

Práctica 5

Álvaro Miranda García

2023-03-14

1) Aplica un contraste de hipótesis basado en la media a:

y1

y1 e y2

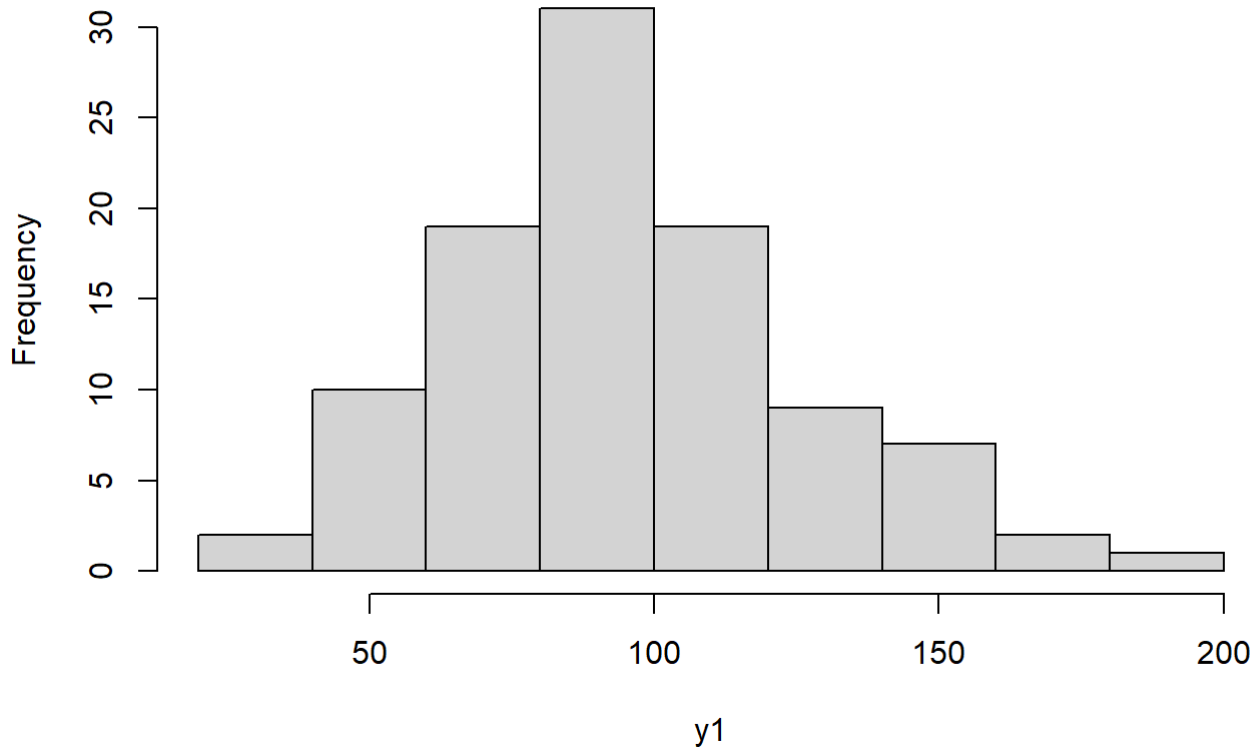
```
set.seed (1)

z = rnorm (100)
x = rpois (100,10.3)
y = rbinom (100, 1, .25)

y1 = 5*z+x*10+rnorm (100,2,1)
y2 = 5*z+x*12+rnorm (100,2,1)

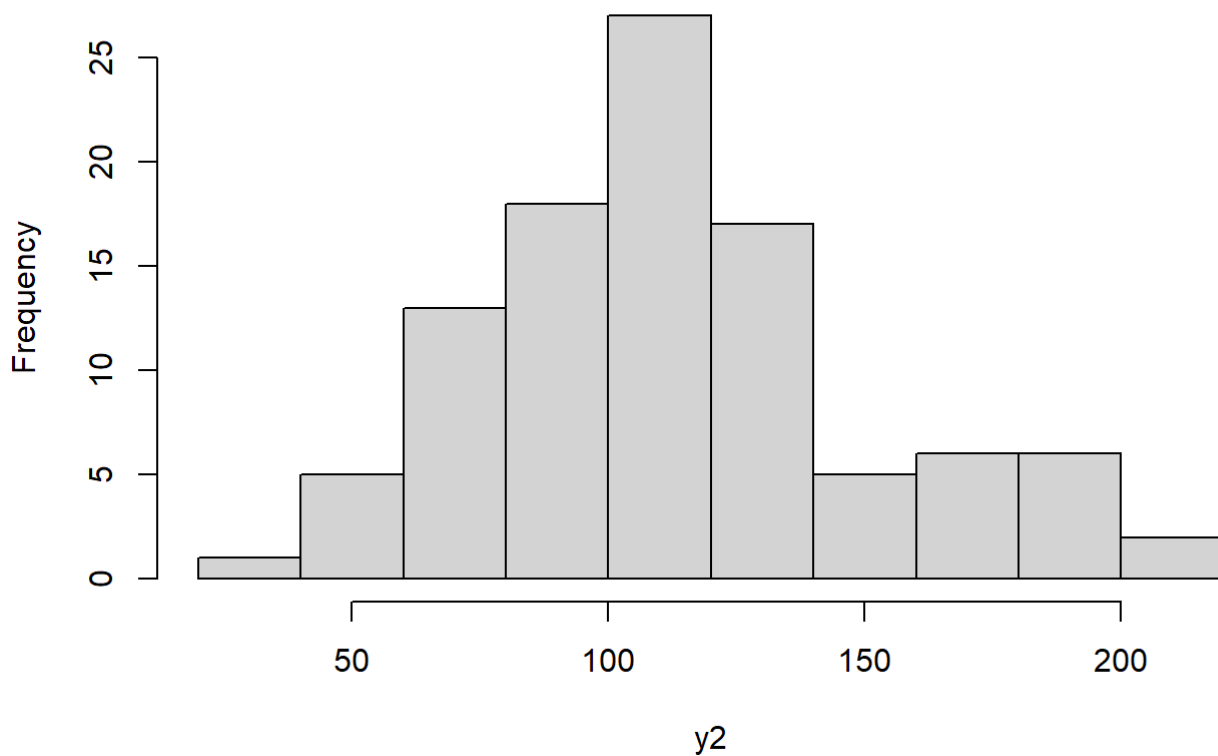
hist (y1)
```

Histogram of y1



```
hist (y2)
```

Histogram of y2



```
t.test (y1)
```

```
##  
## One Sample t-test  
##  
## data: y1  
## t = 30.231, df = 99, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 89.09856 101.61594  
## sample estimates:  
## mean of x  
## 95.35725
```

```
t.test (y2)
```

```
##
## One Sample t-test
##
## data: y2
## t = 30.454, df = 99, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 106.4637 121.3036
## sample estimates:
## mean of x
## 113.8837
```

```
shapiro.test(y1)
```

```
##
## Shapiro-Wilk normality test
##
## data: y1
## W = 0.97469, p-value = 0.05106
```

```
shapiro.test(y2)
```

```
##
## Shapiro-Wilk normality test
##
## data: y2
## W = 0.97289, p-value = 0.03696
```

Aplicando o el test de Shapiro Wilk, y1 tiene un valor p mayor de 0,05 (0.05106), por lo que no podríamos descartar una distribución normal (H0). y2 tiene un valor p de 0.03696, menor de 0.05, por lo que con alguna probabilidad podriamos descartar una distribución normal, por tanto, sería H1. Con t.test () lo comprobamos.

2) ¿Por qué decimos que la correlación lineal es una prueba de correlación paramétrica? ¿En qué se diferencian las pruebas paramétricas de las no paramétricas?

Una correlación paramétrica se aplica para casos en donde la distribución de los datos sigue una curva Gausiana o normal. Las pruebas no paramétricas u ordinales se encargan de analizar datos que no tienen una distribución particular y se basa en una hipótesis, pero los datos no están organizados de forma normal.

3) Calcula la correlación entre las variables almacenadas en la tabla 'data'. ¿Qué variables se encuentran más asociadas?

```
library(readxl)
datos <- read_excel("C:/P7/datos.xlsx")
View(datos)

data1 = as.data.frame(datos)

View (data1)
```

```
cor (data1)
```

```
##          longitud      ancho      grosor      peso
## longitud 1.000000000  0.402029764  0.004679832  0.95558944
## ancho    0.402029764  1.000000000 -0.001294837  0.50833451
## grosor   0.004679832 -0.001294837  1.000000000 -0.05827998
## peso     0.955589444  0.508334506 -0.058279977  1.00000000
```

Longitud y ancho = 0.4020 Longitud y grosor = 0.0046 Longitud y peso = 0.9555

Ancho y grosor = -0.0012 Ancho y peso = 0.5083 Peso y grosor = -0.0582

Longitud y peso, al ser su coeficiente de correlación más cercano a 1, se encuentran más asociadas.

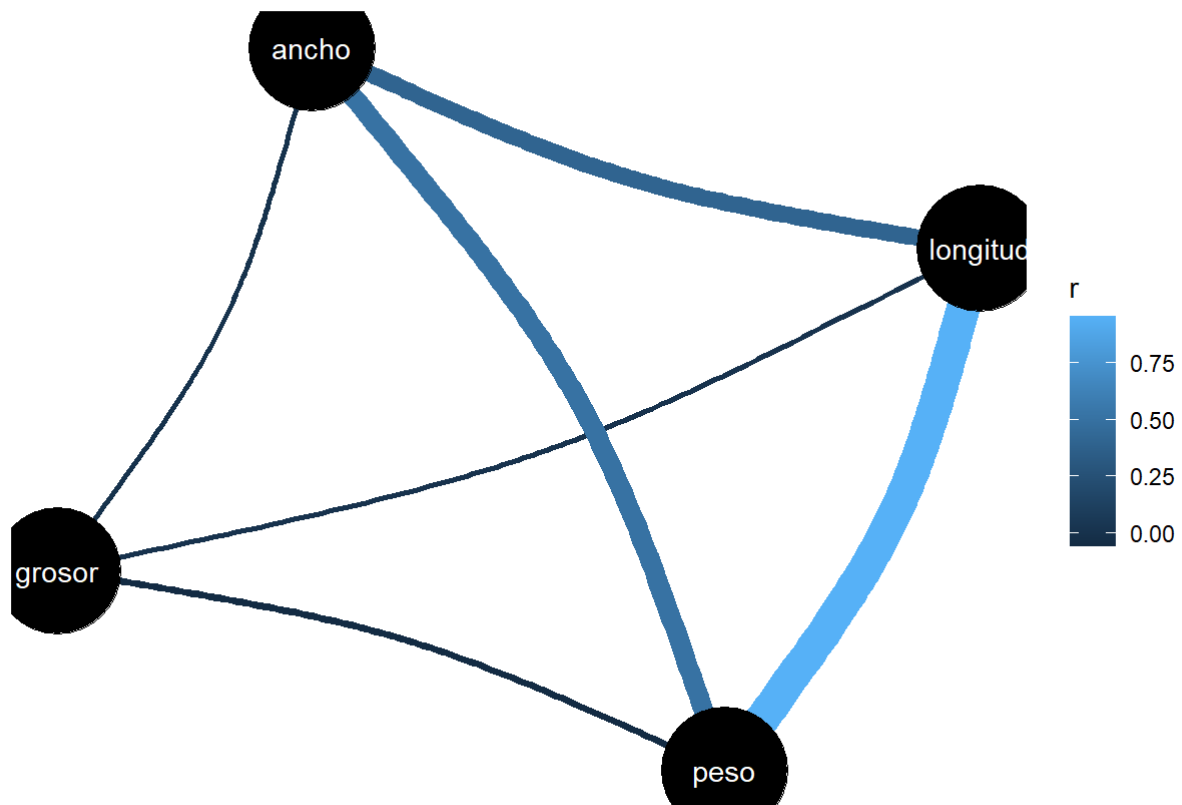
4) Calcula los coeficientes de correlación de las variables junto con el nivel de significancia (p-value) en 1 solo gráfico. Interpreta los resultados.

```
library (correlation)
resultados = correlation(data1)
resultados
```

```
## # Correlation Matrix (pearson-method)
##
## Parameter1 | Parameter2 |      r |      95% CI |      t(38) |      p
## -----|-----|-----|-----|-----|-----
## longitud  |      ancho |    0.40 | [ 0.10, 0.63] |      2.71 | 0.040*
## longitud  |     grosor | 4.68e-03 | [-0.31, 0.32] |      0.03 | > .999
## longitud  |      peso |    0.96 | [ 0.92, 0.98] |     19.99 | < .001***
## ancho     |     grosor | -1.29e-03 | [-0.31, 0.31] |  -7.98e-03 | > .999
## ancho     |      peso |    0.51 | [ 0.23, 0.71] |      3.64 | 0.004**
## grosor    |      peso |   -0.06 | [-0.36, 0.26] |     -0.36 | > .999
##
## p-value adjustment method: Holm (1979)
## Observations: 40
```

```
library (ggplot2)
library (ggpubr)

plot (resultados)
```



5) Emplea una función para obtener en una matriz de correlación lineal, IC 95% y pvalue de todas las variables en el data frame 'data'.

```
cor.test(data1$longitud, data1$ancho)
```

```
##
## Pearson's product-moment correlation
##
## data: data1$longitud and data1$ancho
## t = 2.7066, df = 38, p-value = 0.01012
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.1034798 0.6341239
## sample estimates:
## cor
## 0.4020298
```

```
cor.test(data1$grosor, data1$peso)
```

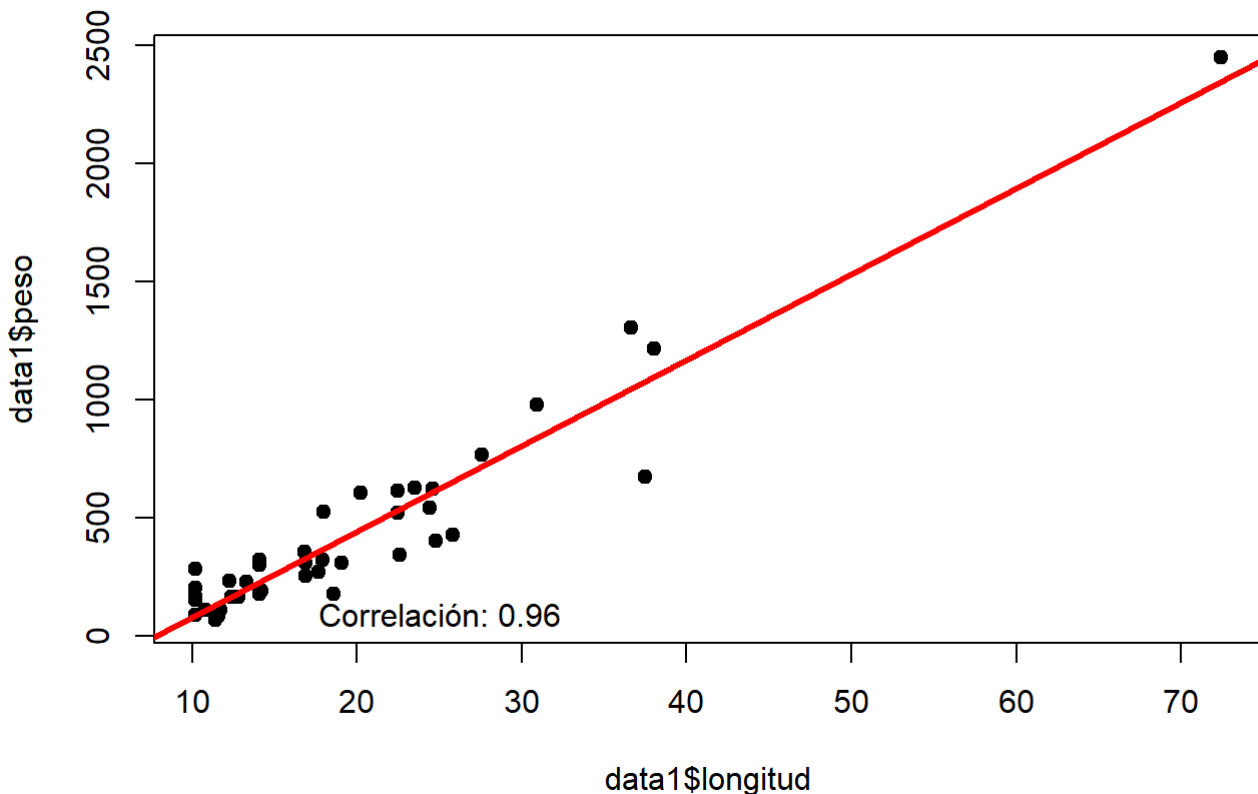
```
##
## Pearson's product-moment correlation
##
## data: data1$grosor and data1$peso
## t = -0.35987, df = 38, p-value = 0.7209
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.3631956 0.2579117
## sample estimates:
## cor
## -0.05827998
```

6) Visualiza gráficamente la correlación lineal existente entre las variables 'longitud' y 'peso'.

```
plot(data1$longitud, data1$peso, pch = 19, col = "black")

# Línea de regresión
abline(lm(data1$peso ~ data1$longitud), col = "red", lwd = 3)

# Correlación de Pearson
text(paste("Correlación:", round(cor(data1$peso, data1$longitud), 2)), x = 25, y = 95)
```

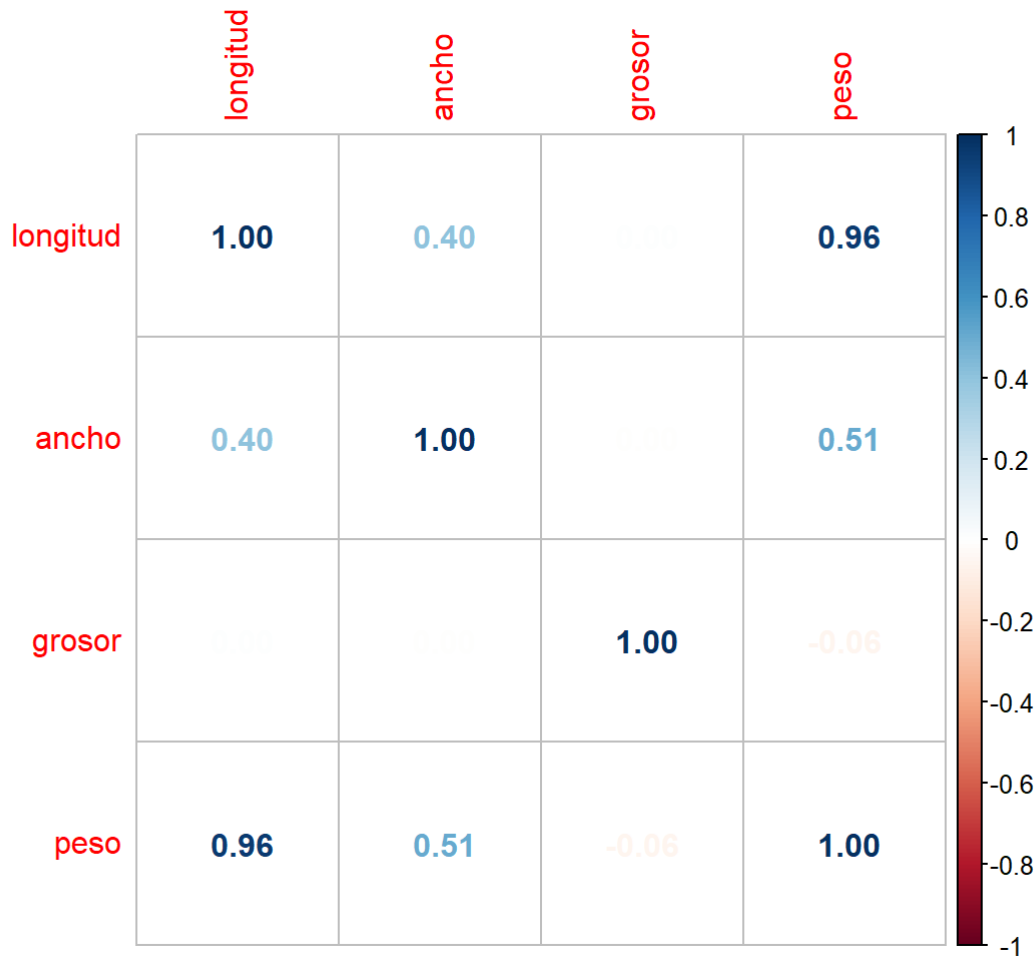


7) Emplea la librería `corrplot()` para visualizar la correlación entre variables.

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
corrplot(cor (data1), method = 'number')
```



8) A partir de la siguiente secuencia de valores numéricos: • Distancia (km):

1.1,100.2,90.3,5.4,57.5,6.6,34.7,65.8,57.9,86.1 • Número de cuentas (valor absoluto):

110,2,6,98,40,94,31,5,8,10

- Crea 2 vectores: 'distancia' y 'n_piezas' para almacenarlos en un data frame
- Calcula el coeficiente de correlación
- Calcula el nivel de significancia
- Calcula el Intervalo de confianza al 95% en relación con el coeficiente de correlación
- ¿Qué intensidad y dirección presentan ambas variables?
- ¿Es significativa esta relación?
- Resulta apropiado afirmar la correlación (o no) entre variables con un tamaño muestral tan reducido (n=10).

```
#a.
distancia <- c(1.1,100.2,90.3,5.4,57.5,6.6,34.7,65.8,57.9,86.1)
n_piezas <- c(110,2,6,98,40,94,31,5,8,10)

df <- data.frame(distancia, n_piezas)
```

```
#b.
cor(df$distancia, df$n_piezas)
```

```
## [1] -0.9249824
```

```
#c.

cor.test(df$distancia, df$n_piezas)
```

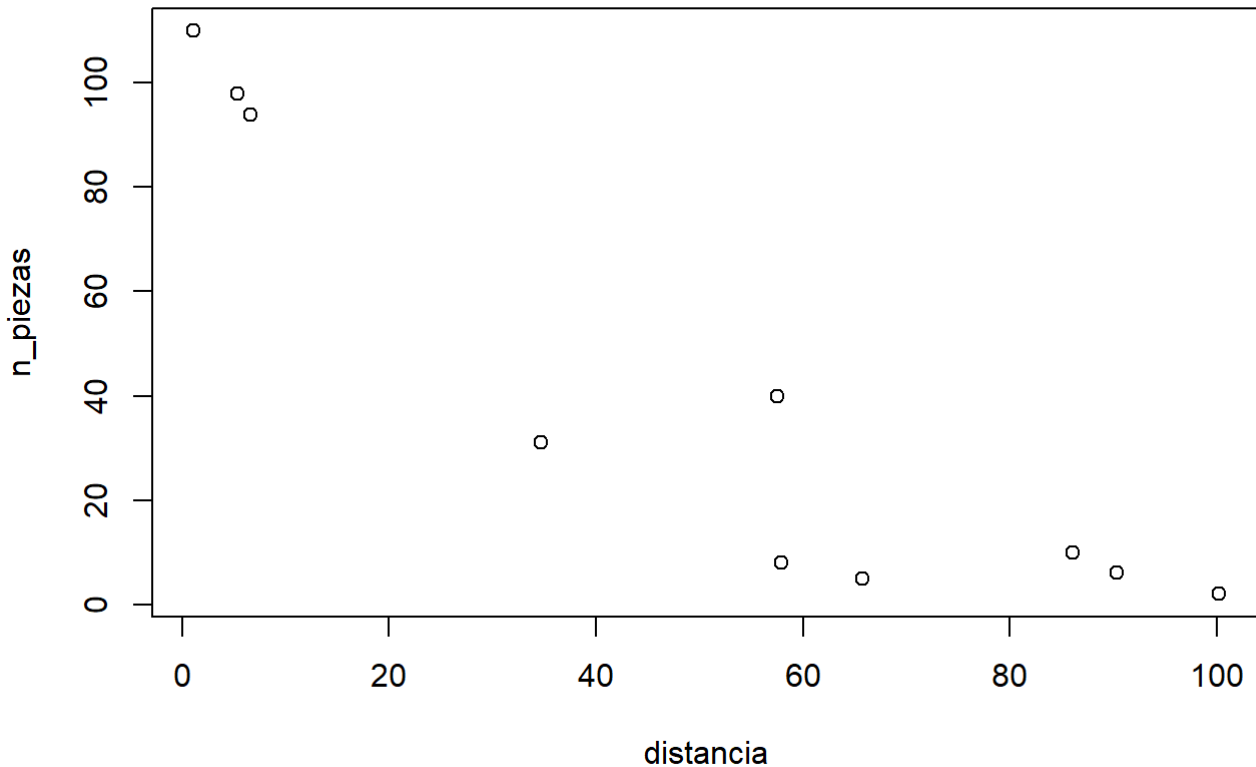
```
##
## Pearson's product-moment correlation
##
## data: df$distancia and df$n_piezas
## t = -6.8847, df = 8, p-value = 0.0001265
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.9824414 -0.7072588
## sample estimates:
## cor
## -0.9249824
```

#d.

```
correlation::correlation(df)
```

```
## # Correlation Matrix (pearson-method)
##
## Parameter1 | Parameter2 |      r |      95% CI | t(8) |      p
## -----
## distancia |   n_piezas | -0.92 | [-0.98, -0.71] | -6.88 | < .001***
##
## p-value adjustment method: Holm (1979)
## Observations: 10
```

#Así es la representación gráfica:
plot (df)



e. El coeficiente (-0.92) indica que se acerca a una relación lineal inversa, pues se acerca a -1.

f. Sí, la relación es significativa, al ser el p-value < -0,05.

g. No, es complicado afirmar una correlación entre variables con un tamaño muestral tan reducido.

9) Explícame con un ejemplo en R la diferencia entre una relación lineal y monótona entre 2 variables.

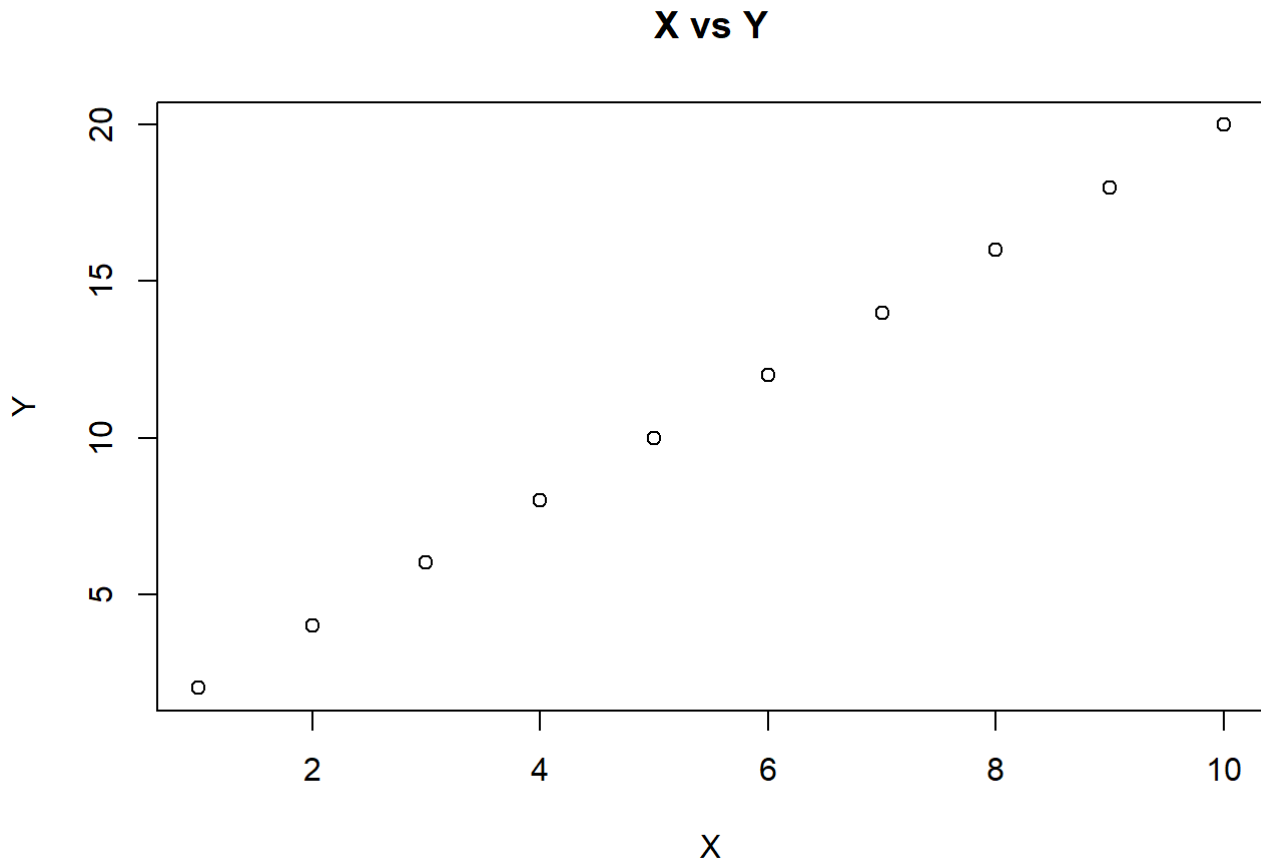
Una relación lineal entre dos variables significa que existe una relación entre las dos variables de tal forma que una sube o baja de manera proporcional a la otra. Estas se pueden representar en una gráfica como una línea recta.

Un ejemplo en R de esta relación lineal sería:

```
# Generar datos
x <- c(1,2,3,4,5,6,7,8,9,10)

y <- c(2,4,6,8,10,12,14,16,18,20)

# Representación gráfica
plot(x, y, main="X vs Y",
      xlab="X", ylab="Y")
```



10) ¿Qué tipo de prueba de correlación se aplica a las variables que experimentan una relación monótona? Expón un ejemplo en R.

La monotonía de una función indica que es fácil de adivinar.

Una prueba de correlación adecuada para variables que experimentan una relación monótona es la correlación de Spearman. Esta prueba mide la fuerza y dirección de la asociación entre dos variables ordinales.

Un ejemplo de la prueba de correlación de Spearman en R es:

```
# Crear datos
x <- c(1, 2, 4, 5, 7, 8)
y <- c(2, 3, 5, 6, 8, 9)

# Calcular la correlación de Spearman
cor.test(x, y, method = "spearman")
```

```
##
## Spearman's rank correlation rho
##
## data: x and y
## S = 0, p-value = 0.002778
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 1
```