

Clase 12 Regresión lineal

OCE 386 - Introducción al análisis de datos con R.

Dr. José A. Gallardo | Pontificia Universidad Católica de
Valparaíso

25 October 2021

PLAN DE LA CLASE

1.- Introducción

- ▶ Regresión lineal ¿Qué es y para qué sirve?
- ▶ Correlación v/s causalidad.
- ▶ Repaso ecuación de regresión lineal.
- ▶ Repaso Betas y causalidad.
- ▶ Interpretación Regresión lineal con R.

2.- Práctica con R y Rstudio cloud

- ▶ Realizar análisis de regresión lineal.
- ▶ Realizar gráficas avanzadas con ggplot2.
- ▶ Elaborar un reporte dinámico en formato pdf.

REGRESIÓN LINEAL

¿Qué es la regresión lineal y para que sirve?

Herramienta estadística que permite determinar si existe una relación (asociación) entre una variable predictora (independiente) y la variable respuesta (dependiente).

```
## # A tibble: 27 x 8
```

```
##      Year `CSIRO - Adjust~` `CSIRO - Lower ~` `CSIRO - Upper ~`
```

```
##      <dbl>          <dbl>          <dbl>          <dbl>
```

```
##  1  1993          6.30          6.04          6.56
```

```
##  2  1994          6.51          6.25          6.77
```

```
##  3  1995          6.62          6.36          6.88
```

```
##  4  1996          6.78          6.52          7.04
```

```
##  5  1997          7.06          6.80          7.32
```

```
##  6  1998          6.67          6.40          6.94
```

```
##  7  1999          7.00          6.74          7.26
```

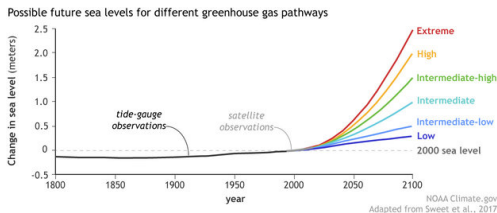
```
##  8  2000          7.06          6.79          7.33
```

```
##  9  2001          7.27          7.01          7.53
```

```
## 10  2002          7.37          7.10          7.64
```

REGRESIÓN LINEAL: PREDICCIÓN.

La ecuación de la regresión permite, bajo ciertos supuestos, predecir el valor de una variable respuesta “y” a partir de una o más variable predictoras “x”.



REGRESIÓN LINEAL: BETAS

Betas miden la influencia sobre la variable

LINEA DE REGRESIÓN.

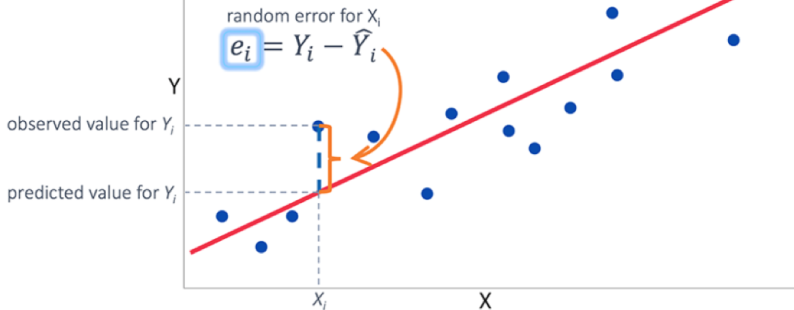
Línea de regresión: Corresponde a los valores “ajustados” o estimados de “y” en función de “x” y de los estimadores de mínimos cuadrados de

```
## function (save = "default", status = 0, runLast = TRUE)
## .Internal(quit(save, status, runLast))
## <bytecode: 0x7f8f85da2b68>
## <environment: namespace:base>
```

RESIDUOS Y MÉTODOS DE MÍNIMOS CUADRADOS

Method of Least Squares

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



COCIENTE DE DETERMINACIÓN

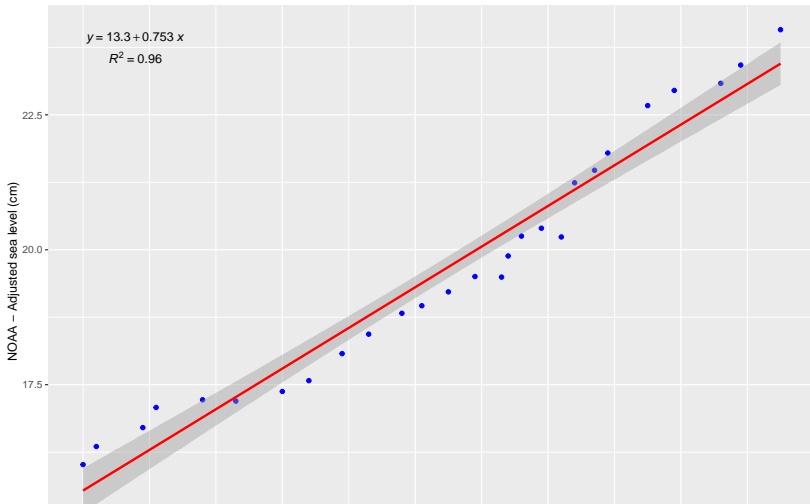
donde:

CORRELACIÓN NO IMPLICA CAUSALIDAD

¿El consumo de helados incrementa el nivel del mar?

¿Si dejamos de tomar helados disminuirá el nivel del mar?

¿Qué factor “z” puede explicar la correlación entre consumo de helados y nivel del mar?



PRUEBAS DE HIPÓTESIS

Prueba de hipótesis del coeficiente de regresión y el intercepto

La hipótesis nula en ambos casos es que el coeficiente del intercepto

Prueba de hipótesis del modelo completo

La hipótesis nula es si los coeficientes son iguales a 0.

REGRESIÓN LINEAL SIMPLE CON R

```
reg <- lm(`NOAA - Adjusted sea level (cm)` ~ Year, data = s  
summary(reg)
```

```
##
```

```
## Call:
```

```
## lm(formula = `NOAA - Adjusted sea level (cm)` ~ Year, da
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -0.84027 -0.18800 -0.07101  0.25306  0.65581
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -5.683e+02  1.650e+01  -34.45  <2e-16 ***
```

```
## Year          2.931e-01  8.223e-03   35.64  <2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
##
```

PREDICCIÓN LINEAL PROMEDIO

Predicción del nivel del mar promedio - próximos 3 años.

```
predict.lm(reg, newdata=data.frame(Year=c(2022,2023,2024)))
```

##		fit	lwr	upr
## 1	24.30088	23.99951	24.60224	
## 2	24.59394	24.27726	24.91062	
## 3	24.88700	24.55485	25.21915	

PREDICCIÓN LINEAL FUERA DEL RANGO OBSERVADO

Predicción del nivel del mar - próximos 3 años.

```
predict.lm(reg, newdata=data.frame(Year=c(5022,5023,5024)))
```

##		fit	lwr	upr
## 1	903.4872	852.4058	954.5687	
## 2	903.7803	852.6819	954.8787	
## 3	904.0734	852.9580	955.1887	

PRÁCTICA ANÁLISIS DE DATOS

- ▶ Guía de trabajo práctico disponible en drive y Rstudio.cloud.

Clase_12

- ▶ El trabajo práctico se realiza en Rstudio.cloud.

Guía 12 Regresión lineal

RESUMEN DE LA CLASE

- ▶ **Elaborar hipótesis para una regresión lineal**