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, X2Imagen 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 = {(-2.833), (-1.833), (-1.833), (1.166), (2.166), (3.166)}

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

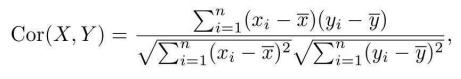
Desv mult = {(-2.833)(-2.333), (-1.833)(-1.333), (-1.833)(-0.333), (1.166)(0.666), (2.166)(1.666), (3.166)(1.666)}

= {6.611, 2.444, 0.611, 0.777, 3.611, 5.277}

Suma\_desv\_mult = 6.611 + 2.444 + 0.611 + 0.777 + 3.611 + 5.277 = 19.331

Cov(x,y) = 19.331/ 6 = 3.221

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

Correlación puede ser escrita también como:

Cov(x,y) = 19.331/ 6 = 3.221

√(Σ(xi - x)^2 ) = √(30.828) = 5.552

√(Σ(xi - x)^2) = √(13.328) = 3.650

= = = = 0.158

1.5. Explica la relación entre covarianza y correlación.

La relación entre la covarianza y la correlación es directamente proporcional.

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

Vamos a crear 2 grupos, es decir, k=2 (2 clusters).