

Regresión lineal

DIMAS

18/11/2020

Regresión Lineal Simple

El modelo a nivel poblacional supone $E[Y|X_i]$ están perfectamente alineadas. Si están perfectamente alineadas, se puede representar en una recta. Donde β_1 es la ordenada al origen y β_2 es la pendiente.

$$E[Y|X_i] = \beta_1 + \beta_2 X_i$$

Si se desea tener los valores reales de Y_i , entonces se tiene la siguiente expresión

$$Y_i = \beta_1 + \beta_2 X_i + \epsilon_i$$

Donde ϵ_i es el componente de error aleatorio y se supone que $\epsilon_i \sim N(0, \sigma^2)$

Entonces

$$Y_i = E[Y|X_i] + \epsilon_i$$

En el caso de tener una muestra, la expresión es la siguiente

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Si se desea saber a Y_i , entonces

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\epsilon}_i$$

Por lo que

$$Y_i = \hat{Y}_i + \hat{\epsilon}_i$$

El término de error estimado $\hat{\epsilon}_i$ se puede calcular como

$$\hat{\epsilon}_i = Y_i - \hat{Y}_i$$

Se desea encontrar las $\hat{\beta}_1$ y $\hat{\beta}_2$ que hacen que $\sum \hat{\epsilon}_i^2$ sea mínima

$$\hat{\beta}_2 = \frac{\sum n X_i Y_i - \sum X_i \sum Y_i}{n \sum (X_i^2) - (\sum X_i)^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Ejemplo

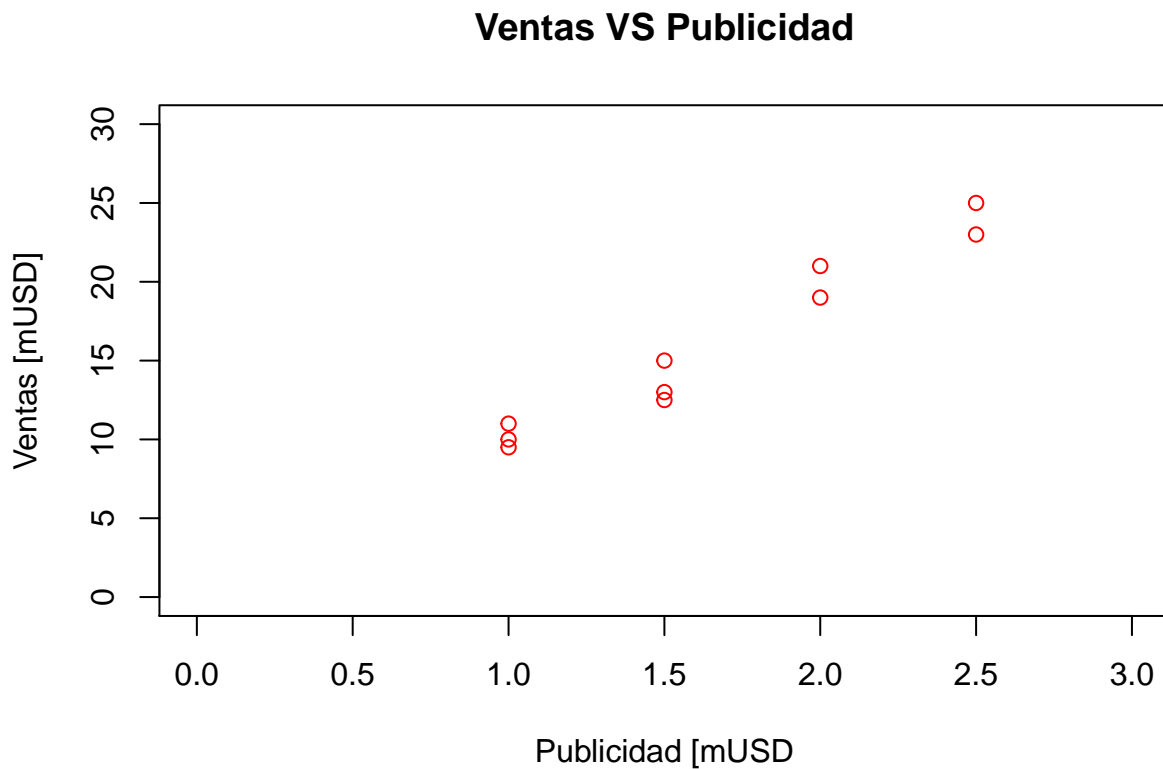
Para la siguiente muestra de *ventas VS publicidad* , encuentre la mejor recta que lo represente.

```
v <- c(10, 11, 9.5, 15, 13, 12.5, 21, 19, 25, 23)      #Ventas en Miles de Unidades (EJE Y) Podríamos
p <- c(1, 1, 1, 1.5, 1.5, 1.5, 2, 2, 2.5, 2.5) #Publicidad en Miles de Unidades (EJE X )

tabla_1 <- data.frame(Publicidad=p,Ventas=v )
tabla_1
```

```
##      Publicidad Ventas
## 1          1.0    10.0
## 2          1.0    11.0
## 3          1.0     9.5
## 4          1.5    15.0
## 5          1.5    13.0
## 6          1.5    12.5
## 7          2.0    21.0
## 8          2.0    19.0
## 9          2.5    25.0
## 10         2.5    23.0
```

```
plot(v~p, col="Red", main="Ventas VS Publicidad", xlab="Publicidad [mUSD]", ylab="Ventas [mUSD]", xlim=c
```



Calculo de $\hat{\beta}_1$ y $\hat{\beta}_2$

```
n <- length(v)

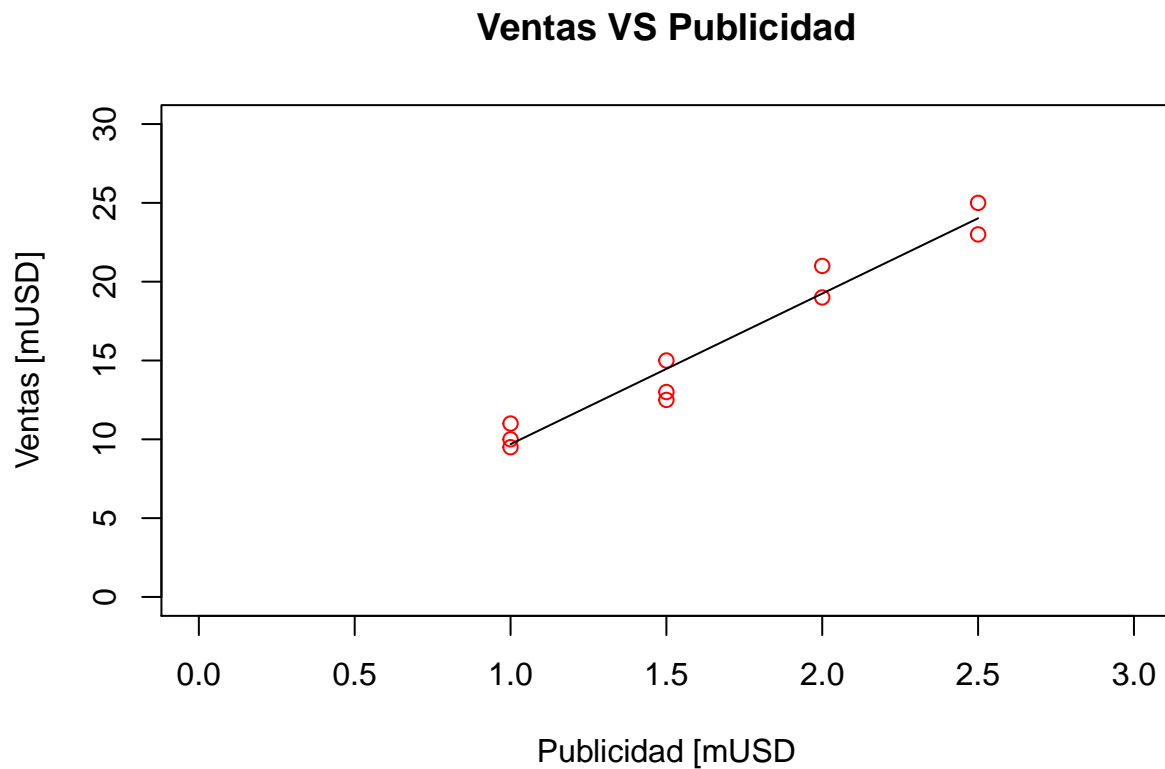
b_2 <- (n*sum(p*v)-sum(p)*sum(v))/(n*sum(p^2)-sum(p)^2)
b_1 <- mean(v)-b_2*mean(p)
```

La respuesta es:

$\hat{\beta}_1 = 0.14$ y $\hat{\beta}_2 = 9.55$

```
x_i <- seq(1, 2.5, by=.5)
y_i <- b_1 + (b_2*x_i)

plot(v~p, col="Red", main="Ventas VS Publicidad", xlab="Publicidad [mUSD]", ylab="Ventas [mUSD]", xlim=c(1, 2.5))
lines(x_i, y_i)
```



Solución usando el comando *lm*

```
m_1 <- lm(Ventas~Publicidad, data= tabla_1 ) #Reconoce los encabezados de la tabla

#lm(tabla_1$Ventas ~ tabla_1$Publicidad )
#lm(v ~ p )
#lm(y~x)

m_1 #En donde intercept es la pendientes b1 y publicidad b2
```

```
##
```

```
## Call:
## lm(formula = Ventas ~ Publicidad, data = tabla_1)
##
## Coefficients:
## (Intercept)    Publicidad
##      0.1364      9.5537
```

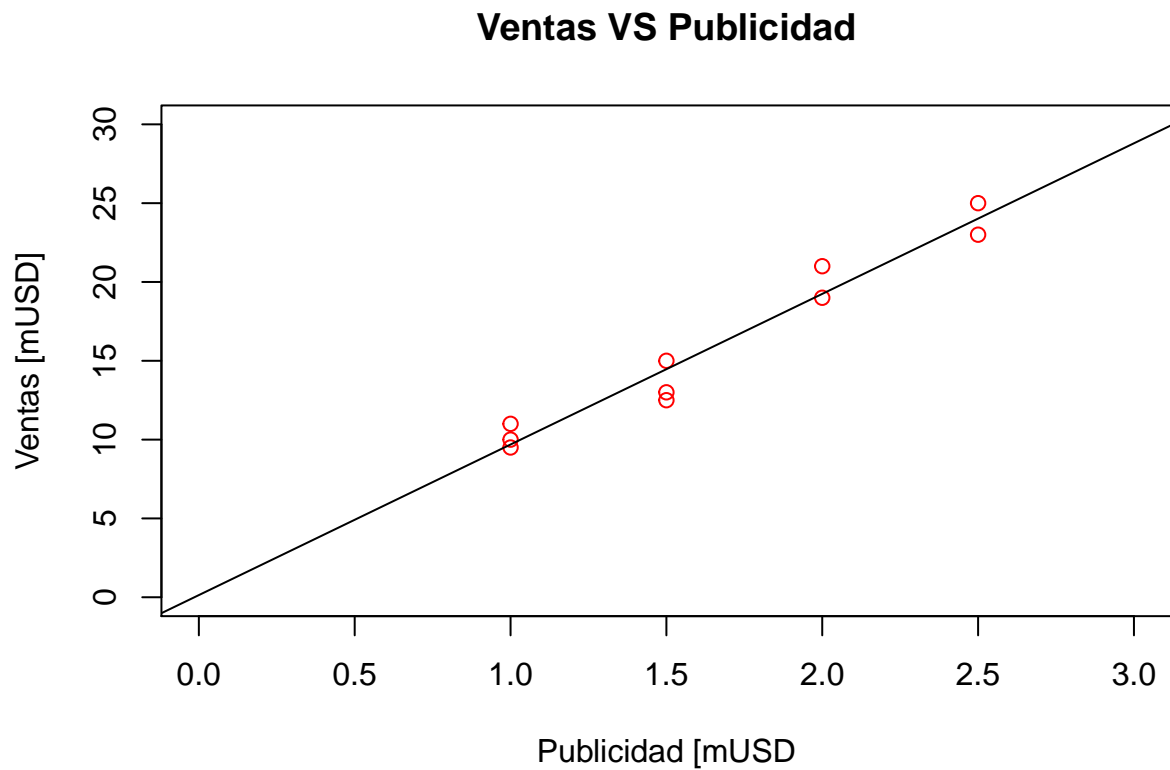
```
#m_1$residuals           son los errores
#m_1$fitted.values       son valores ajustados
```

```
tabla_2 <- data.frame(Publicidad=p, Ventas= v, Ventas_estimadas = m_1$fitted.values, Errores_estimadas=
```

```
tabla_2
```

```
##      Publicidad Ventas Ventas_estimadas Errores_estimadas
## 1          1.0   10.0         9.690083         0.3099174
## 2          1.0   11.0         9.690083         1.3099174
## 3          1.0    9.5         9.690083        -0.1900826
## 4          1.5   15.0        14.466942         0.5330579
## 5          1.5   13.0        14.466942        -1.4669421
## 6          1.5   12.5        14.466942        -1.9669421
## 7          2.0   21.0        19.243802         1.7561983
## 8          2.0   19.0        19.243802        -0.2438017
## 9          2.5   25.0        24.020661         0.9793388
## 10         2.5   23.0        24.020661        -1.0206612
```

```
plot(v ~ p, col="Red", main="Ventas VS Publicidad", xlab="Publicidad [mUSD]", ylab="Ventas [mUSD]", xlim=
abline(m_1)
```



Ejemplo Práctico

Importar datos a R

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 3.6.3
```

```
Base <- read_excel("Base de datos encuestas.xlsx")
#View(Base)
```

Análisis de los datos

```
str(Base) #structure
```

```
## tibble [366 x 38] (S3: tbl_df/tbl/data.frame)
##  $ MARCA TEMPORAL                : chr [1:366] "2020/03/02 12
##  $ NOMBRE DE USUARIO              : chr [1:366] "JULSUAREZ19@G
##  $ ¿CUAL ES TU EDAD?              : num [1:366] 20 20 21 20 20
##  $ ¿CUAL ES TU SEXO BIOLOGICO?    : chr [1:366] "HOMBRE" "MUJE
##  $ ¿QUE CARRERA ESTUDIAS?         : chr [1:366] "MECANICA" "ME
##  $ ¿TE GUSTA TU CARRERA?          : chr [1:366] "SI" "SI" "SI"
##  $ ¿EN DONDE HICISTE TU BACHILLERATO? : chr [1:366] "CCH" "PREPA U
```

```

## $ ¿CUAL FUE TU PROMEDIO DE BACHILLERATO? : num [1:366] 9.2 9 9.9 7.7 8
## $ ¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA? : num [1:366] 8.7 8.9 9.2 8.5
## $ ¿CUAL ES TU AVANCE DE CREDITOS? : chr [1:366] "40% - 50%" "50% - 60%" "60% - 70%" "70% - 80%" "80% - 90%" "90% - 100%"
## $ ¿CUAL ES TU NIVEL DE INGLES? : chr [1:366] "20% - 30%" "30% - 40%" "40% - 50%" "50% - 60%" "60% - 70%" "70% - 80%" "80% - 90%" "90% - 100%"
## $ ¿CUANTAS MATERIAS INSCRIBES EN PROMEDIO AL SEMESTRE? : num [1:366] 5 6 6 7 5 5 5 5 5 5
## $ ¿CUANTAS MATERIAS HAS REPROBADO? : num [1:366] 0 0 1 2 3 3 3 3 3 3
## $ ¿CUANTAS HORAS AL DIA, PASAS EN PROMEDIO EN LA FACULTAD? : num [1:366] 5 8 8 8 9 10 9 9 9 9
## $ ¿REALIZAS ALGUNA ACTIVIDAD EXTRACURRICULAR? : chr [1:366] "SI" "NO" "SI" "NO" "SI" "NO" "SI" "NO" "SI" "NO"
## $ ¿CUANTAS HORAS EN PROMEDIO, LE DEDICAS A ESA ACTIVIDAD POR SEMANA? : num [1:366] 15 0 6 8 2 1 8 8 8 8
## $ ¿CUANTAS HORAS AL DIA TE TOMA TRANSPORTARTE A LA ESCUELA? : num [1:366] 2.5 2 4 1.5 1.5 1.5 1.5 1.5 1.5 1.5
## $ ¿QUE TIPO DE TRANSPORTE UTILIZAS PARA IR A LA ESCUELA? : chr [1:366] "TRANSPORTE PUBLICO" "TRANSPORTE PRIVADO" "TRANSPORTE COMUNITARIO" "TRANSPORTE INDIVIDUAL" "TRANSPORTE COMUNITARIO" "TRANSPORTE INDIVIDUAL"
## $ ¿ERES FORANEO? : chr [1:366] "NO" "NO" "NO" "NO" "NO" "NO" "NO" "NO" "NO" "NO"
## $ ¿CON QUIEN VIVES? : chr [1:366] "AMBOS PADRES" "PADRE" "MADRE" "OTRO" "AMBOS PADRES" "PADRE" "MADRE" "OTRO" "AMBOS PADRES" "PADRE" "MADRE" "OTRO"
## $ ¿CUAL ES EL GRADO MAXIMO DE ESTUDIOS DE TU MADRE? : chr [1:366] "BACHILLERATO" "SECUNDARIA" "PRIMARIA" "OTRO" "BACHILLERATO" "SECUNDARIA" "PRIMARIA" "OTRO" "BACHILLERATO" "SECUNDARIA" "PRIMARIA" "OTRO"
## $ ¿CUAL ES EL GRADO MAXIMO DE ESTUDIOS DE TU PADRE? : chr [1:366] "BACHILLERATO" "SECUNDARIA" "PRIMARIA" "OTRO" "BACHILLERATO" "SECUNDARIA" "PRIMARIA" "OTRO" "BACHILLERATO" "SECUNDARIA" "PRIMARIA" "OTRO"
## $ ¿CUANTAS PERSONAS HABITAN CONTIGO? : num [1:366] 7 3 4 5 5 5 4 3 3 3
## $ ¿CUANTAS CON HABITACION PROPIA? : chr [1:366] "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI"
## $ ¿ERES RESPONSABLE DE ALGUNA MASCOTA? : chr [1:366] "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI"
## $ ¿CUANTAS CON ALGUNA BECA? : chr [1:366] "NO" "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI"
## $ ¿ESTUDIAS Y TRABAJAS? : chr [1:366] "SI" "NO" "NO" "NO" "SI" "NO" "NO" "NO" "SI" "NO" "NO"
## $ ¿CUANTAS HORAS A LA SEMANA TRABAJAS? : num [1:366] 15 0 0 0 4 2 0 0 0 0
## $ ¿CUAL ES EL INGRESO PROMEDIO MENSUAL DE TU FAMILIA? : chr [1:366] "$20,001 - $50,000" "$10,001 - $20,000" "$5,001 - $10,000" "$1,001 - $5,000" "$0 - $1,000" "$20,001 - $50,000" "$10,001 - $20,000" "$5,001 - $10,000" "$1,001 - $5,000" "$0 - $1,000"
## $ 27. ¿CUANTO GASTAS SEMANALMENTE EN COSAS RELACIONADAS CON LA ESCUELA? : num [1:366] 500 400 200 100 50 0 0 0 0 0
## $ 28. ¿TE ENCUENTRAS EN UNA RELACION CON ALGUNA PERSONA? : chr [1:366] "NO" "SI" "NO" "SI" "NO" "SI" "NO" "SI" "NO" "SI"
## $ 29. ¿TIENES HIJOS? : chr [1:366] "NO" "NO" "NO" "NO" "NO" "NO" "NO" "NO" "NO" "NO"
## $ 30. ¿CUANTOS HIJOS TIENES? : num [1:366] NA NA NA NA NA NA NA NA NA NA
## $ 31. ¿TE HACES CARGO AL 100% DE TUS HIJOS? : chr [1:366] NA NA NA NA NA NA NA NA NA NA
## $ 32. ¿CUANTAS HORAS DIARIAS DUERMES EN PROMEDIO? : num [1:366] 6 5 6 6 6 5 6 6 6 6
## $ 33. ¿CUANTAS CON LAPTOP PROPIA? : chr [1:366] "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI" "SI"
## $ ¿CUANTAS VECES AL MES CONSUMES ALCOHOL? : chr [1:366] "6" "1" "1" "4" "6" "1" "1" "4" "6" "1" "1" "4"
## $ ¿QUE TIPO DE BEBIDA ES LA QUE MAS CONSUMES? : chr [1:366] "VODKA, TEQUILA" "VODKA, TEQUILA" "VODKA, TEQUILA" "VODKA, TEQUILA" "VODKA, TEQUILA" "VODKA, TEQUILA" "VODKA, TEQUILA" "VODKA, TEQUILA" "VODKA, TEQUILA" "VODKA, TEQUILA"

```

```
names(Base) #headers de "Base"
```

```

## [1] "MARCA TEMPORAL"
## [2] "NOMBRE DE USUARIO"
## [3] "¿CUAL ES TU EDAD?"
## [4] "¿CUAL ES TU SEXO BIOLOGICO?"
## [5] "¿QUE CARRERA ESTUDIAS?"
## [6] "¿TE GUSTA TU CARRERA?"
## [7] "¿EN DONDE HICISTE TU BACHILLERATO?"
## [8] "¿CUAL FUE TU PROMEDIO DE BACHILLERATO?"
## [9] "¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA?"
## [10] "¿CUAL ES TU AVANCE DE CREDITOS?"
## [11] "¿CUAL ES TU NIVEL DE INGLES?"
## [12] "¿CUANTAS MATERIAS INSCRIBES EN PROMEDIO AL SEMESTRE?"
## [13] "¿CUANTAS MATERIAS HAS REPROBADO?"
## [14] "¿CUANTAS HORAS AL DIA, PASAS EN PROMEDIO EN LA FACULTAD?"
## [15] "¿REALIZAS ALGUNA ACTIVIDAD EXTRACURRICULAR?"
## [16] "¿CUANTAS HORAS EN PROMEDIO, LE DEDICAS A ESA ACTIVIDAD POR SEMANA?"
## [17] "¿CUANTAS HORAS AL DIA TE TOMA TRANSPORTARTE A LA ESCUELA?"
## [18] "¿QUE TIPO DE TRANSPORTE UTILIZAS PARA IR A LA ESCUELA?"
## [19] "¿ERES FORANEO?"
## [20] "¿CON QUIEN VIVES?"

```

```
## [21] "¿CUAL ES EL GRADO MAXIMO DE ESTUDIOS DE TU MADRE?"
## [22] "¿CUAL ES EL GRADO MAXIMO DE ESTUDIOS DE TU PADRE?"
## [23] "¿CUANTAS PERSONAS HABITAN CONTIGO?"
## [24] "¿CUENTAS CON HABITACION PROPIA?"
## [25] "¿ERES RESPONSABLE DE ALGUNA MASCOTA?"
## [26] "¿CUENTAS CON ALGUNA BECA?"
## [27] "¿ESTUDIAS Y TRABAJAS?"
## [28] "¿CUANTAS HORAS A LA SEMANA TRABAJAS?"
## [29] "¿CUAL ES EL INGRESO PROMEDIO MENSUAL DE TU FAMILIA?"
## [30] "27.\t¿CUANTO GASTAS SEMANALMENTE EN COSAS RELACIONADAS CON LA ESCUELA?"
## [31] "28.\t¿TE ENCUENTRAS EN UNA RELACION CON ALGUNA PERSONA?"
## [32] "29.\t¿TIENES HIJOS?"
## [33] "30.\t¿CUANTOS HIJOS TIENES?"
## [34] "31.\t¿TE HACES CARGO AL 100% DE TUS HIJOS?"
## [35] "32.\t¿CUANTAS HORAS DIARIAS DUERMES EN PROMEDIO?"
## [36] "33.\t¿CUENTAS CON LAPTOP PROPIA?"
## [37] "¿CUANTAS VECES AL MES CONSUMES ALCOHOL?"
## [38] "¿QUE TIPO DE BEBIDA ES LA QUE MAS CONSUMES?"
```

`summary(Base)`

```
## MARCA TEMPORAL      NOMBRE DE USUARIO  ¿CUAL ES TU EDAD?
## Length:366          Length:366          Min.    :17.00
## Class :character     Class :character    1st Qu.:20.00
## Mode  :character     Mode  :character    Median  :21.00
##                                     Mean    :21.04
##                                     3rd Qu.:22.00
##                                     Max.    :32.00
##
## ¿CUAL ES TU SEXO BIOLOGICO? ¿QUE CARRERA ESTUDIAS? ¿TE GUSTA TU CARRERA?
## Length:366           Length:366           Length:366
## Class :character     Class :character    Class :character
## Mode  :character     Mode  :character    Mode  :character
##
##
##
## ¿EN DONDE HICISTE TU BACHILLERATO? ¿CUAL FUE TU PROMEDIO DE BACHILLERATO?
## Length:366           Min.    : 7.00
## Class :character     1st Qu.: 8.40
## Mode  :character     Median  : 8.80
##                                     Mean    : 8.77
##                                     3rd Qu.: 9.20
##                                     Max.    :10.00
##
## ¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA? ¿CUAL ES TU AVANCE DE CREDITOS?
## Min.    :6.400          Length:366
## 1st Qu.:8.062          Class :character
## Median :8.505          Mode  :character
## Mean    :8.505
## 3rd Qu.:8.960
## Max.    :9.950
##
## ¿CUAL ES TU NIVEL DE INGLES?
```

```

## Length:366
## Class :character
## Mode :character
##
##
##
##
## ¿CUANTAS MATERIAS INSCRIBES EN PROMEDIO AL SEMESTRE?
## Min. : 4.000
## 1st Qu.: 5.000
## Median : 6.000
## Mean : 6.038
## 3rd Qu.: 6.000
## Max. :50.000
##
## ¿CUANTAS MATERIAS HAS REPROBADO?
## Min. : 0.000
## 1st Qu.: 0.000
## Median : 1.000
## Mean : 2.309
## 3rd Qu.: 3.750
## Max. :15.000
##
## ¿CUANTAS HORAS AL DIA, PASAS EN PROMEDIO EN LA FACULTAD?
## Min. : 4.000
## 1st Qu.: 7.000
## Median : 8.000
## Mean : 8.481
## 3rd Qu.:10.000
## Max. :45.000
##
## ¿REALIZAS ALGUNA ACTIVIDAD EXTRACURRICULAR?
## Length:366
## Class :character
## Mode :character
##
##
##
##
## ¿CUANTAS HORAS EN PROMEDIO, LE DEDICAS A ESA ACTIVIDAD POR SEMANA?
## Min. : 0.000
## 1st Qu.: 0.000
## Median : 1.500
## Mean : 3.822
## 3rd Qu.: 6.000
## Max. :42.000
##
## ¿CUANTAS HORAS AL DIA TE TOMA TRANSPORTARTE A LA ESCUELA?
## Min. : 0.100
## 1st Qu.: 1.000
## Median : 2.000
## Mean : 2.371
## 3rd Qu.: 3.000
## Max. :45.000

```



```

##
## ¿QUE TIPO DE TRANSPORTE UTILIZAS PARA IR A LA ESCUELA? ¿ERES FORANEO?
## Length:366 Length:366
## Class :character Class :character
## Mode :character Mode :character
##
##
##
##
## ¿CON QUIEN VIVES? ¿CUAL ES EL GRADO MAXIMO DE ESTUDIOS DE TU MADRE?
## Length:366 Length:366
## Class :character Class :character
## Mode :character Mode :character
##
##
##
##
## ¿CUAL ES EL GRADO MAXIMO DE ESTUDIOS DE TU PADRE?
## Length:366
## Class :character
## Mode :character
##
##
##
##
## ¿CUANTAS PERSONAS HABITAN CONTIGO? ¿CUENTAS CON HABITACION PROPIA?
## Min. : 0.00 Length:366
## 1st Qu.: 3.00 Class :character
## Median : 4.00 Mode :character
## Mean : 3.88
## 3rd Qu.: 4.00
## Max. :47.00
##
##
## ¿ERES RESPONSABLE DE ALGUNA MASCOTA? ¿CUENTAS CON ALGUNA BECA?
## Length:366 Length:366
## Class :character Class :character
## Mode :character Mode :character
##
##
##
##
## ¿ESTUDIAS Y TRABAJAS? ¿CUANTAS HORAS A LA SEMANA TRABAJAS?
## Length:366 Min. : 0.000
## Class :character 1st Qu.: 0.000
## Mode :character Median : 0.000
## Mean : 6.433
## 3rd Qu.: 8.000
## Max. :72.000
## NA's :134
##
## ¿CUAL ES EL INGRESO PROMEDIO MENSUAL DE TU FAMILIA?
## Length:366
## Class :character
## Mode :character
##

```

```

##
##
##
## 27.\t¿CUANTO GASTAS SEMANALMENTE EN COSAS RELACIONADAS CON LA ESCUELA?
## Min. : 20.0
## 1st Qu.: 200.0
## Median : 300.0
## Mean : 393.7
## 3rd Qu.: 500.0
## Max. :1500.0
##
## 28.\t¿TE ENCUENTRAS EN UNA RELACION CON ALGUNA PERSONA? 29.\t¿TIENES HIJOS?
## Length:366 Length:366
## Class :character Class :character
## Mode :character Mode :character
##
##
##
##
## 30.\t¿CUANTOS HIJOS TIENES? 31.\t¿TE HACES CARGO AL 100% DE TUS HIJOS?
## Min. :0.00000 Length:366
## 1st Qu.:0.00000 Class :character
## Median :0.00000 Mode :character
## Mean :0.00571
## 3rd Qu.:0.00000
## Max. :1.00000
## NA's :191
## 32.\t¿CUANTAS HORAS DIARIAS DUERMES EN PROMEDIO?
## Min. : 2.000
## 1st Qu.: 5.000
## Median : 6.000
## Mean : 5.689
## 3rd Qu.: 6.000
## Max. :10.000
##
## 33.\t¿CUENTAS CON LAPTOP PROPIA? ¿CUANTAS VECES AL MES CONSUMES ALCOHOL?
## Length:366 Length:366
## Class :character Class :character
## Mode :character Mode :character
##
##
##
##
## ¿QUE TIPO DE BEBIDA ES LA QUE MAS CONSUMES?
## Length:366
## Class :character
## Mode :character
##
##
##
##

```

“Cuál fue tu promedio de bachillerato?”

```
p_b <- Base$`¿CUAL FUE TU PROMEDIO DE BACHILLERATO?`
```

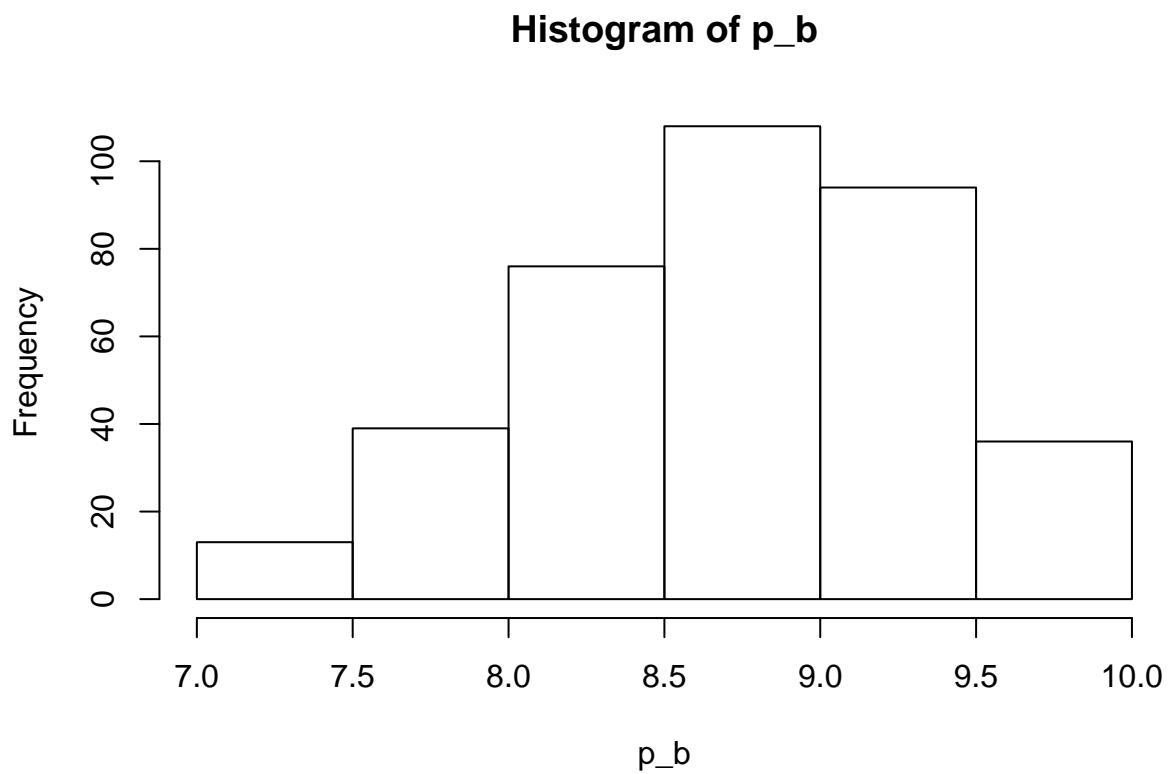
```
mean(p_b)
```

```
## [1] 8.77
```

```
sd(p_b)
```

```
## [1] 0.6209242
```

```
hist(p_b)
```



“Cuál fue tu promedio de licenciatura?”

```
p_l <- Base$`¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA?`
```

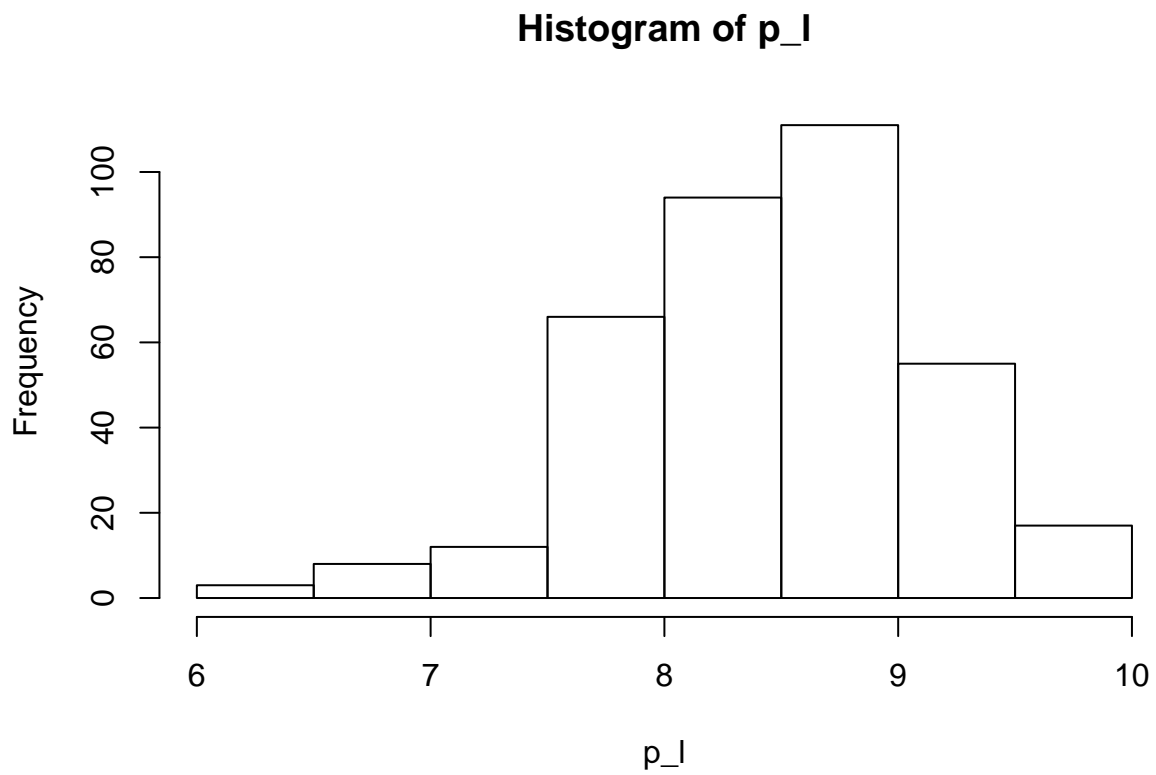
```
mean(p_l)
```

```
## [1] 8.505464
```

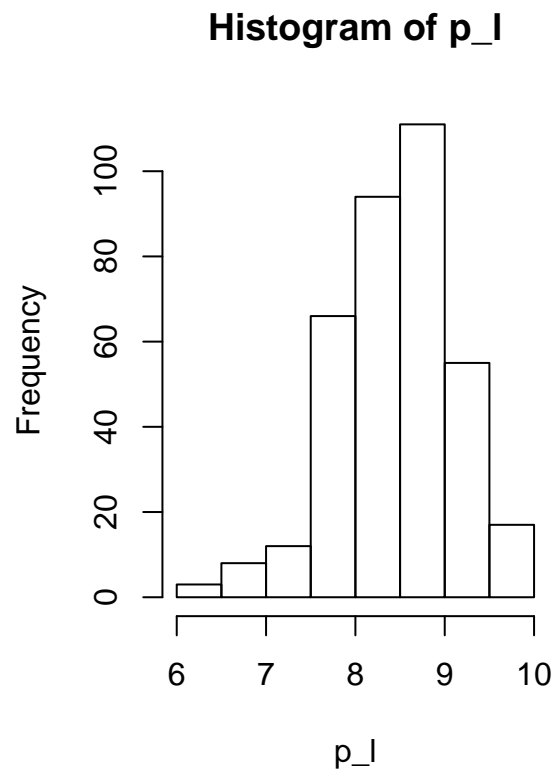
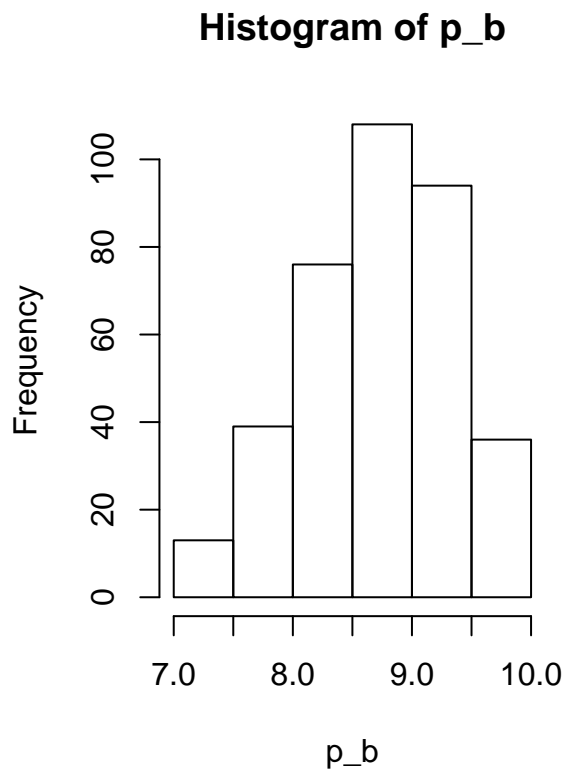
```
sd(p_l)
```

```
## [1] 0.637307
```

```
hist(p_l)
```



```
par(mfrow=c(1,2)) #para emparejar dos graficas en una
#mfrow o mfcoll para asignar el número de renglones, columnas
hist(p_b)
hist(p_l)
```



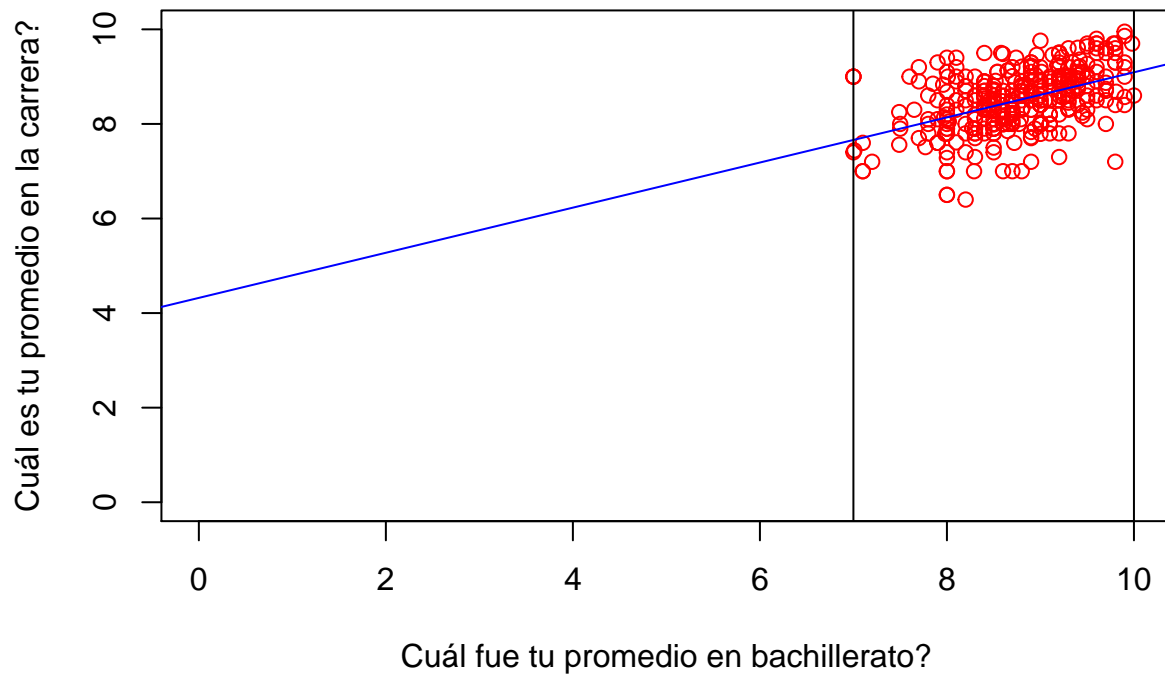
```
dev.off() # funcion para reiniciar
```

```
## null device
##      1
```

Regresión lineal

```
m_2 <- lm('¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA?' ~ '¿CUAL FUE TU PROMEDIO DE BACHILLERATO?', data=Base)
#m_2 <- lm(Base$'¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA?' ~ Base$'¿CUAL FUE TU PROMEDIO DE BACHILLERATO?', data=Base)
plot((Base$'¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA?' ~ Base$'¿CUAL FUE TU PROMEDIO DE BACHILLERATO?'),
      abline(m_2, col="Blue")
      abline(v=c(7,10))
```

Bachillerato VS Carrera



```
cor(Base$'¿CUAL FUE TU PROMEDIO DE BACHILLERATO?', Base$'¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA?')
```

```
## [1] 0.4648126
```

Resultados

```
m_2 #intercept es la B0 por la cruza el eje de las y cuando x=0
```

```
##
## Call:
## lm(formula = '¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA?' ~ '¿CUAL FUE TU PROMEDIO DE BACHILLERATO?',
##     data = Base)
##
## Coefficients:
##                (Intercept)
##                   4.3215
## '¿CUAL FUE TU PROMEDIO DE BACHILLERATO?'
##                   0.4771
```

```
#intercept es B0 y es el valor que minimiza los errores al cuadrado
```

```
summary(m_2)
```

```
##
```

```
## Call:
## lm(formula = '¿CUAL ES TU PROMEDIO ACTUAL EN LA CARRERA?' ~ '¿CUAL FUE TU PROMEDIO DE BACHILLERATO?',
## data = Base)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.83353 -0.37665 -0.00373  0.33595  1.33896
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   4.32150     0.41878   10.32   <2e-16
## '¿CUAL FUE TU PROMEDIO DE BACHILLERATO?' 0.47708     0.04763   10.02   <2e-16
##
## (Intercept)                  ***
## '¿CUAL FUE TU PROMEDIO DE BACHILLERATO?' ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5651 on 364 degrees of freedom
## Multiple R-squared:  0.2161, Adjusted R-squared:  0.2139
## F-statistic: 100.3 on 1 and 364 DF, p-value: < 2.2e-16
```

```
confint(m_2, level=0.95)
```

```
##                                2.5 %    97.5 %
## (Intercept)                   3.4979719 5.1450352
## '¿CUAL FUE TU PROMEDIO DE BACHILLERATO?' 0.3834071 0.5707459
```

$$Y_i = \beta_1 + \beta_2 X_i$$

$P_A = \text{Promedio Actual}$, $P_B = \text{Promedio Bachillerato}$

$$P_A = 4.3214 + 0.4771 P_B$$

#Esta sección ya podría considerarse como un pronóstico

```
P_B <- 9
P_A=4.3214 +0.4771*P_B
P_A
```

```
## [1] 8.6153
```

Covarianza

Correlación ~ “Desviación estándar”

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$-1 \leq \rho \leq 1$$

Es la fuerza de la relación lineal entre dos variables aleatorias. Es la fuerza de como covarian las dos variables aleatorias.

Fuerza de relación lineal: que tan bien se aproximan a una recta los valores de las dos variables.

Correlación no implica causalidad

```
x_p <- c(1,2,2,3,3,3,4,4,5,6,6)
y_p <- c(1.5, 1.5, 2.5, 2.5, 3, 3.5, 3.5, 4.5, 5.5, 6, 6.5)

r_xy <- cor(x_p, y_p)
r_yx <- cor(y_p, x_p)

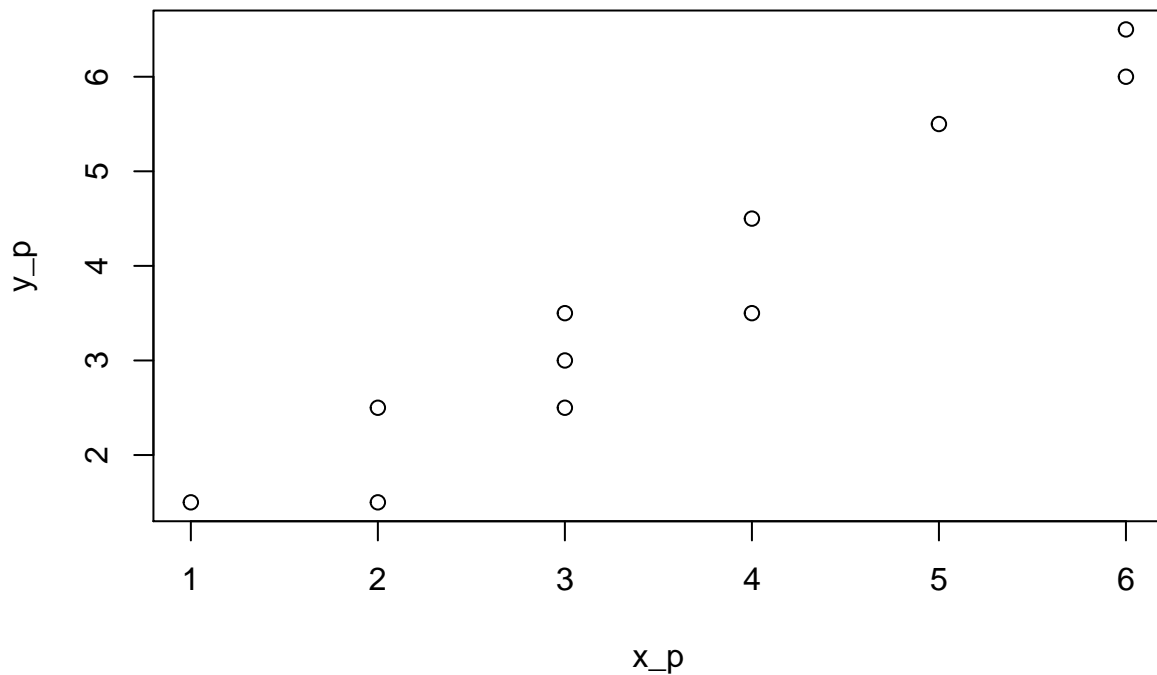
sd(x_p)
```

```
## [1] 1.634848
```

```
sd(y_p)
```

```
## [1] 1.735983
```

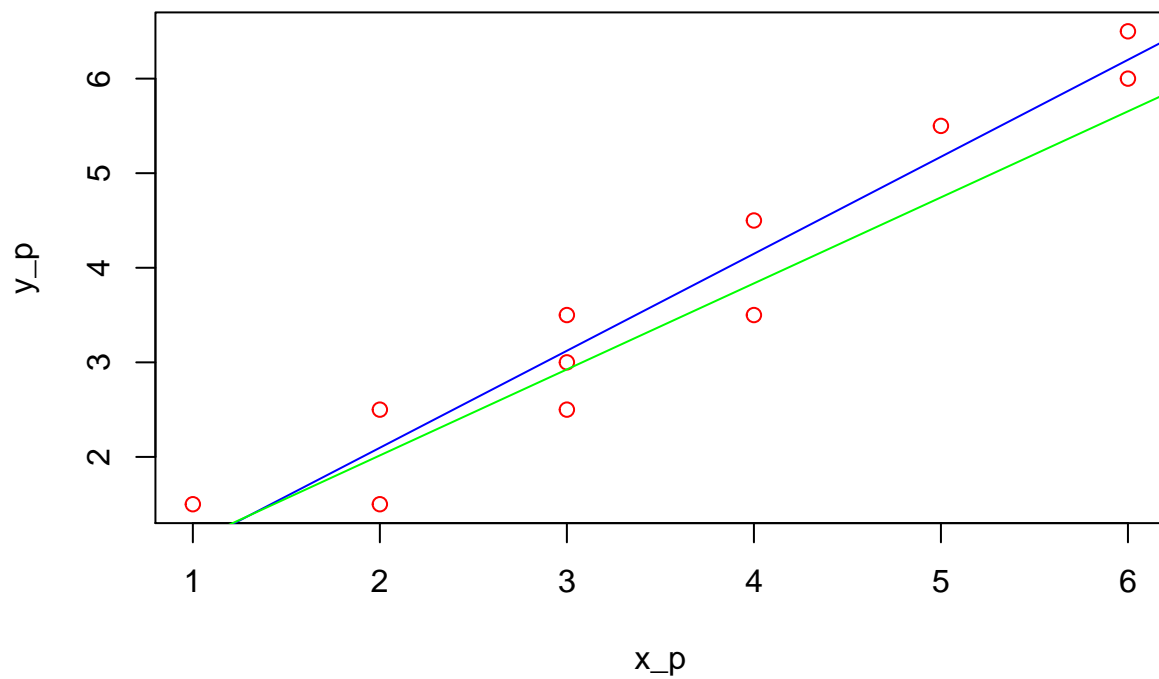
```
plot(y_p~x_p)
```




```
ml_yx <- lm(y_p~x_p)
ml_xy <- lm(x_p~y_p)
```

Un modelo lineal tampoco representa causalidad

```
plot(y_p~x_p, col="Red")
abline(ml_yx, col="Blue")
abline(ml_xy, col="Green")
```



La única manera de demostrar la causalidad es por conceptos y teorías externas.