

# Práctica dirigida 8

## Contents

I. Análisis de regresión lineal simple (RLS): ideas clave	1
II. Aplicación práctica	2
Practica resolviendo los siguientes ejercicios en casa	11
FACULTAD DE CIENCIAS SOCIALES - PUCP	

Curso: POL 278 - Estadística para el análisis político 1 | Semestre 2025 - 1

---

## I. Análisis de regresión lineal simple (RLS): ideas clave

Técnica estadística que predice el valor de una variable con los valores de otra. La regresión lineal simple es un método útil para predecir una respuesta cuantitativa **Y** partiendo de una sola variable predictora **X**, asumiendo que hay una relación aproximadamente lineal entre X e Y. Matemáticamente, esta relación lineal se representa como

$$Y = a + bX + E$$

- Y = variable dependiente o explicada. Variable cuyos valores se desea predecir o resumir. Un modelo de regresión lineal tiene como variable dependiente una variable numérica
- a = Constante: ordenada en el origen, valor esperado de “Y” cuando X=0
- b = Pendiente: mide el cambio de la variable “Y” por cada unidad de cambio de “X”. Su magnitud sirve para predecir en cuánto aumentará “y” cada vez que “x” se incremente en una unidad. Su signo puede ser positivo o negativo, y en esto la interpretación coincide con la correlación.
- X = variable utilizada para predecir el valor de la variable dependiente. También se denomina variable predictora o variable explicativa. Las variables explicativas que son parte del modelo suelen ser numéricas o intervalares; sin embargo, es posible incorporar variables explicativas ordinales o categóricas.
- E = Corresponde a las desviaciones de los valores verdaderos de Y con respecto a los valores esperados de “Y” (diferencia entre lo observado y estimado por el modelo). Asumimos que es independiente de “X”.

La relación entre las variables depende de la pendiente:

- Si b es positivo, Y aumenta cuando X aumenta. Es una relación directa / positiva.
- Si b es negativo, Y aumenta cuando X disminuye. Es una relación inversa / negativa.
- Si b es cero. Y no cambia cuando X varía. No existe relación entre las variables.

Asimismo, con el método de la regresión lineal se puede responder las siguientes preguntas:

- Analizar si hay una **asociación** entre las variables mediante un test de independencia estadística.
- Analizar la **dirección** de la asociación (directa o inversa).

- Evaluar la **fuerza** de la asociación usando una medida de asociación llamada *correlación de Pearson*.
- Estimar una ecuación de regresión que “**predice**” los valores de la variable dependiente para valores de la variable independiente.

## II. Aplicación práctica

Para la sesión de hoy trabajaremos con datos del 2017 de una base de datos que contiene variables obtenidas de las siguientes bases: - Freedom in the World - V-Dem - Democracy Index - Global Corruption Barometer

Información sobre las bases: - Freedom in the World es elaborado por Freedom House y analiza los siguientes aspectos: el proceso electoral, las políticas pluriculturales y la participación, el funcionamiento del gobierno, la libertad de expresión y de creencia, los derechos de asociación y organización, el estado de derecho, la autonomía personal y los derechos individuales.

- V-Dem es publicada por el V-Dem Institute. En ella se describe la calidad de los gobiernos a partir de información de 542 indicadores. La sata describe todos los aspectos de un gobierno, brindando énfasis en la calidad de la democracia, la inclusividad y otros indicadores económicos.
- Democracy Index es elaborado por The Economist y utiliza 60 indicadores, agrupados en cinco categorías: proceso electoral y pluralismo, libertades civiles, funcionamiento del gobierno, participación política y política cultural. A partir de estas categorías posiciona a los países en alguno de los cuatro tipos de régimen: Democracia plena, democracia imperfecta, régimen híbrido y régimen autoritario.
- Global Corruption Barometer es publicado por Transparencia Internacional y contiene información proveniente de la opinión pública ciudadana

```
library(rio)
data=import("BD_corrupcion.xlsx")
```

### ¿La percepción de corrupción en los funcionarios impacta en el puntaje del índice de la democracia?

Ejercicio 1: Impacto de la percepción de la corrupción de funcionarios en el puntaje de democracia participativa (según V-Dem), por país.

Las variables que utilizaremos serán:

**c\_funcionarios** = Porcentaje de personas que opinan que la mayoría o todos los funcionarios del gobierno están envueltos en corrupción

**Part** = Indicador de democracia participativa - Vdem

### Gráfico de dispersión:

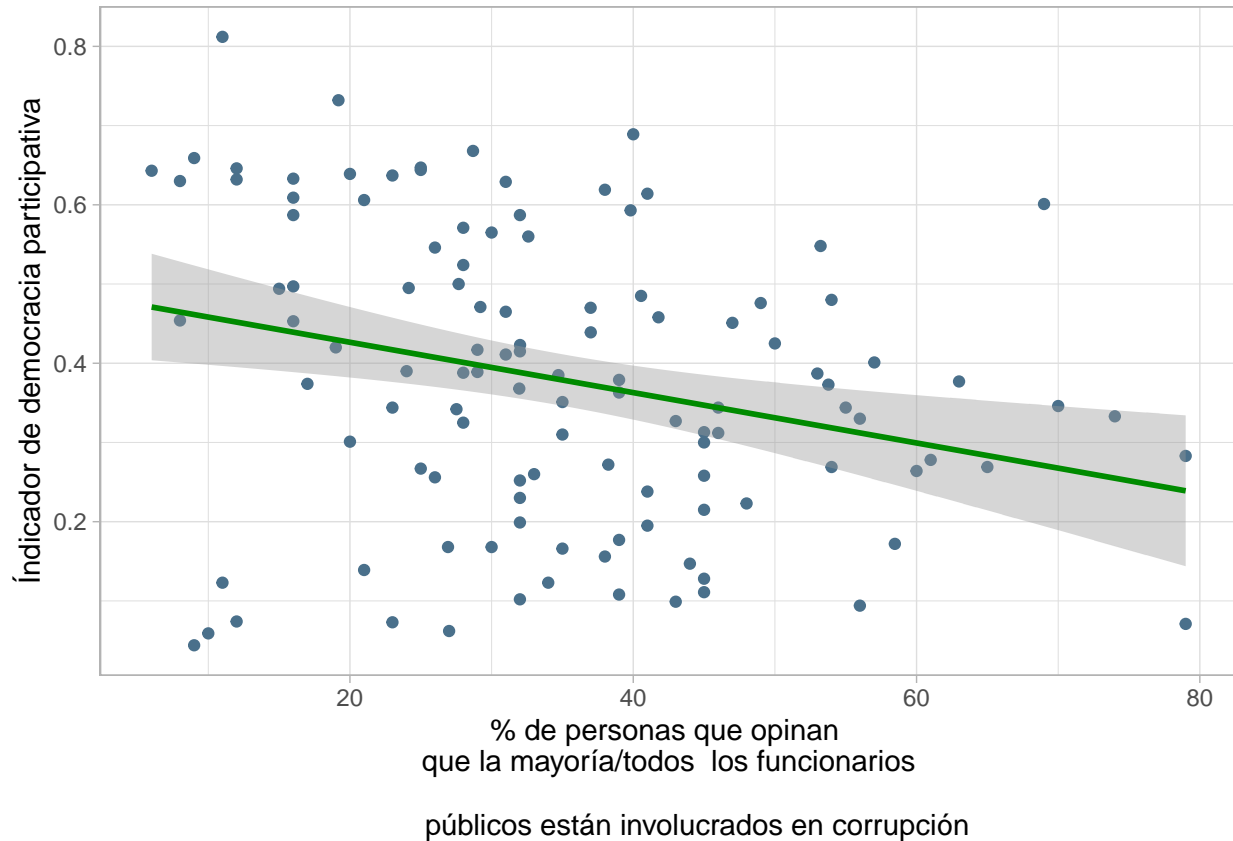
Elaboramos el gráfico de dispersión y la recta

```
library(dplyr)
library(ggplot2)
data %>%
  ggplot(aes (x=c_funcionarios*100, y=Part)) +
  geom_point(colour="skyblue4") +
  xlab("% de personas que opinan \n que la mayoría/todos los funcionarios \n
    públicos están involucrados en corrupción") +
  ylab("Índice de democracia participativa")+ theme_light() +
  geom_smooth(method="lm", se = T, colour="green4")

## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



A primera vista parece haber un leve impacto, y la relación entre nuestras variables se perfila como negativa. Comprobémoslo ...

### Creación del modelo

```
modelo1=lm(Part~c_funcionarios,data=data)
summary(modelo1)
```

```
##
## Call:
## lm(formula = Part ~ c_funcionarios, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41744 -0.12712  0.00909  0.13936  0.35691
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.49004    0.03944   12.425 < 2e-16 ***
## c_funcionarios -0.31779    0.10258   -3.098  0.00245 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1768 on 114 degrees of freedom
## (2 observations deleted due to missingness)
## Multiple R-squared:  0.07765,    Adjusted R-squared:  0.06956
## F-statistic: 9.598 on 1 and 114 DF,  p-value: 0.002453
```

## Interpretación

*¿El porcentaje de personas que creen que los funcionarios son corruptos influyen en el indicador de democracia participativa?*

Para interpretar los resultados del modelo debemos tener presente lo siguiente:

**Primero: p-value** Sabremos si la variable independiente impacta en la dependiente al revisar la significancia del p valor.

Establezcamos nuestras hipótesis:

- H0: El modelo de regresión no es válido
- H1: El modelo de regresión es válido (variable X aporta al modelo)

Como el p valor es **0.002453**, entonces podemos afirmar que hay suficiente evidencia para rechazar la H0, por lo que concluimos que el modelo sí es válido como modelo de predicción. Es decir, podemos decir que hay evidencia estadística suficiente para afirmar que existe una relación significativa entre el índice de democracia participativa y la percepción de funcionarios corruptos por país.

En otras palabras, podemos decir que la percepción de funcionarios corruptos del país *sí influye* en el puntaje obtenido en el índice de democracia participativa.

**Segundo: pendiente/b** Explica cómo es el efecto de x en y. Para ello analizamos el valor del parámetro de la pendiente.

En este caso, al ser este valor **-0.31779**, concluimos que cada vez que el valor de la percepción de corrupción en el país aumenta en 1, el puntaje del índice de democracia participativa disminuye en 0.318. Es decir, tenemos una *relación inversa o negativa*.

**Tercero: R<sup>2</sup>** Analizar cuánto de la variabilidad de la variable dependiente (y) es explicada por la variable independiente (x), para ello revisamos el **R<sup>2</sup>** (Multiple R-squared). Los valores van de 0 a 1. Mientras más cercano esté el R<sup>2</sup> a 1, mayor será la variabilidad explicada. El R<sup>2</sup> es un indicador de ajuste del modelo.

En nuestro modelo, este arrojó el valor de **0.07765**, por lo que podemos concluir que aproximadamente el **7.8%** (0.07765\*100) de la variabilidad en el índice de democracia participativa puede ser explicado por el porcentaje de personas que opinan que la mayoría/todos los funcionarios están envueltos en corrupción.

Esto quiere decir que la cantidad de variabilidad explicada es relativamente baja. Por lo que podemos afirmar que hay otros factores más importantes o complejos que afectan la percepción de corrupción en cada país.

**Cuarto: Ecuación de la recta** Hallar la ecuación de la recta del modelo. Para lograrlo, revisemos los dos valores de la tabla que se encuentran en la columna de “Estimate”, el valor de la primera fila es el del intercepto (a) y el de la segunda es el de la pendiente (b).

Del segundo paso, ya conocíamos que el valor de la pendiente es **-0.31779**. Si volvemos a revisar nuestra tabla podemos observar que en el cruce de Estimate e Intercept está el valor de **0.49004**, este sería nuestro intercepto. Ahora, armemos nuestra ecuación de la recta:

$$\hat{Y} = 0.49004 - 0.31779 \cdot X$$

Donde:

- X = Porcentaje de personas que opinan que la mayoría o todos los funcionarios del gobierno están envueltos en corrupción - (independiente)
- Y = Indicador de democracia participativa - (dependiente)

**¿Qué se quiere saber?:** Queremos saber si el % de personas que opinan que los funcionarios son corruptos influye en el indicador de democracia participativa.

Esa ecuación crea una línea recta en el diagrama de dispersión que representa la relación entre ambas variables y además indica que el cambio esperado en nuestra variable dependiente (Porcentaje de personas que opinan que la mayoría o todos los funcionarios del gobierno están envueltos en corrupción) por cada cambio de una unidad en nuestra variable independiente (Indicador de democracia participativa). Así, con esta ecuación se puede estimar el valor de Y para cualquier valor de X.

También podemos obtener los coeficientes de intercepción/intercepto y pendiente de la siguiente forma:

```
modelo1$coefficients
```

```
##      (Intercept) c_funcionarios
##      0.4900430      -0.3177901
```

**Quinto: Predecir** Podemos utilizar la relación linear establecida por la ecuación para estimar el valor de la variable dependiente (Y) para un valor dado de la variable independiente (X).

Por ejemplo, si queremos calcular el valor de porcentaje de personas que opinan que la mayoría/todos los funcionarios están envueltos en corrupción cuando el índice de democracia participativa (0 a 1) es 0.4, solo tenemos que reemplazar el valor de x por 0.4 en nuestra ecuación.

Sustituyendo el valor de “x” en la ecuación, tenemos:

$$\hat{Y} = 0.49004 - 0.31779 * 0.4$$

$$\hat{Y} = 0.49004 - 0.12712$$

$$\hat{Y} = 0.36292$$

Por lo tanto, utilizando la ecuación de la recta, cuando el porcentaje de personas que cree que existen funcionarios corruptos es 40% (0.4) , podemos predecir que el índice de democracia será de, aproximadamente, 0.36 puntos.

Asimismo, se puede afirmar que, de acuerdo al modelo, por cada unidad de aumento del % de percepción de corrupción, el índice de democracia se reduce en 0.32 puntos.

Ejercicio 2: Impacto de la percepción de derechos políticos en el puntaje del índice de libertad (según Freedom House), por país

Las variables que utilizaremos serán:

**Der\_pol** = Puntaje del indicador de derechos políticos

**Libertad** = Puntuación de libertad - Freedom House

**¿Los derechos políticos impactan en el puntaje de libertad?**

**Gráfico de dispersión:**

Elaboramos el gráfico de dispersión y la recta

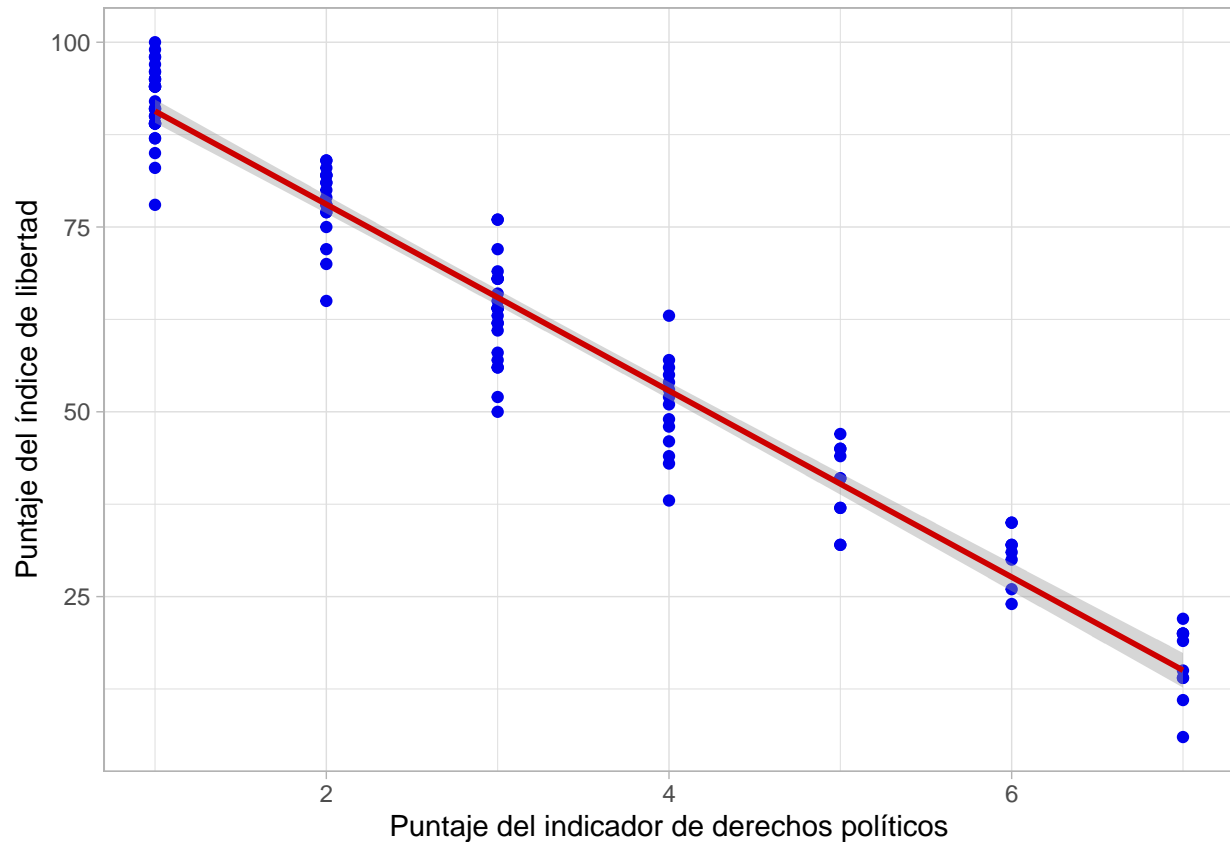
```
library(dplyr)
library(ggplot2)
data %>%
  ggplot(aes (x=Der_pol, y=Libertad)) +
  geom_point(colour="blue2") +
  xlab("Puntaje del indicador de derechos políticos") +
```

```
ylab("Puntaje del índice de libertad")+ theme_light() +
geom_smooth(method="lm", se = T, colour="red3")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 4 rows containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



A primera vista parece haber un impacto fuerte, y la relación entre nuestras variables se perfila como negativa o inversa.

### Creación del modelo

```
modelo2=lm(Libertad~Der_pol,data=data)
summary(modelo2)
```

```
##
## Call:
## lm(formula = Libertad ~ Der_pol, data = data)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-15.4645	-3.2478	0.3189	4.1001	10.5355

```
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 103.2894      1.0177  101.50  <2e-16 ***
## Der_pol     -12.6083      0.2749  -45.86  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.669 on 112 degrees of freedom
## (4 observations deleted due to missingness)
## Multiple R-squared:  0.9494, Adjusted R-squared:  0.949
## F-statistic: 2103 on 1 and 112 DF, p-value: < 2.2e-16
```

## Interpretación

**Primero: p-value** Sabremos si la variable independiente impacta en la dependiente al revisar la significancia del p valor.

Establezcamos nuestras hipótesis:

- H0: El modelo de regresión no es válido
- H1: El modelo de regresión es válido (variable X aporta al modelo)

Como el p valor es **2.2e-16**, entonces podemos afirmar que hay suficiente evidencia para rechazar la H0, por lo que concluimos que el modelo sí es válido como modelo de predicción. Es decir, podemos decir que hay evidencia estadística suficiente para afirmar que existe una relación significativa entre el puntaje del indicador de derechos políticos y el puntaje del índice de Libertad, por país.

En otras palabras, podemos decir que la percepción de la existencia de derechos políticos *sí influye* en el puntaje obtenido en el índice de libertad.

**Segundo: pendiente/b** Explica cómo es el efecto de x en y. Para ello analizamos el valor del parámetro de la pendiente.

En este caso, al ser este valor **-12.6083**, concluimos que cada vez que el valor de la percepción de la existencia de derechos políticos en el país aumenta en 1, el puntaje del índice de libertad disminuye en 12.61 puntos. Es decir, tenemos una *relación inversa o negativa*.

**Tercero: R<sup>2</sup>** Analizar cuánto de la variabilidad de la variable dependiente (y) es explicada por la variable independiente (x), para ello revisamos el **R<sup>2</sup>** (Multiple R-squared). Los valores van de 0 a 1. Mientras más cercano esté el R<sup>2</sup> a 1, mayor será la variabilidad explicada. El R<sup>2</sup> es un indicador de ajuste del modelo.

En nuestro modelo, este arrojó el valor de **0.9494**, por lo que podemos concluir que aproximadamente el **95%** (0.9494\*100) de la variabilidad en el puntaje de percepción de derechos políticos respecto del índice de libertad.

Esto quiere decir que la cantidad de variabilidad explicada es muy alta. Por tanto, podemos afirmar que la percepción de la existencia de derechos políticos tiene el impacto mayoritario sobre el puntaje del índice de la libertad.

**Cuarto: Ecuación de la recta** Hallar la ecuación de la recta del modelo. Para lograrlo, revisemos los dos valores de la tabla que se encuentran en la columna de “Estimate”, el valor de la primera fila es el del intercepto (a) y el de la segunda es el de la pendiente (b).

Del segundo paso, ya conocíamos que el valor de la pendiente es **-12.6083**. Si volvemos a revisar nuestra tabla podemos observar que en el cruce de Estimate e Intercept está el valor de **103.2894**, este sería nuestro intercepto. Ahora, armemos nuestra ecuación de la recta:

$$\hat{Y} = 103.2894 - 12.6083 \cdot X$$

Donde:

- X = Puntaje del indicador de derechos políticos - (independiente)
- Y = Indicador de libertad, según Freedom House - (dependiente)

También podemos obtener los coeficientes de intercepción/intercepto y pendiente de la siguiente forma:

```
modelo2$coefficients
```

```
## (Intercept)      Der_pol  
##    103.28943    -12.60832
```

**Quinto: Predecir** Podemos utilizar la relación linear establecida por la ecuación para estimar el valor de la variable dependiente (Y) para un valor dado de la variable independiente (X).

Por ejemplo, si queremos calcular el valor de porcentaje de personas que opinan que la mayoría/todos los funcionarios están envueltos en corrupción cuando el índice de democracia participativa (0 a 1) es 0.4, solo tenemos que reemplazar el valor de x por 0.4 en nuestra ecuación.

Sustituyendo el valor de “x” en la ecuación, tenemos:

$$\hat{Y} = 103.2894 - 12.6083 * X * 6$$

$$\hat{Y} = 103.2894 - 75.6498$$

$$\hat{Y} = 27.6396$$

Por lo tanto, utilizando la ecuación de la recta, cuando el puntaje del indicador de derechos políticos en un país es 6 , podemos predecir que aproximadamente el puntaje del indicador de libertad será de 27.7 puntos en ese país.

Ejercicio 3: Impacto de la percepción de funcionamiento del gobierno en el puntaje del índice de democracia (según Democracy Index), por país

Las variables que utilizaremos serán:

**Fun\_gob** = Puntaje del indicador de funcionamiento del gobierno

**Democracia** = Puntuación del indice de democracia - Democracy Index

**¿La percepción del funcionamiento del gobierno influye en el puntaje del indice de la democracia (Democracy Index)?**

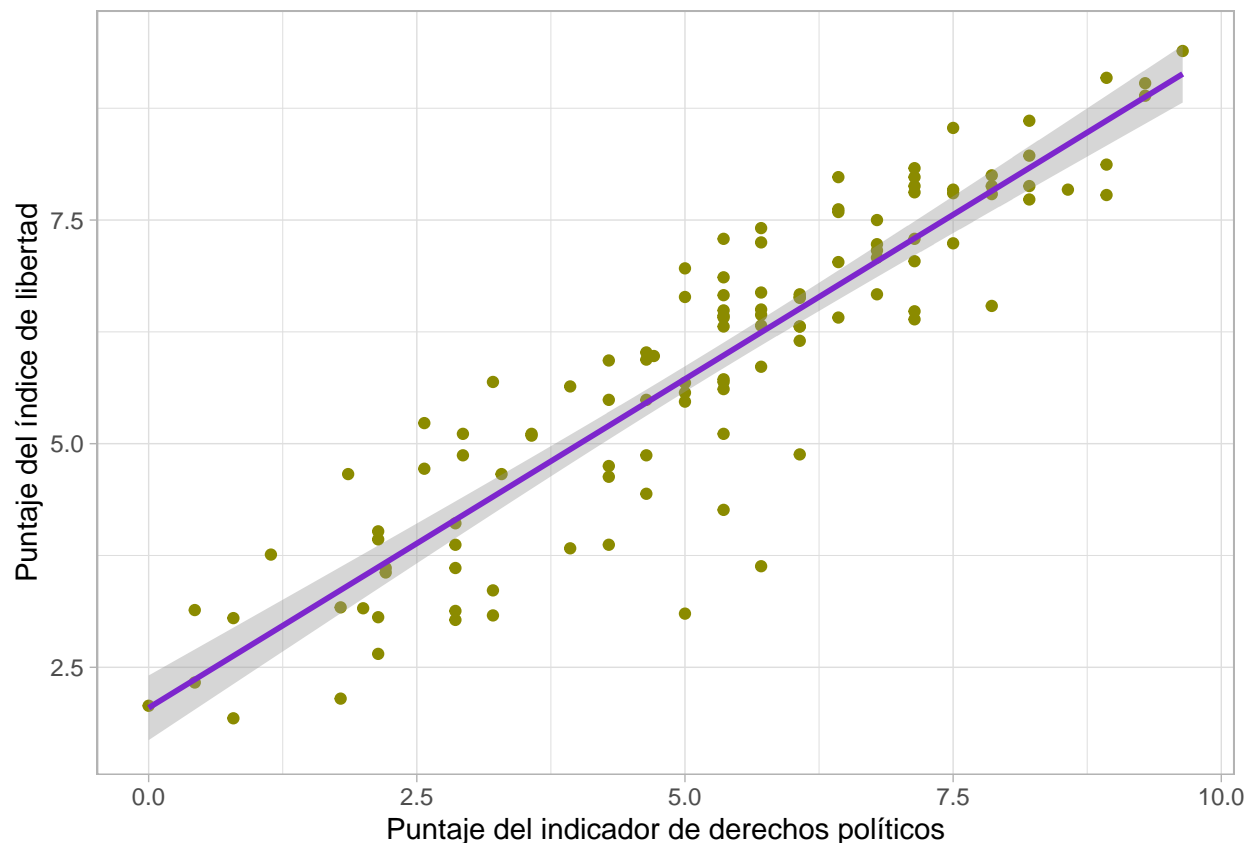
**Gráfico de dispersión:**

Elaboramos el gráfico de dispersión y la recta

```
library(dplyr)
library(ggplot2)
data %>%
  ggplot(aes (x=Fun_gob, y=Democracia)) +
  geom_point(colour="yellow4") +
  xlab("Puntaje del indicador de derechos políticos") +
  ylab("Puntaje del índice de libertad")+ theme_light() +
  geom_smooth(method="lm", se = T, colour="purple3")

## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 4 rows containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).
```





A primera vista parece haber un impacto moderado o mediano, y la relación entre nuestras variables se perfila como positiva o directa.

### Creación del modelo

```
modelo3=lm(Democracia~Fun_gob,data=data)
summary(modelo3)
```

```
##
## Call:
## lm(formula = Democracia ~ Fun_gob, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.62244 -0.36708  0.09834  0.48131  1.30293
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.0470     0.1825   11.22  <2e-16 ***
## Fun_gob       0.7351     0.0324   22.69  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7665 on 112 degrees of freedom
## (4 observations deleted due to missingness)
## Multiple R-squared:  0.8213, Adjusted R-squared:  0.8197
```

```
## F-statistic: 514.7 on 1 and 112 DF,  p-value: < 2.2e-16
```

## Interpretación

**Primero: p-value** Sabremos si la variable independiente impacta en la dependiente al revisar la significancia del p valor.

Establezcamos nuestras hipótesis:

- H0: El modelo de regresión no es válido
- H1: El modelo de regresión es válido (variable X aporta al modelo)

Como el p valor es **2.2e-16**, entonces podemos afirmar que hay suficiente evidencia para rechazar la H0, por lo que concluimos que el modelo sí es válido como modelo de predicción. Es decir, podemos decir que hay evidencia estadística suficiente para afirmar que existe una relación significativa entre el puntaje del indicador del funcionamiento del gobierno y el puntaje del índice de la democracia, por país.

En otras palabras, podemos decir que la percepción acerca del funcionamiento del gobierno *sí influye* en el puntaje obtenido en el índice de la democracia.

**Segundo: pendiente/b** Explica cómo es el efecto de x en y. Para ello analizamos el valor del parámetro de la pendiente.

En este caso, al ser este valor **0.7351**, concluimos que cada vez que el valor de la percepción de un buen funcionamiento del gobierno en el país aumenta en 1, el puntaje del índice de democracia aumenta en 0.74 puntos. Es decir, tenemos una *relación directa o positiva*.

**Tercero: R<sup>2</sup>** Analizar cuánto de la variabilidad de la variable dependiente (y) es explicada por la variable independiente (x), para ello revisamos el **R<sup>2</sup>** (Multiple R-squared). Los valores van de 0 a 1. Mientras más cercano esté el R2 a 1, mayor será la variabilidad explicada. El R2 es un indicador de ajuste del modelo.

En nuestro modelo, este arrojó el valor de **0.8213**, por lo que podemos concluir que aproximadamente el **82%** (0.8213\*100) del puntaje en el índice de la democracia es explicado por la percepción del funcionamiento del gobierno del país.

Esto quiere decir que la cantidad de variabilidad explicada es alta (sin llegar a ser muy alta). Por tanto, podemos afirmar que la percepción de un buen funcionamiento del gobierno tiene un impacto considerable sobre el puntaje del índice de la democracia.

**Cuarto: Ecuación de la recta** Hallar la ecuación de la recta del modelo. Para lograrlo, revisemos los dos valores de la tabla que se encuentran en la columna de "Estimate", el valor de la primera fila es el del intercepto (a) y el de la segunda es el de la pendiente (b).

Del segundo paso, ya conocíamos que el valor de la pendiente es **+0.7351**. Si volvemos a revisar nuestra tabla podemos observar que en el cruce de Estimate e Intercept está el valor de **2.0470**, este sería nuestro intercepto. Ahora, armemos nuestra ecuación de la recta:

$$\hat{Y} = 2.0470 + 0.7351 \cdot X$$

Donde:

- X = Puntaje del indicador de funcionamiento del gobierno - (independiente)
- Y = Indicador de democracia, según el Democracy Index - (dependiente)

También podemos obtener los coeficientes de intercepción/intercepto y pendiente de la siguiente forma:

```
modelo3$coefficients
```

```
## (Intercept)      Fun_gob  
##    2.0470146    0.7350849
```

**Quinto: Predecir** Sustituyendo el valor de “x” en la ecuación, tenemos:

$$\hat{Y} = 2.0470 + 0.7351 * X * 1.5$$

$$\hat{Y} = 2.0470 + 1.10265$$

$$\hat{Y} = 3.14965$$

Por lo tanto, utilizando la ecuación de la recta, cuando el puntaje del indicador de funcionamiento del gobierno en un país es 1.5, podemos predecir que aproximadamente el puntaje del indicador de democracia será de 3.15 puntos en ese país.

## **Practica resolviendo los siguientes ejercicios en casa**

Ejercicio 1: Impacto del % de personas que creen que los congresistas están envueltos en actos de corrupción (c\_congreso) en el índice de participación política de V-Dem (Part).

Ejercicio 2: Impacto de la percepción ciudadana acerca de los procesos electorales (Proceso\_elect) en el índice de la democracia de Democracy Index (Democracia).

Ejercicio 3: Impacto de la percepción sobre libertades civiles (Lib\_civiles) en el puntaje del índice de la libertad de Freedom House (Libertad).