

# Práctica Dirigida N°1 de Análisis de Regresión

Asencios Menacho Soledad

2024-09-11

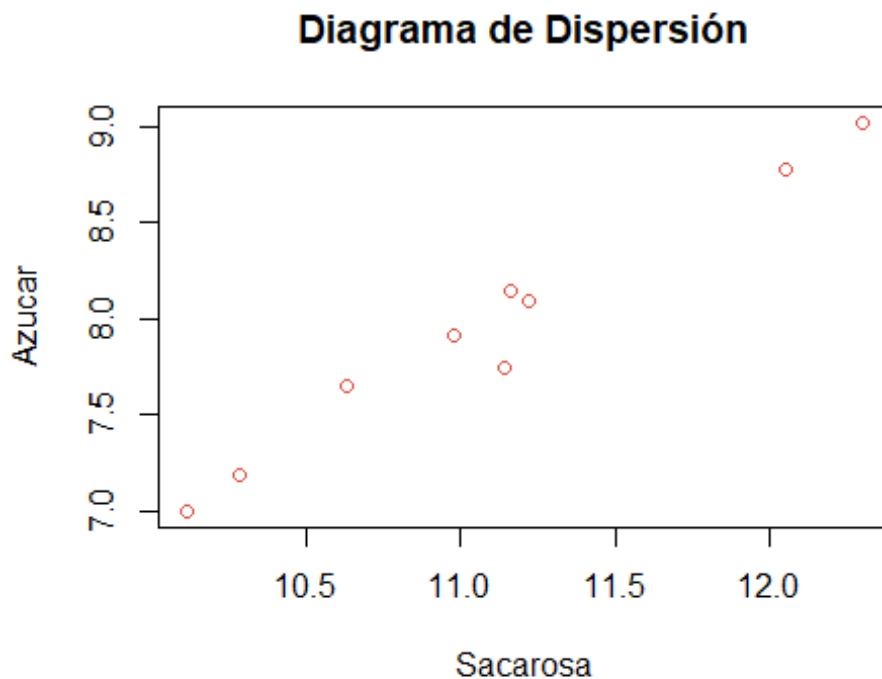
## CASO 02: Empresa Azucarera del Norte

La empresa azucarera del Norte llevo a cabo un estudio con la finalidad de analizar la influencia de algunos factores sobre el rendimiento de azúcar. Se desea determinar si existe relación entre el porcentaje de azúcar por caña (Y) con el porcentaje de sacarosa por caña (X). Los resultados obtenidos se presentan en la siguiente tabla:

### Análisis de Correlación

#### a) Elabore el gráfico de dispersión

```
plot(PD1$sacarosa, PD1$azucar,  
     main = "Diagrama de Dispersión",  
     col = "#e74c3c",  
     xlab = "Sacarosa",  
     ylab = "Azucar")
```



Se puede apreciar una relación directa entre el azúcar y la sacarosa.

**b) Estime el coeficiente de correlación de Pearson.**

```
cor(PD1$sacarosa, PD1$azucar, method = "p")
```

```
## [1] 0.9852942
```

Hay una correlación lineal directa positiva fuerte entre el azúcar y la sacarosa.

**Pruebe a un nivel de significación de 0.03 si el coeficiente de correlación es significativamente diferente de cero.**

```
cor.test(PD1$sacarosa, PD1$azucar,  
         method = "p",  
         alternative = "t")
```

```
##  
## Pearson's product-moment correlation  
##  
## data: PD1$sacarosa and PD1$azucar  
## t = 15.257, df = 7, p-value = 1.252e-06  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.9291982 0.9970143  
## sample estimates:  
## cor  
## 0.9852942
```

- 1) Planteo de hipótesis

$$H_0: \rho_0 = 0$$

$$H_1: \rho_1 \neq 0$$

- 2) Nivel de significancia

$$\alpha = 0.03$$

- 3) Estadística de prueba

$$T = \frac{\bar{x} - \mu}{S\sqrt{(n)}}$$

- 4) Determinación de la región crítica

```
n <- nrow(PD1)
```

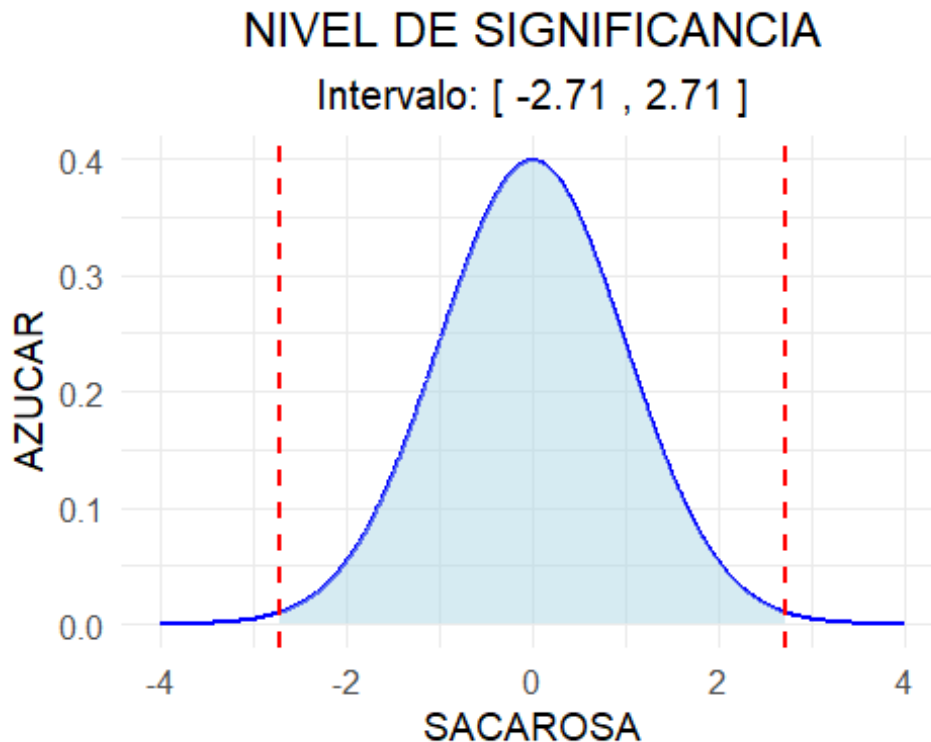
```
qt(0.03/2, n-2)
```

```
## [1] -2.714573
```

```
qt(c(0.03/2, 1-0.03/2), n-2)
```

```
## [1] -2.714573 2.714573
```

- 5) Gráfico



- 6) Interpretación valor p extremadamente bajo ( $1.2526 \times 10^{-6}$ ), se puede afirmar que existe un alto grado de confianza con una fuerte correlación positiva entre las variables sacarosa y azúcar.
- d) Pruebe a un nivel de significación de 0.03 si el coeficiente de correlación es menor a 0.8.

e)

d) Halle un intervalo del 97% de confianza para  $\rho$ .

```
cor.test(PD1$sacarosa, PD1$azucar,
         method = "p",
         alternative = "t",
         conf.level = 0.97)$conf

## [1] 0.9164998 0.9974844
## attr(,"conf.level")
## [1] 0.97
```

El intervalo 0.9164998 0.9974844 brinda una confianza de 97% de contener a la correlación entre el azúcar y la sacarosa.

### Análisis de Regresión Lineal Simple

f) Estime la ecuación de la recta e interprete sus coeficientes.

```
modelo <- lm(azucar ~ sacarosa, data = PD1)
modelo
```

```
##
## Call:
## lm(formula = azucar ~ sacarosa, data = PD1)
##
## Coefficients:
## (Intercept)      sacarosa
##      -1.9805         0.8948
```

$Y = -1.9805 + 0.8948X$   $b_0 = -1.9805$   $b_1 = 0.8948$  Al incrementar la sacarosa en 1%, se espera que el número medio de azúcar se incremente en 0.8948%

g) Realice el Análisis de Varianza. Use .

```
anova(modelo)

## Analysis of Variance Table
##
## Response: azucar
##           Df Sum Sq Mean Sq F value    Pr(>F)
## sacarosa   1 3.4028   3.4028   232.76 1.252e-06 ***
## Residuals  7 0.1023   0.0146
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1) Planteo de hipótesis

$$H_0: \beta_0 = 0$$

$$H_1: \beta_1 \neq 0$$

2) Nivel de significancia

$$\alpha = 0.03$$

3) Estadística de prueba

$$T = \frac{CMReg}{CME} \sim F_{1,n-2}$$

4) Determinación de la región crítica

```
n <- nrow(PD1)
n

## [1] 9

pf(0.03/2,1,n-2, lower.tail = F)

## [1] 0.9059653

pf(232.76,1,7, lower.tail = F)

## [1] 1.251841e-06
```

A un  $\alpha = 0.03$  se puede afirmar que el porcentaje del azúcar es explicado por el porcentaje de la sacarosa.

```
summary(modelo)
```

```
##
## Call:
## lm(formula = azucar ~ sacarosa, data = PD1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.237664 -0.028127 -0.005644  0.065506  0.144440
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.98053     0.65207  -3.037   0.0189 *
## sacarosa      0.89481     0.05865  15.257 1.25e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1209 on 7 degrees of freedom
## Multiple R-squared:  0.9708, Adjusted R-squared:  0.9666
## F-statistic: 232.8 on 1 and 7 DF,  p-value: 1.252e-06
```

h) Halle intervalos del 97% de confianza para  $\beta_0$  y  $\beta_1$ .

```
confint(modelo, level = 0.97)

##              1.5 %       98.5 %
## (Intercept) -3.750631 -0.2104243
## sacarosa      0.735599  1.0540224
```

El intervalos 0.735599 1.0540224 brinda la confianza de contener  $\beta_1$

i) Halle un intervalo del 97% de confianza para  $\sigma_2$

```
sig <- summary(modelo)$sigma
n <- nrow(PD1)
LI <- (n-2)*sig^2/qchisq(1-0.03/2,n-2)
LS <- (n-2)*sig^2/qchisq(0.03/2,n-2)
c(LI, LS)

## [1] 0.005881705 0.072144753
```

El intervalo 0.005881705 0.072144753 brinda la confianza de contener  $\sigma_2$

j) Halle e interprete el coeficiente de determinación.

```
correlacion <- cor(PD1$sacarosa, PD1$azucar)
correlacion

## [1] 0.9852942

determinacion <- correlacion^2
determinacion

## [1] 0.9708046
```

Coeficiente de correlación (r): 0.9852942 Coeficiente de determinación ( $R^2$ ): 0.9708046

Se interpreta que el 97% de la variabilidad en el porcentaje de azúcar puede ser explicada por la variabilidad en el porcentaje de sacarosa. Esto indica una relación lineal positiva fuerte entre ambas variables.

- k) Halle un intervalo del 97% de confianza para el porcentaje medio azúcar por caña cuando se tiene un 12% de porcentaje de sacarosa por caña.

```
summary(modelo)$ r.sq*100
```

```
## [1] 97.08046
```

La temperatura explica aproximadamente el 29% de las diferencias en la renta de bicicletas.

```
# 'modelo' es mi modelo de regresión
```

```
nueva_observacion <- data.frame(sacarosa = 12)
```

```
prediccion <- predict(modelo, newdata = nueva_observacion, interval =  
"confidence", level = 0.97)
```

```
prediccion
```

```
##          fit          lwr          upr
```

```
## 1 8.757201 8.576497 8.937906
```

Esto significa que con un 97% de confianza, podemos decir que el porcentaje de azúcar con respecto a la sacarosa en 12% estará entre 8.576497 8.937906.