Correlación lineal

Natalia Jimenez Guillen

2024-03-17

library(readxl)  
data <-as.data.frame(read\_excel("C:/data.xls"))  
View(data)  
print(data)

## id peso altura  
## 1 1 23.47 34.72  
## 2 2 5.81 87.59  
## 3 3 67.92 16.28  
## 4 4 12.36 52.94  
## 5 5 38.09 11.67  
## 6 6 54.73 65.83  
## 7 7 89.24 28.15  
## 8 8 16.57 43.26  
## 9 9 73.10 78.94  
## 10 10 42.85 25.09  
## 11 11 6.28 59.37  
## 12 12 91.63 37.62  
## 13 13 29.04 20.48  
## 14 14 50.17 91.75  
## 15 15 18.92 6.82  
## 16 16 81.36 74.56  
## 17 17 37.59 48.29  
## 18 18 64.28 13.75  
## 19 19 10.73 82.36  
## 20 20 45.81 39.21  
## 21 21 79.05 57.83  
## 22 22 27.36 22.67  
## 23 23 58.29 83.94  
## 24 24 14.67 30.17  
## 25 25 33.92 63.45  
## 26 26 70.18 47.92  
## 27 27 8.73 18.26  
## 28 28 96.51 70.83  
## 29 29 21.84 9.51  
## 30 30 47.36 54.29

#Actividad 1

La correlación lineal es una medida estadística que describe la relación entre dos variables numéricas. Indica si las variables tienden a cambiar juntas (correlación positiva), en direcciones opuestas (correlación negativa) o no tienen relación (correlación nula).

#Actividad 2 La correlación lineal es una prueba de correlación paramétrica porque asume que los datos siguen una distribución específica (normal) y que cumplen con ciertas condicioneS. Las pruebas paramétricas y las no paramétricas se diferencian esencialmente en que las primeras asumen que los datos siguen una distribución especifica y cumplen con ciertas condiciones, mientras que las no paramétricas no hacen suposiciones sobre la distribución de los datos y son más flexibles.

#Actividad 3

correlación\_data <- cor(data)  
print(correlación\_data)

## id peso altura  
## id 1.000000000 0.08321171 0.003792534  
## peso 0.083211709 1.00000000 0.147768260  
## altura 0.003792534 0.14776826 1.000000000

Captura de pantalla de un celular

Descripción generada automáticamente

Imagen que contiene Gráfico

Descripción generada automáticamenteInterfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

#Actividad 5

library(correlation)  
matriz <-correlation(data)  
matriz

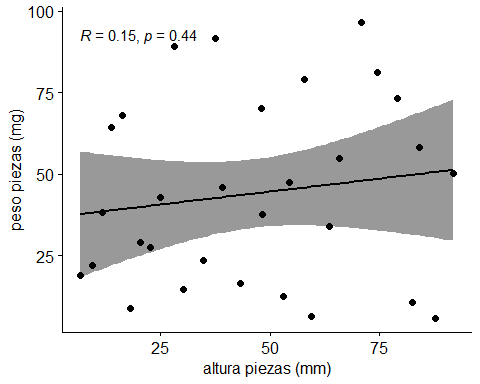
## # Correlation Matrix (pearson-method)  
##   
## Parameter1 | Parameter2 | r | 95% CI | t(28) | p  
## -------------------------------------------------------------------  
## id | peso | 0.08 | [-0.29, 0.43] | 0.44 | > .999  
## id | altura | 3.79e-03 | [-0.36, 0.36] | 0.02 | > .999  
## peso | altura | 0.15 | [-0.22, 0.48] | 0.79 | > .999  
##   
## p-value adjustment method: Holm (1979)  
## Observations: 30

#Actividad 6

library(ggpubr)

## Loading required package: ggplot2

library(ggplot2)  
ggscatter(data, x= "altura", y="peso",  
 add = "reg.line", conf.int = TRUE,  
 cor.coef=TRUE, cor.method = "pearson",  
 xlab = "altura piezas (mm)", ylab = "peso piezas (mg)")

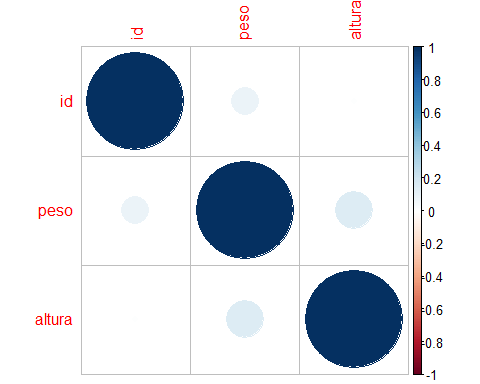


#Actividad 7

library(corrplot)

## corrplot 0.92 loaded

corrplot(cor(data))

 #Actividad 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

#(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)  
dataframe8 <-data.frame(distancia, n\_piezas)  
print(dataframe8)

## distancia n\_piezas  
## 1 1.1 110  
## 2 100.2 2  
## 3 90.3 6  
## 4 5.4 98  
## 5 57.5 40  
## 6 6.6 94  
## 7 34.7 31  
## 8 65.8 5  
## 9 57.9 8  
## 10 86.1 10

#(b)

correlacion\_dataframe8 <-cor(dataframe8)  
View(correlacion\_dataframe8)

#(c)

significancia\_dataframe8 <-cor.test(dataframe8$distancia, dataframe8$n\_piezas)$p.value  
print(significancia\_dataframe8)

## [1] 0.0001264706

#(d)

intervaloconfianza <- cor.test(dataframe8$distancia,dataframe8$n\_piezas)$conf.int  
print(intervaloconfianza)

## [1] -0.9824414 -0.7072588  
## attr(,"conf.level")  
## [1] 0.95

#(e)

intensidad\_direccion <- cor(dataframe8$distancia, dataframe8$n\_piezas)  
print(intensidad\_direccion)

## [1] -0.9249824

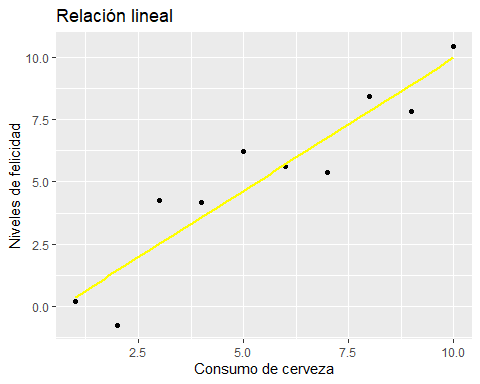
#(f) La relacion entre las dos variables es negativa debido a que el resultado del ejercicio anterior es más cercano a -1. Por lo tanto, esto singifica que cuando una variable tiende a disminuir cuando la otra aumenta .

#(g) Una muestra tan pequeña, implica que la interpretación de la correlación entre las variables debe ser más cautelosa. Las muestras reducidas pueden resultar en estimaciones menos conciables y más susceptibles a variaciones aleatorias. Siempre que sea posible, es mas conveniente obtener mas datos. Por lo tanto no es apropiado trabajar con tan pocos datos.

#Ejercicio 9 Relación lineal. Cuantas mas cervezas se consumen, mayor es el rango de felicidad. Es decir una variable aumenta al mismo tiempo que la otra

Cerveza <- seq(1, 10, by = 1)  
Felicidad <- Cerveza + rnorm(length(Cerveza), mean = 0, sd =1)  
  
Relacion\_lineal <- data.frame(Cerveza, Felicidad)  
ggplot(Relacion\_lineal, aes(x=Cerveza, y = Felicidad)) + geom\_point() + geom\_smooth(method ="lm", se= FALSE, color = "yellow") + labs(title = "Relación lineal", x="Consumo de cerveza", y= "Niveles de felicidad")

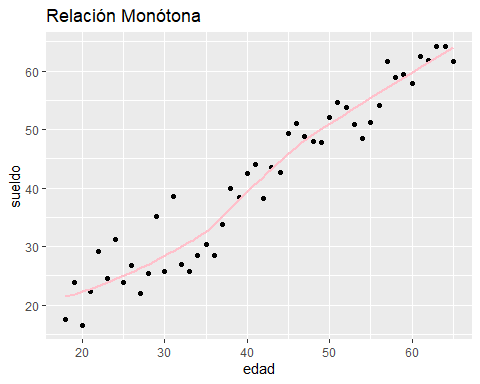
## `geom\_smooth()` using formula = 'y ~ x'



Relación monotona. Cuando aumenta la edad, es posible que su sueldo tambien aumente o disminuya, pero no en una tasa constante.

edad <- seq(18,65, by=1)  
sueldo <- edad + rnorm(length(edad), mean = 0, sd=5)  
  
relacionmontona <- data.frame(edad, sueldo)  
  
ggplot(relacionmontona, aes(x = edad, y= sueldo)) + geom\_point() + geom\_smooth(method = "loess", se = FALSE, color = "pink")+  
 labs(title = "Relación Monótona", x="edad", y = "sueldo")

## `geom\_smooth()` using formula = 'y ~ x'

 #(10) La correlación de Spearman se enfoca en la dirección y fuerza de la relación monotona entre las variables. El tipo de prueba Spearman es una prueba de correlación no paramétrica de la prueba de correlación de Pearson. Apropiada cuando la relación entre las variables no es lineal.

set.seed(123)  
estudio <- c(6, 7, 5, 8, 4)   
trabajo <- c(2, 3, 1, 5, 4)  
dataframecorr <-data.frame(estudio, trabajo)  
prueba\_correlacion <- cor.test(dataframecorr$estudio, dataframecorr$trabajo, method = "spearman")  
print(prueba\_correlacion)

##   
## Spearman's rank correlation rho  
##   
## data: dataframecorr$estudio and dataframecorr$trabajo  
## S = 12, p-value = 0.5167  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.4