

# Codigo

Alan de Jesús Felipe Cruz

2025-03-12

## R Markdown

```
# Cargamos los datos
library(readr)
datos <- read_csv("contaminacion.csv")

## Rows: 41 Columns: 8
## -- Column specification -----
## Delimiter: ","
## chr (1): Estado
## dbl (7): S02, temp, manu, popul, wind, precip, predays
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
tail(datos, 4)
```

```
## # A tibble: 4 x 8
##   Estado      S02  temp  manu popul  wind precip predays
##   <chr>    <dbl> <dbl> <dbl> <dbl> <dbl>   <dbl>   <dbl>
## 1 St. Louis    56  55.9   775   622    9.5    35.9     105
## 2 Washington   29  57.3   434   757    9.3    38.9     111
## 3 Wichita      8  56.6   125   277   12.7    30.6     82
## 4 Wilmington  36  54     80    80     9     40.2     114
```

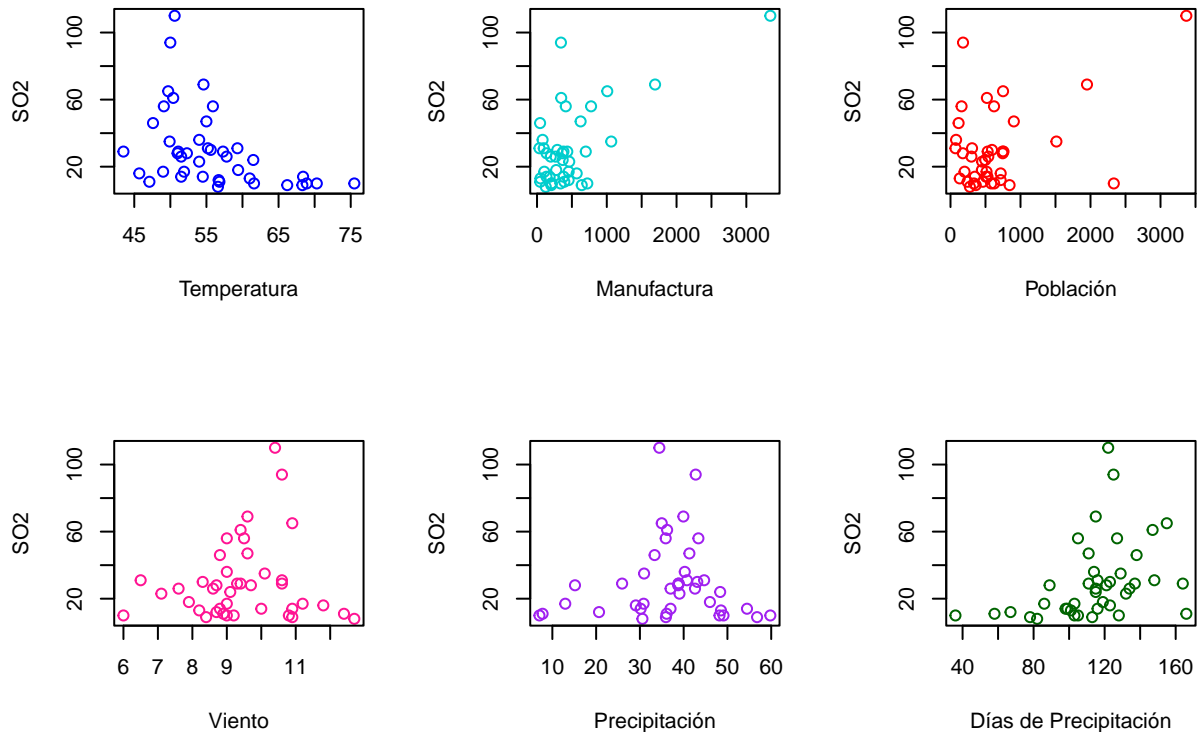
### a) Elección de variables X's y Y

```
y <- datos$S02
x1 <- datos$temp
x2 <- datos$manu
x3 <- datos$popul
x4 <- datos$wind
x5 <- datos$precip
x6 <- datos$predays
```

### b) Gráficas de los datos

```
par(mfrow = c(2, 3), cex=0.7)
plot(x1, y, xlab = "Temperatura", ylab = "S02", col="blue")
plot(x2, y, xlab = "Manufactura", ylab = "S02", col="cyan3")
plot(x3, y, xlab = "Población", ylab = "S02", col="red")
plot(x4, y, xlab = "Viento", ylab = "S02", col="deeppink" )
```

```
plot(x5, y, xlab = "Precipitación", ylab = "SO2", col="purple")
plot(x6, y, xlab = "Días de Precipitación", ylab = "SO2", col="darkgreen")
```



### c) Modelo para los datos

```
modelo.1 <- lm(y ~ x1+x2+x3+x4+x5+x6)
summary(modelo.1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.030  -9.265  -1.031   6.448  53.715
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 105.208843  51.192329   2.055  0.047610 *
## x1          -1.306404   0.664539  -1.966  0.057521 .
## x2           0.044603   0.011073   4.028  0.000299 ***
## x3          -0.018551   0.009893  -1.875  0.069369 .
## x4          -2.864006   1.937692  -1.478  0.148599
## x5           0.545474   0.378144   1.443  0.158310
## x6          -0.036340   0.169122  -0.215  0.831149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.23 on 34 degrees of freedom
## Multiple R-squared:  0.6419, Adjusted R-squared:  0.5787
```

```
## F-statistic: 10.16 on 6 and 34 DF, p-value: 1.961e-06
```

```
# Estimación de parámetros
```

```
cat("Los betas son: \n")
```

```
## Los betas son:
```

```
cat("Beta_0: ", coef(modelo.1)[1], "\n")
```

```
## Beta_0: 105.2088
```

```
cat("Beta_1: ", coef(modelo.1)[2], "\n")
```

```
## Beta_1: -1.306404
```

```
cat("Beta_2: ", coef(modelo.1)[3], "\n")
```

```
## Beta_2: 0.0446034
```

```
cat("Beta_3: ", coef(modelo.1)[4], "\n")
```

```
## Beta_3: -0.01855099
```

```
cat("Beta_4: ", coef(modelo.1)[5], "\n")
```

```
## Beta_4: -2.864006
```

```
cat("Beta_5: ", coef(modelo.1)[6], "\n")
```

```
## Beta_5: 0.5454745
```

```
cat("Beta_6: ", coef(modelo.1)[7], "\n")
```

```
## Beta_6: -0.03633992
```

#### d) analisis de varianza del modelo

##### Varianza explicada por el modelo

```
# Variabilidad explicada por el modelo
```

```
anov <- anova(modelo.1)
```

```
sc<- as.data.frame(anov$`Sum Sq`)
```

```
scregr <- sc[1,]+sc[2,]+sc[3,]+sc[4,]+sc[5,]+sc[6,]
```

```
Scerror <- sc[7,]
```

```
Sctotal <- scregr+Scerror
```

```
varexpl <- (scregr/Sctotal)*100
```

```
cat("La variabilidad explicada por el modelo es:", varexpl, "%\n")
```

```
## La variabilidad explicada por el modelo es: 64.19193 %
```

```
anov
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: y
```

##	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
## x1	1	4143.3	4143.3	17.8516	0.0001691	***
## x2	1	7234.8	7234.8	31.1714	2.999e-06	***
## x3	1	1400.9	1400.9	6.0356	0.0192801	*
## x4	1	303.7	303.7	1.3085	0.2606586	
## x5	1	1053.1	1053.1	4.5374	0.0404806	*
## x6	1	10.7	10.7	0.0462	0.8311492	

```
## Residuals 34 7891.3    232.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Verifique si la variable explicativa influye de forma conjunta y lineal sobre Y

Para esto simplemente obtenemos el p-value del modelo con lm

```
summary(modelo.1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x3 + x4 + x5 + x6)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.030  -9.265  -1.031   6.448  53.715
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 105.208843   51.192329   2.055 0.047610 *
## x1          -1.306404    0.664539  -1.966 0.057521 .
## x2           0.044603    0.011073   4.028 0.000299 ***
## x3          -0.018551    0.009893  -1.875 0.069369 .
## x4          -2.864006    1.937692  -1.478 0.148599
## x5           0.545474    0.378144   1.443 0.158310
## x6          -0.036340    0.169122  -0.215 0.831149
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.23 on 34 degrees of freedom
## Multiple R-squared:  0.6419, Adjusted R-squared:  0.5787
## F-statistic: 10.16 on 6 and 34 DF,  p-value: 1.961e-06
```

El p-value es menor que  $\alpha = 0.05$  por lo tanto, el modelo es significativo en su conjunto.

### e) Significancia estadística de los betas

```
coeficientes <- c("Intercepto", "x1", "x2", "x3", "x4", "x5", "x6")
for (i in 1:length(coeficientes)) {
  if((summary(modelo.1)$coefficients[,4][i] < 0.05) == TRUE){
    cat("El coeficiente de",coeficientes[i]," es significativo en el modelo.\n")
  }
}
```

```
## El coeficiente de Intercepto es significativo en el modelo.
## El coeficiente de x2 es significativo en el modelo.
```

### Intervalos de confianza para los beta's

```
print("Los intervalos de confianza son:\n")
```

```
## [1] "Los intervalos de confianza son:\n"
```

```
confint(modelo.1)
```

```
##              2.5 %          97.5 %
```

```
## (Intercept)  1.17351283 2.092442e+02
## x1          -2.65690935 4.410045e-02
## x2           0.02209946 6.710734e-02
## x3          -0.03865514 1.553148e-03
## x4          -6.80186997 1.073857e+00
## x5          -0.22300726 1.313956e+00
## x6          -0.38003710 3.073573e-01
```

#### g) Predicción de nuevos datos

Para esto eliminamos las variables que no son significativas y creamos un nuevo modelo.

```
# Predicción de un nuevo dato
newdato <- data.frame(x1=0,x2=500,x3=0,x4=0,x5=0,x6=0)
predict(modelo.1, newdato)
```

```
##          1
## 127.5105
```

#### h) Supuestos para el modelo

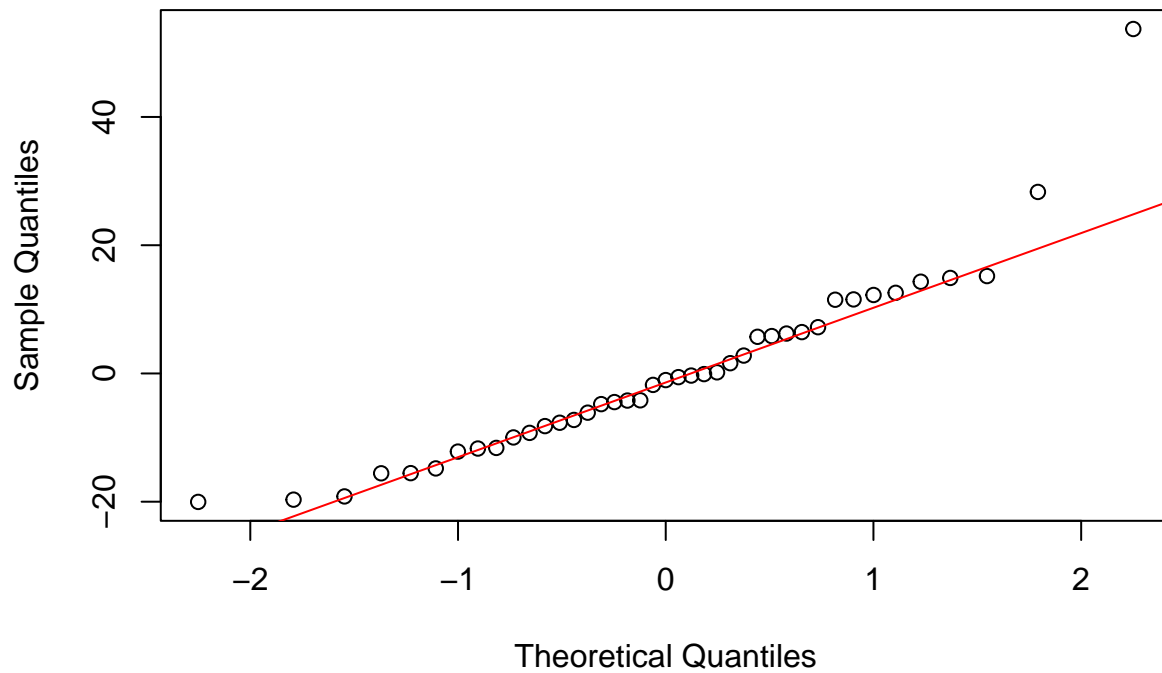
```
# Multicolinealidad
datos.m.n <- data.frame(y,x1,x2,x3,x4,x5,x6)
# Matriz de correlación
cor(datos.m.n)
```

```
##          y          x1          x2          x3          x4          x5
## y  1.00000000 -0.4336002  0.64527371  0.41539614  0.09469045  0.05429434
## x1 -0.43360020  1.0000000 -0.19111834  0.02165780 -0.34973963  0.38625342
## x2  0.64527371 -0.1911183  1.00000000  0.89578504  0.23922805 -0.03418952
## x3  0.41539614  0.0216578  0.89578504  1.00000000  0.23621108  0.01828827
## x4  0.09469045 -0.3497396  0.23922805  0.23621108  1.00000000 -0.01299438
## x5  0.05429434  0.3862534 -0.03418952  0.01828827 -0.01299438  1.00000000
## x6  0.37008844 -0.4319225  0.12958023  0.01755794  0.16169758  0.49526465
##          x6
## y  0.37008844
## x1 -0.43192247
## x2  0.12958023
## x3  0.01755794
## x4  0.16169758
## x5  0.49526465
## x6  1.00000000
```

No existen problemas de multicolinealidad.

```
# Prueba gráfica para la normalidad de residuos
residuos <- residuals(modelo.1)
qqnorm(residuos, main = "Gráfica de normalidad de los residuos.")
qqline(residuos, col = "red")
```

## Gráfica de normalidad de los residuos.



```
# Prueba analítica  
shapiro.test(residuos)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  residuos  
## W = 0.90168, p-value = 0.001859
```

La prueba indica que no hay normalidad en los residuos, y en el gráfico igual se puede ver.

## Homocedasticidad

```
library(lmtest)
```

```
## Cargando paquete requerido: zoo  
##  
## Adjuntando el paquete: 'zoo'  
## The following objects are masked from 'package:base':  
##  
## as.Date, as.Date.numeric
```

```
bptest(modelo.1)
```

```
##  
## studentized Breusch-Pagan test  
##  
## data:  modelo.1  
## BP = 3.9324, df = 6, p-value = 0.6858
```

Hay homocedasticidad de varianzas.

i) Concluya con base en el fenómeno físico y el análisis estadístico.