1.1. ¿Cuál es la media, mediana y desviación estándar?, y la moda y los

valores repeticiones de la moda para los datos categóricos.

* X1:

Media:

n = 6

1+1+0+5+6+4/6 = 2.833

Mediana:

0 1 1 4 5 6 = (1+4)/2 = 2.5

Moda:

0 1 1 4 5 6 = 1 (dato que más se repite)

Desviación estándar:

0-2.833 = (-2.833) ^2 = 8.025

1-2.833 = (-1.833) ^2= 3.359

1-2.833 = (-1.833) ^2 = 3.359

4-2.833 = (1.167) ^2 = 1.361

5-2.833 = (2.167) ^2 = 4.695

6-2.833 = (3.167) ^2 = 10.029

√(Σ(xi - x)^2 / n) = √(30.831/6) = √5.138 = 2.266

* X2:

Media:

n = 6

0+1+2+3+4+4/6 = 2.333

Mediana:

0 1 2 3 4 4 = (2+3)/2 = 2.5

Moda:

0 1 2 3 4 4 = 4 (dato que más se repite)

Desviación estándar:

0-2.333 = (-2.333) ^2 = 5.442

1-2.333 = (-1.333) ^2 = 1.776

2-2.333 = (-0.333) ^2 = 0.110

3-2.333 = (0.667) ^2 = 0.444

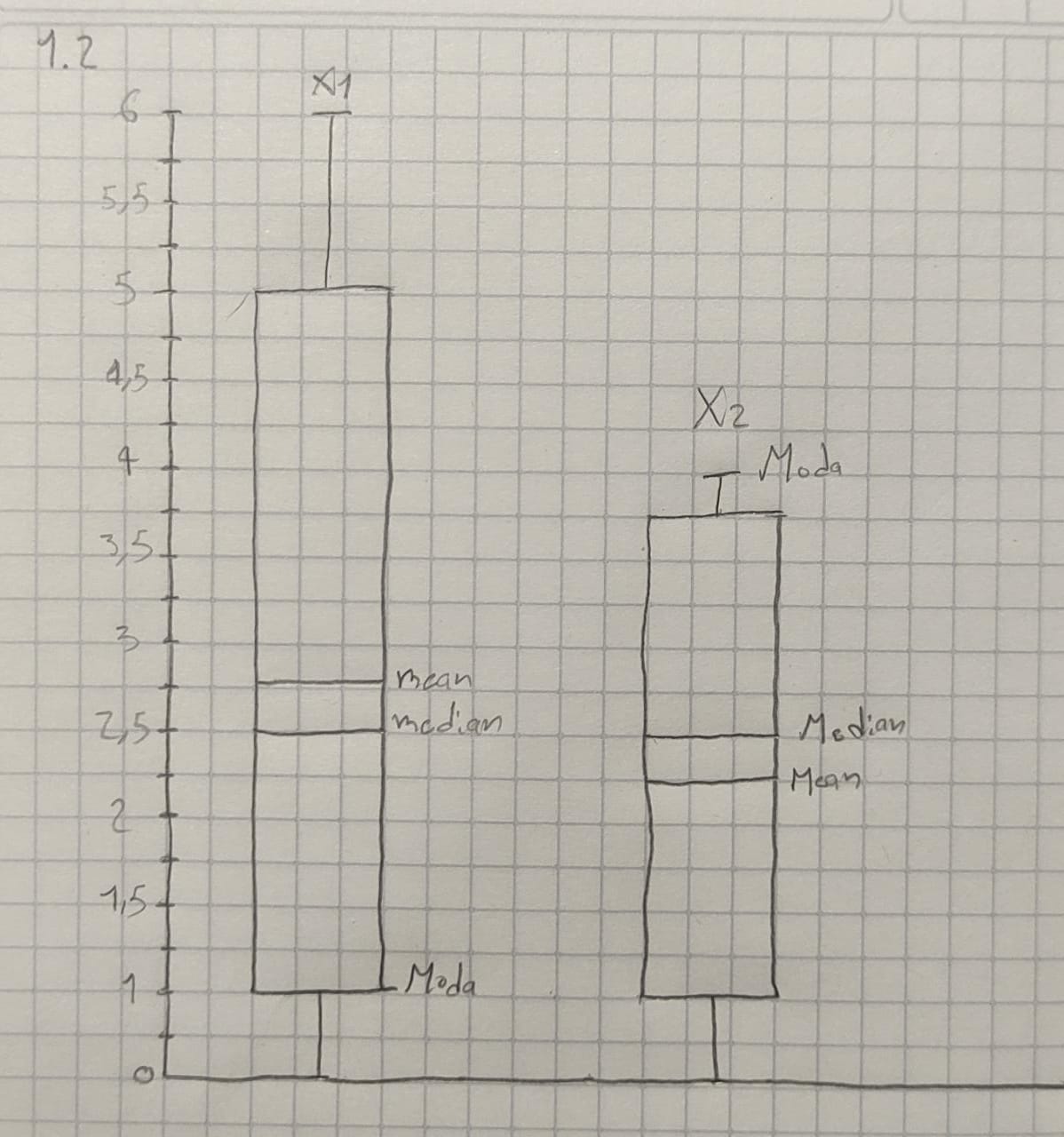
4-2.333 = (1.667) ^2 = 2.778

4-2.333 = (1.667) ^2 = 2.778

√(Σ(xi - x)^2 / n) = √(13.328/6) = √2.221 =1.490

1.2. Dibujar un boxplot a mano. Utilizando los datos de la tabla 1 y las

siguientes proporciones.



1.3. Cuál es la covarianza entre las 2 variables X1, X2.Imagen que contiene Texto

Descripción generada automáticamente

Media x1 = (0+1+1+4+5+6) / 6 = 2.833

Media x2 = (0+1+2+3+4+4) / 6 = 2.333

Desv x1 = {(-1.833), (-1.833), (-2.833), (2.166), (3.166), (1.166)}

Desv x2 = {(1.666), (0.666), (1.666), (-1.333), (-0.333) (-2.333)}

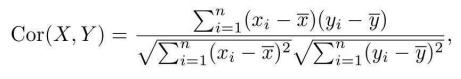
Desv mult = {(-1.833)(1.666), (-1.833)(0.666), (-2.833)(1.666), (2.166)(-1.333), (3.166)(-0.333), (1.166)(-2.333)}

= {-3.05, -1.22, -4.71, -2.88, -1.05, -2.72}

Suma\_desv\_mult = -3.05 + -1.22 + -4.71 + -2.88 + -1.05 + -2.72

Cov(x,y) = -15.63/ 6 = -2.605

1.4. Cuál es la correlación entre la variable x1 y x2 (Calcularla a mano).

Correlación puede ser escrita también como:

Σ(xi - x) (yi - y) = -15.63

√(Σ(xi - x)^2 ) = √(30.828) = 5.552

√(Σ(yi - y)^2) = √(13.328) = 3.650

= = = = -0.77

1.5. Explica la relación entre covarianza y correlación.

La relación entre la covarianza y la correlación es inversamente proporcional entre ambas, debido a que tanto la covarianza como la correlación son negativas.

1.6. Calcule el resultado del algoritmo K-means sobre este set de datos.

k=2 (2 clusters).

Centroides iniciales tomados: (1,1) y (5,3).

Asignación a clusters:

(0,0): Pertenece al primer cluster, ya que la distancia al centroide (1,1) es menor que la distancia al centroide (5,3)

(1,1): Pertenece al primer cluster, ya que la distancia al centroide (1,1) es menor que la distancia al centroide (5,3)

(1,2): Pertenece al primer cluster, ya que la distancia al centroide (1,1) es menor que la distancia al centroide (5,3)

(4,3): Pertenece al segundo cluster, ya que la distancia al centroide (5,3) es menor que la distancia al centroide (1,1)

(5,4): Pertenece al segundo cluster, ya que la distancia al centroide (5,3) es menor que la distancia al centroide (1,1)

(6,4): Pertenece al segundo cluster, ya que la distancia al centroide (5,3) es menor que la distancia al centroide (1,1)

Centroide del primer cluster: ((0+1+1)/3, (0+1+2)/3) = (0.67,1)

Centroide del segundo cluster: ((4+5+6)/3, (3+4+4)/3) = (5,3.67)

SEGUNDA ITERACION.

Asignación a clusters:

(0,0): Pertenece al primer cluster, ya que la distancia al centroide (0.67,1) es menor que la distancia al centroide (5,3.67)

(1,1): Pertenece al primer cluster, ya que la distancia al centroide (0.67,1) es menor que la distancia al centroide (5,3.67)

(1,2): Pertenece al primer cluster, ya que la distancia al centroide (0.67,1) es menor que la distancia al centroide (5,3.67)

(4,3): Pertenece al segundo cluster, ya que la distancia al centroide (5,3.67) es menor que la distancia al centroide (0.67,1)

(5,4): Pertenece al segundo cluster, ya que la distancia al centroide (5,3.67) es menor que la distancia al centroide (0.67,1)

(6,4): Pertenece al segundo cluster, ya que la distancia al centroide (5,3.67) es menor que la distancia al centroide (0.67,1)

Centroide del primer cluster: ((0+1+1)/3, (0+1+2)/3) = (0.67,1)

Centroide del segundo cluster: ((4+5+6)/3, (3+4+4)/3) = (5,3.67)