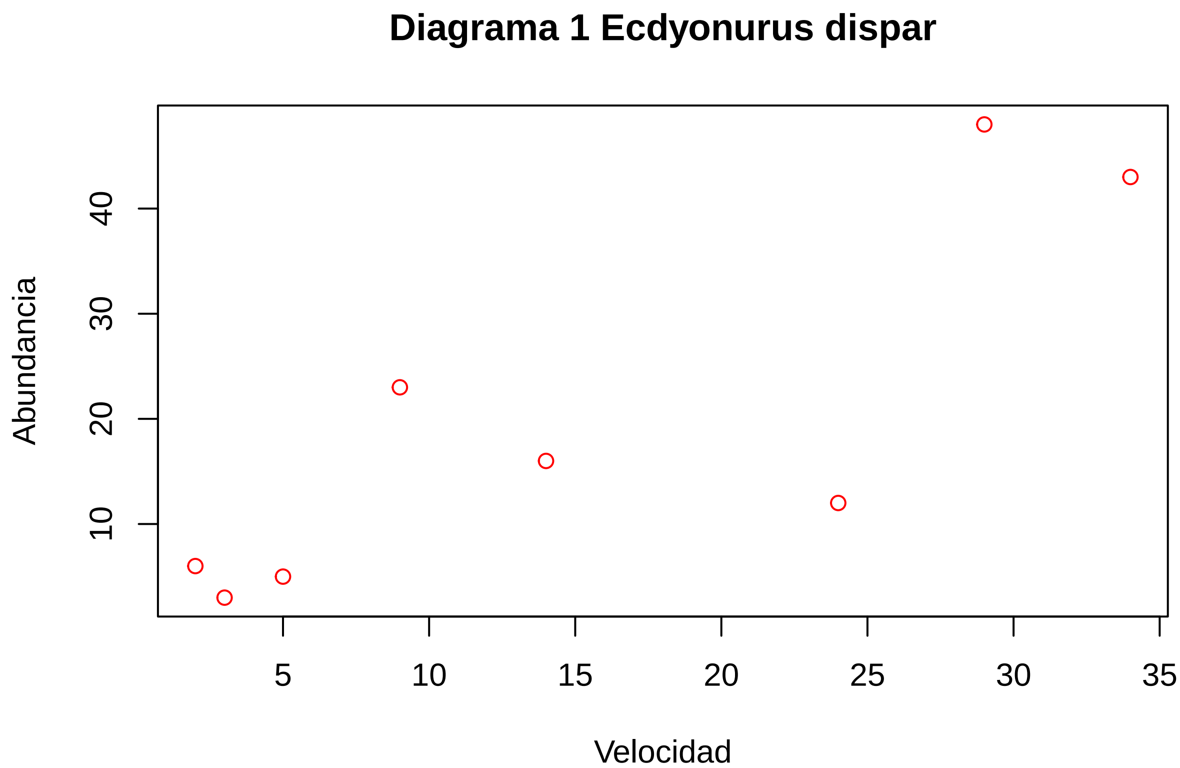
R7.R

Usuario

2025-11-27

```r
################################################################################
################################## PROBLEMA 1 ##################################
################################################################################

#se crea un data.frame de las variables llamado Cuadro_1
Cuadro_1 <- data.frame(
  Speed=c(2,3,5,9,14,24,29,34),
  Abundance=c(6,3,5,23,16,12,48,43)
)

#Se va a crear un diagrama para visualizar como se dispersan los datos
#inicialmente
plot(Cuadro_1$Speed, Cuadro_1$Abundance,
        xlab="Velocidad",
        ylab="Abundancia",
        col="red",
     main="Diagrama 1 Ecdyonurus dispar")
```




```r
#Se checa si mis datos son normales
shapiro.test(Cuadro_1$Speed) #Pvalue = 0.25
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Cuadro_1$Speed
## W = 0.89444, p-value = 0.2572
```

```r
shapiro.test(Cuadro_1$Abundance) #Pvalue = 0.10
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Cuadro_1$Abundance
## W = 0.85403, p-value = 0.1046
```

```r
#Como mis datos si son normales (sobrepasan ambos el 0.05) se va a realizar el
#p-value con el metodo de pearson utilizando "var.test"

var.test(Cuadro_1$Speed,Cuadro_1$Abundance, method="pearson") #Pvalue = 0.40
```

```
##
##  F test to compare two variances
```

```
##
## data:  Cuadro_1$Speed and Cuadro_1$Abundance
## F = 0.51564, num df = 7, denom df = 7, p-value = 0.4019
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.1032331 2.5755740
## sample estimates:
## ratio of variances
##          0.5156398
```

```
#El resultado es una buena correlación positiva de 0.4.

################################################################################
################################## PROBLEMA 2 ##################################
################################################################################

#Se crea un data.frame con todas las variables llamada Cuadro_2
Cuadro_2 <- data.frame(
  Gp=c("T0","T0","T0","T0","T1","T1","T1"),
  Block=c(1,2,3,4,1,2,3),
  pH=c(5.40,5.65,5.14,5.14,5.14,5.10,4.70),
  N=c(.188,.165,.260,.169,.164,.094,.100),
  Dens=c(.92,1.04,0.95,1.10,1.12,1.22,1.52),
  P=c(215,208,300,248,174,129,117),
  Ca=c(16.35,12.25,13.02,11.92,14.17,8.55,8.74),
  Mg=c(7.65,5.15,5.68,7.88,8.12,6.92,8.16),
  K=c(0.72,.71,.68,1.09,.70,.81,.39),
  Na=c(1.14,.94,.60,1.01,2.17,2.67,3.32),
  Conduc=c(1.09,1.35,1.41,1.64,1.85,3.18,4.16)
)

#Ahora se va a realizar un test de P-value y r utilizando cor.test de pH con
#cada uno de los datos a partir de N hasta Na del Cuadro_2 para ver su
#normalidad

cor.test(Cuadro_2$pH,Cuadro_2$N)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Cuadro_2$pH and Cuadro_2$N
## t = 0.94167, df = 5, p-value = 0.3896
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.5156557  0.8830741
## sample estimates:
##       cor
## 0.3881145
```

```
cor.test(Cuadro_2$pH,Cuadro_2$Dens)
```

```
##
##  Pearson's product-moment correlation
```

```
##
## data:  Cuadro_2$pH and Cuadro_2$Dens
## t = -2.7306, df = 5, p-value = 0.04125
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.96468856 -0.04943664
## sample estimates:
##        cor
## -0.7736913
```

```
cor.test(Cuadro_2$pH,Cuadro_2$P)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Cuadro_2$pH and Cuadro_2$P
## t = 1.0367, df = 5, p-value = 0.3474
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.4865625  0.8913418
## sample estimates:
##      cor
## 0.420612
```

```
cor.test(Cuadro_2$pH,Cuadro_2$Ca)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Cuadro_2$pH and Cuadro_2$Ca
## t = 1.5451, df = 5, p-value = 0.183
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3227349  0.9253846
## sample estimates:
##       cor
## 0.5684873
```

```
cor.test(Cuadro_2$pH,Cuadro_2$Mg)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Cuadro_2$pH and Cuadro_2$Mg
## t = -1.7265, df = 5, p-value = 0.1448
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9342417  0.2629006
## sample estimates:
##        cor
## -0.6111533
```

```r
cor.test(Cuadro_2$pH,Cuadro_2$K)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Cuadro_2$pH and Cuadro_2$K
## t = 0.89317, df = 5, p-value = 0.4127
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.5302315  0.8785775
## sample estimates:
##       cor
## 0.3709419
```

```r
cor.test(Cuadro_2$pH,Cuadro_2$Na)
```

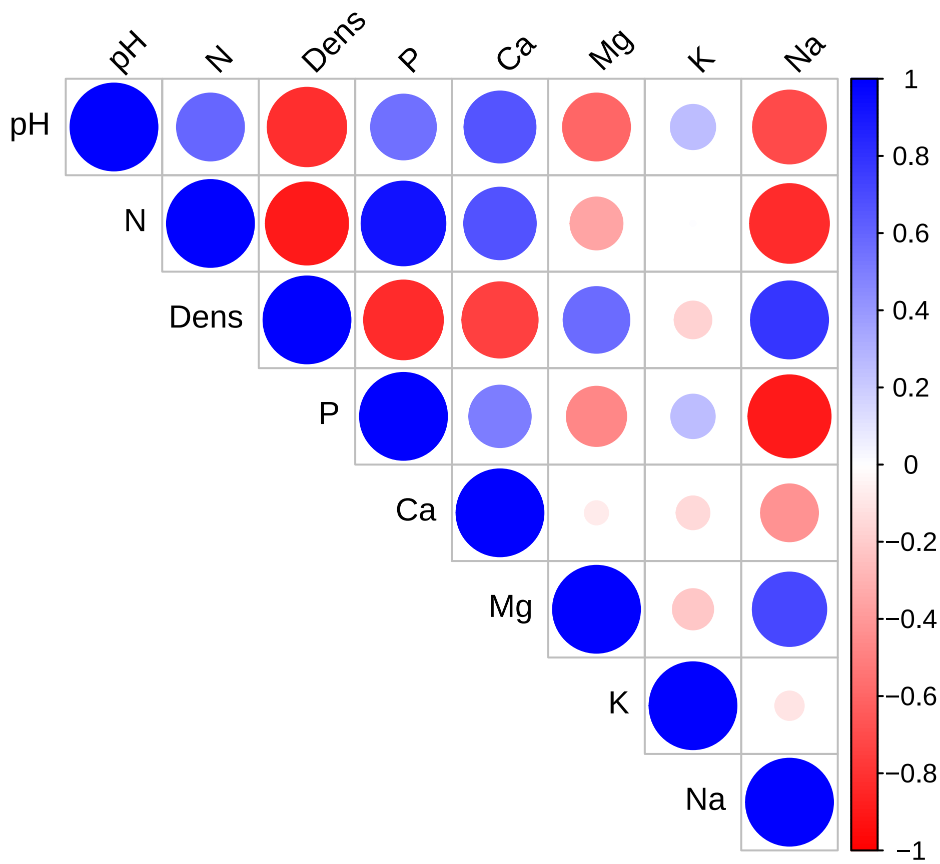
```
##
##  Pearson's product-moment correlation
##
## data:  Cuadro_2$pH and Cuadro_2$Na
## t = -2.2637, df = 5, p-value = 0.07301
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.95360061  0.08965103
## sample estimates:
##       cor
## -0.711438
```

```r
#Ahora se crea otro data.frame sobre los datos recopilados de r y P que se va a
#llamar Cuadro_3
Cuadro_3 <- data.frame(
  pH=c(5.40,5.65,5.14,5.14,5.14,5.10,4.70),
  r=c(0.38,-0.77,0.42,0.56,-0.61,0.37,-0.71),
  P=c(0.38,0.04,0.34,0.18,0.14,0.41,0.07)
)

#para poder crear el grafico, necesitamos descargar el paquete corrplot y cargar
#en la libreria
library(corrplot)
```

```
## corrplot 0.95 loaded
```

```r
#Despues se tienen que leer los datos del data.frame Cuadro_2

datos <- data.frame(Cuadro_2,header=TRUE)

#al hacer la prueba de cor.test, un dato es menor a 0.05, por lo tanto se va a
#calcular la matriz de la correlación con el metodo de spearman

M <- cor(datos[,c("pH","N","Dens","P","Ca","Mg","K","Na")], method="spearman")
```

```r
#Ya por ultimo se va a graficar
corrplot(M, method="circle", type="upper",
         col=colorRampPalette(c("red","white","blue"))(200),
         tl.col="black", tl.srt=45)
```




```r
################################################################################
################################ PROBLEMA 3 ####################################
################################################################################

# Ingresamos manualmente los cuatro conjuntos de datos
anscombe <- data.frame(
  x1 = c(10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5),
  y1 = c(8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.68),
  x2 = c(10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5),
  y2 = c(9.14, 8.14, 8.74, 8.77, 9.26, 8.10, 6.13, 3.10, 9.13, 7.26, 4.74),
  x3 = c(10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5),
  y3 = c(7.46, 6.77, 12.74, 7.11, 7.81, 8.84, 6.08, 5.39, 8.15, 6.42, 5.73),
  x4 = c(8, 8, 8, 8, 8, 8, 8, 19, 8, 8, 8),
  y4 = c(6.58, 5.76, 7.71, 8.84, 8.47, 7.04, 5.25, 12.50, 5.56, 7.91, 6.89)
)

# Mostrar los datos para verificar que se cargaron correctamente
print("Datos de Anscombe:")
```

```
## [1] "Datos de Anscombe:"
```

```r
print(anscombe)
```

```
##    x1    y1 x2   y2 x3    y3 x4    y4
## 1  10  8.04 10 9.14 10  7.46  8  6.58
## 2   8  6.95  8 8.14  8  6.77  8  5.76
## 3  13  7.58 13 8.74 13 12.74  8  7.71
## 4   9  8.81  9 8.77  9  7.11  8  8.84
## 5  11  8.33 11 9.26 11  7.81  8  8.47
## 6  14  9.96 14 8.10 14  8.84  8  7.04
## 7   6  7.24  6 6.13  6  6.08  8  5.25
## 8   4  4.26  4 3.10  4  5.39 19 12.50
## 9  12 10.84 12 9.13 12  8.15  8  5.56
## 10  7  4.82  7 7.26  7  6.42  8  7.91
## 11  5  5.68  5 4.74  5  5.73  8  6.89
```

```r
# Creamos una función para calcular las mismas estadísticas para cada conjunto
calcular_estadisticas <- function(x, y, conjunto) {
  cat("=== CONJUNTO", conjunto, "===\n")
  cat("Media de x:", mean(x), "\n")
  cat("Varianza de x:", var(x), "\n")
  cat("Media de y:", mean(y), "\n")
  cat("Varianza de y:", var(y), "\n")
  cat("Correlación:", cor(x, y), "\n")

  # Ajustamos un modelo de regresión lineal: y = a + bx
  modelo <- lm(y ~ x)
  cat("Recta de regresión: y =", round(coef(modelo)[1], 2), "+", round(coef(
    modelo)[2], 2), "x\n")
  cat("Coeficiente de determinación R²:", summary(modelo)$r.squared, "\n\n")
  # R² indica qué tan bien se ajusta la recta
}

# Aplicamos la función a los cuatro conjuntos de datos
calcular_estadisticas(anscombe$x1, anscombe$y1, "I")
```

```
## === CONJUNTO I ===
## Media de x: 9
## Varianza de x: 11
## Media de y: 7.500909
## Varianza de y: 4.127269
## Correlación: 0.8164205
## Recta de regresión: y = 3 + 0.5 x
## Coeficiente de determinación R²: 0.6665425
```

```r
calcular_estadisticas(anscombe$x2, anscombe$y2, "II")
```

```
## === CONJUNTO II ===
## Media de x: 9
## Varianza de x: 11
## Media de y: 7.500909
## Varianza de y: 4.127629
## Correlación: 0.8162365
```

```
## Recta de regresión: y = 3 + 0.5 x
## Coeficiente de determinación R²: 0.666242
```

```r
calcular_estadisticas(anscombe$x3, anscombe$y3, "III")
```

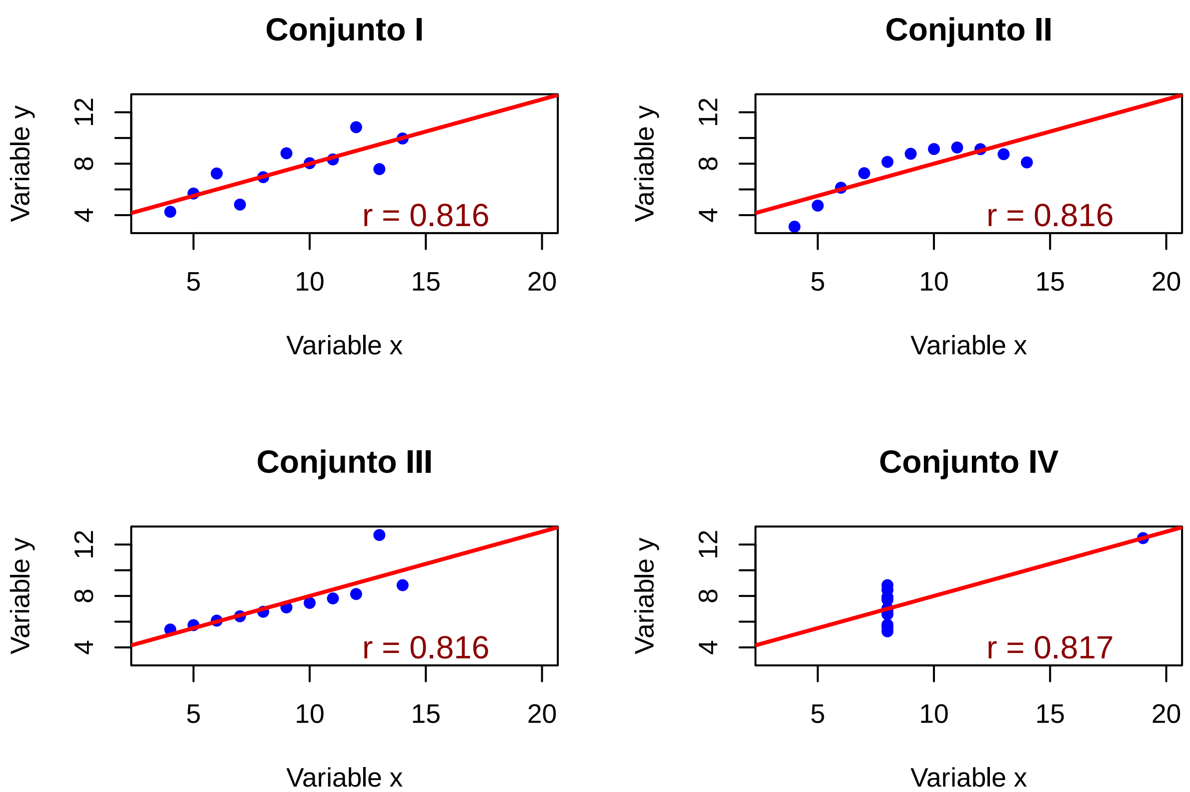
```
## === CONJUNTO III ===
## Media de x: 9
## Varianza de x: 11
## Media de y: 7.5
## Varianza de y: 4.12262
## Correlación: 0.8162867
## Recta de regresión: y = 3 + 0.5 x
## Coeficiente de determinación R²: 0.666324
```

```r
calcular_estadisticas(anscombe$x4, anscombe$y4, "IV")
```

```
## === CONJUNTO IV ===
## Media de x: 9
## Varianza de x: 11
## Media de y: 7.500909
## Varianza de y: 4.123249
## Correlación: 0.8165214
## Recta de regresión: y = 3 + 0.5 x
## Coeficiente de determinación R²: 0.6667073
```

```r
# Configuramos la ventana gráfica para mostrar 4 gráficos en una misma ventana
#(2 filas × 2 columnas)
par(mfrow = c(2, 2))

# Función para crear cada gráfico individual
graficar_conjunto <- function(x, y, conjunto) {
  # Creamos el gráfico de dispersión
  plot(x, y,
       main = paste("Conjunto", conjunto),
       xlab = "Variable x",
       ylab = "Variable y",
       pch = 16,
       col = "blue",
       xlim = c(3, 20),
       ylim = c(3, 13))

  # Añadimos la línea de regresión en color rojo
  abline(lm(y ~ x), col = "red", lwd = 2)

  # Añadimos el valor de correlación como texto en el gráfico
  texto_cor <- paste("r =", round(cor(x, y), 3))
  text(15, 4, texto_cor, col = "darkred", cex = 1.2)  # Posicionamos el texto en
  #la esquina
}

# Creamos los cuatro gráficos
graficar_conjunto(anscombe$x1, anscombe$y1, "I")
```

```r
graficar_conjunto(anscombe$x2, anscombe$y2, "II")
graficar_conjunto(anscombe$x3, anscombe$y3, "III")
graficar_conjunto(anscombe$x4, anscombe$y4, "IV")
```




```r
# Restauramos la configuración gráfica normal (1 gráfico por ventana)
par(mfrow = c(1, 1))

# INTERPRETACIÓN DE RESULTADOS
#Estadísticas casi idénticas para los cuatro conjuntos pero siendo gràficas muy
#diferentes:
#Conjunto I: Relación lineal aparentemente normal.
#Conjunto II: Relación curvilínea (no lineal).
#Conjunto III: Relación lineal con un outlier que influye en la regresión.
#Conjunto IV: Un solo punto extremo que domina la correlación.

#Si solo confiamos en estadísticas como la correlación o la recta de regresión,
#podríamos llegar a conclusiones erróneas. Los cuatro conjuntostienen la misma
#correlación (~0.816) y misma recta de regresión (y = 3.00 + 0.50x), pero
#representan relaciones totalmente diferentes entre las variables.
```