y,z))
  v <- c(v,i)
}
plot(v, w)

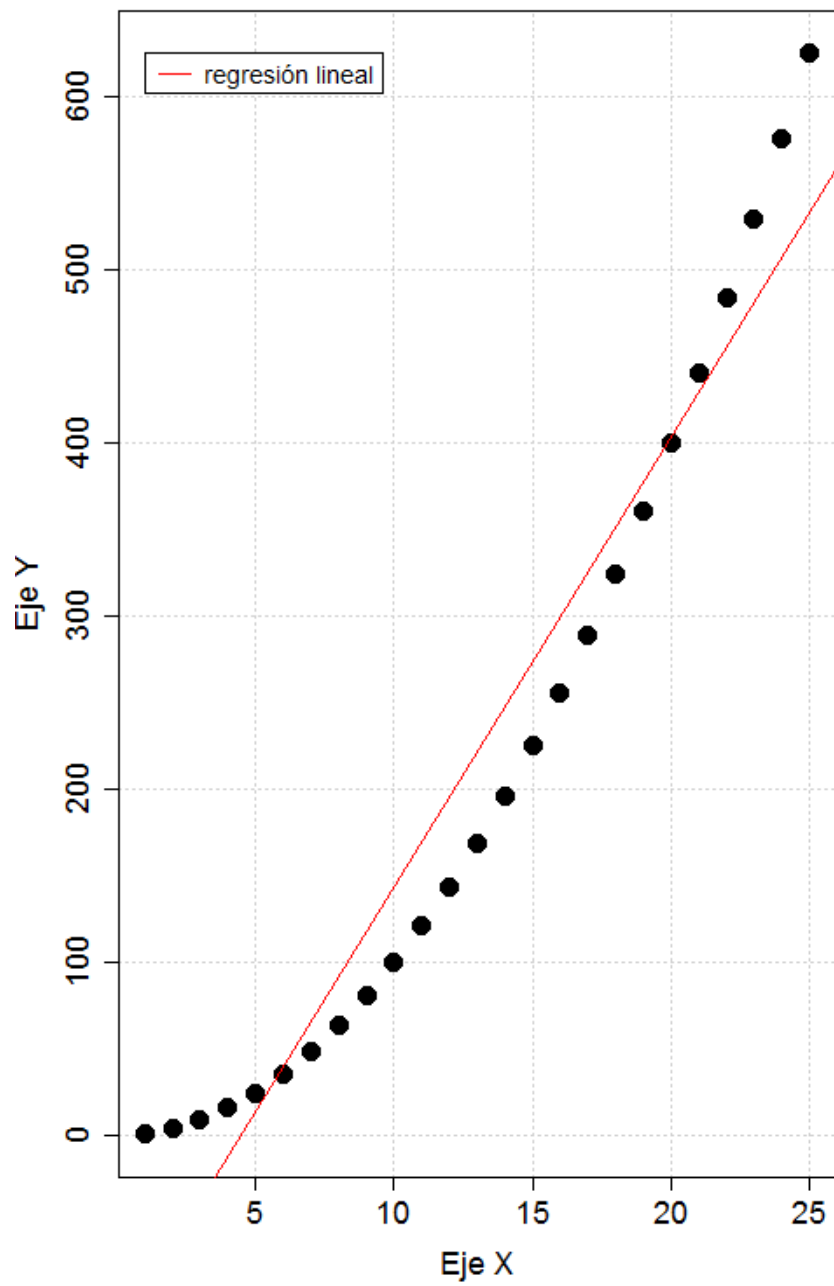
```



**Figura 4.2:** Lambdas y su correlación usando la transformada de Box-Scott en la función  $f(x) = 3x^2$ .

## Referencias

- [1] Satu Elisa Schaeffer. *Modelos probabilistas aplicados*. Sitio en, <https://elisa.dyndns-web.com/teaching/prob/pisis/prob.html>.
- [2] The R Foundation. *The R Project for Statistical Computing*. <https://www.r-project.org/>. 2019.



**Figura 4.3:** Relación entre el conjunto X y el Y al aplicar  $f(x) = x^2$ .