SOLUCIÓN TAREA1

Melanie Giraldo C

1. ¿Cuál es la media, mediana y desviación estándar?, y la moda y los valores repeticiones de la moda para los datos categóricos.

|  |
| --- |
| MODA |
| 3 |
|  |
| MEDIA O PROMEDIO |
| 12.71428571 |
|  |
| MEDIANA |
| 4 |
|  |
| DESVIACION ESTANDAR |
| 16.81687415 |
|  |
|  |
|  |
|  |
| COVARIANZA |
| 3.571428571 |
|  |
|  |
| CORRELACIÓN |
| 0.9375 |

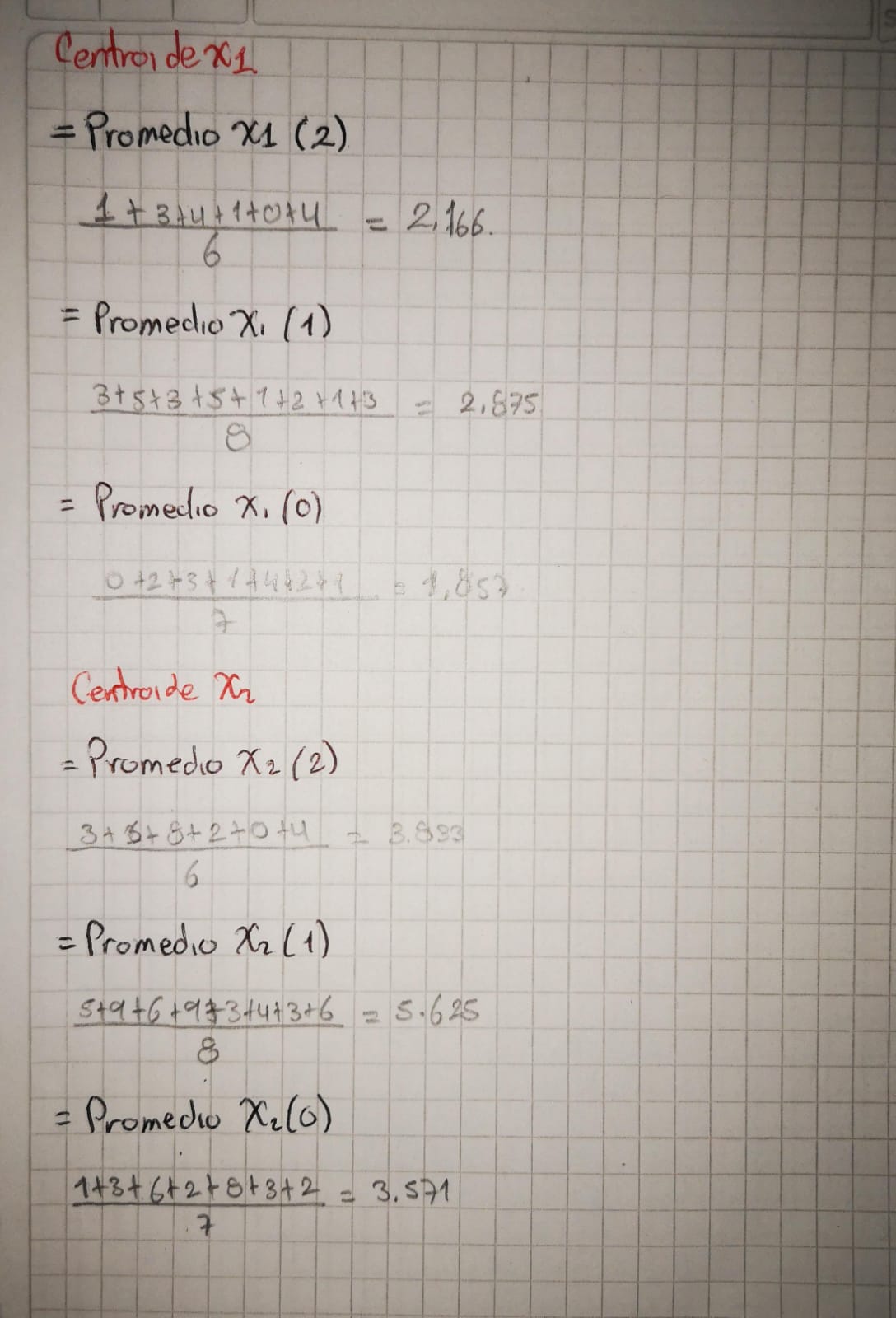
* 1. Explica la relación entre covarianza y correlación

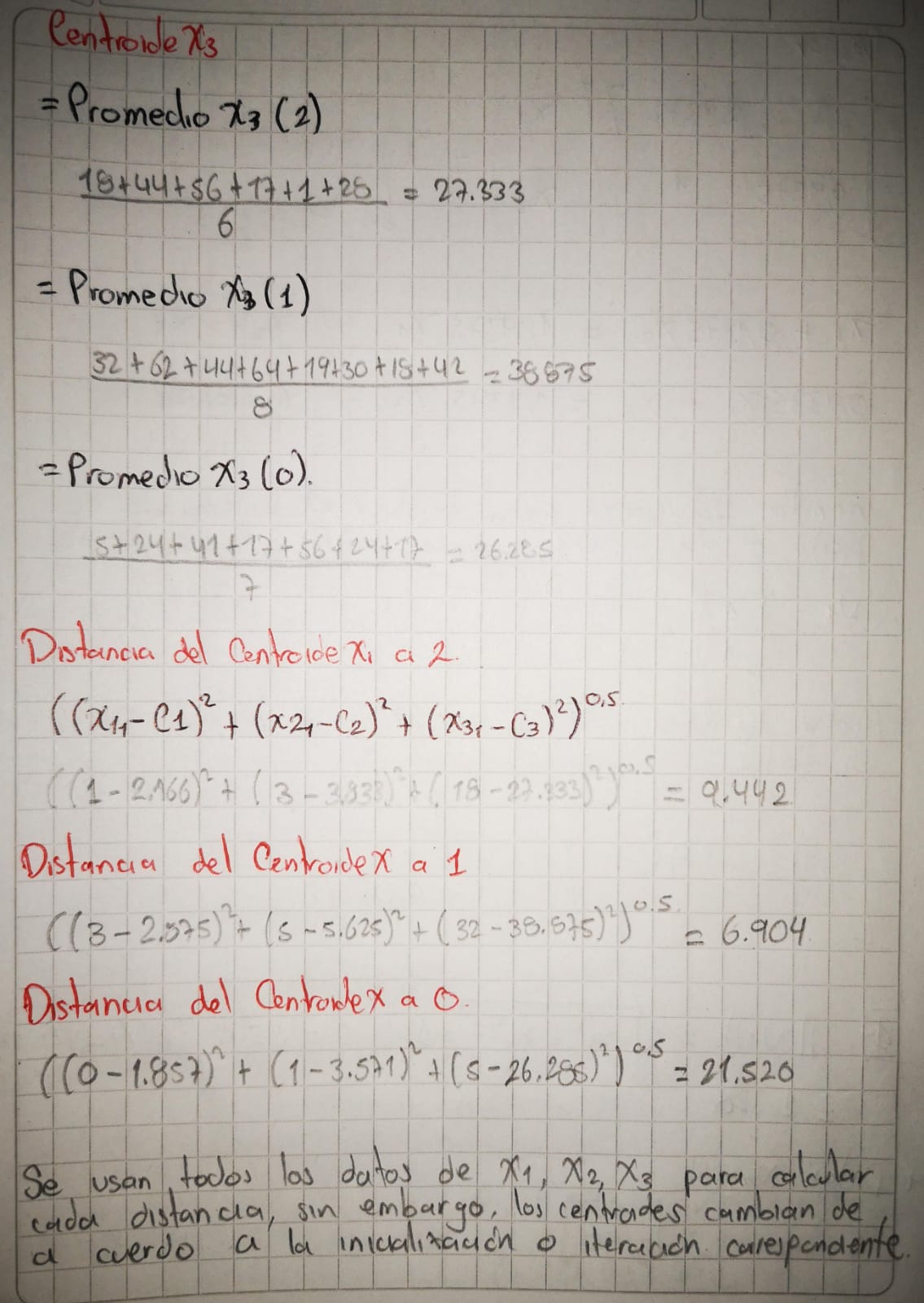
La relación entre la covarianza y la correlación es que en la primera hay una relación directa entre las variables es puesto que es positiva, es probable que si una cambia la otra lo haría en igual medida. En la segunda se aprecia que hay una relación lineal positiva fuerte entre las variables porque está cerca de 1; por lo anterior se puede concluir que entre X1 y X2 existe una relación lineal fuerte aparente.

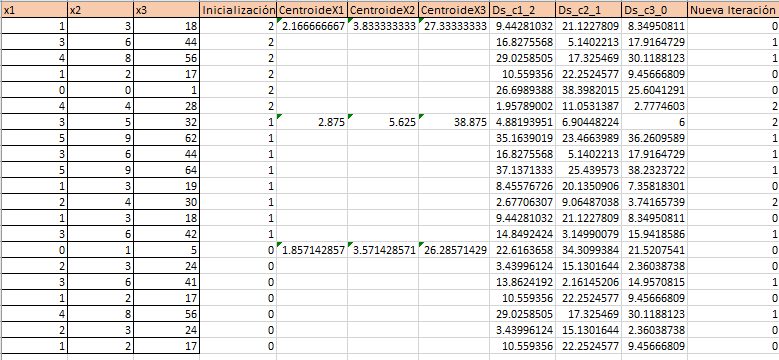
* 1. Kmean

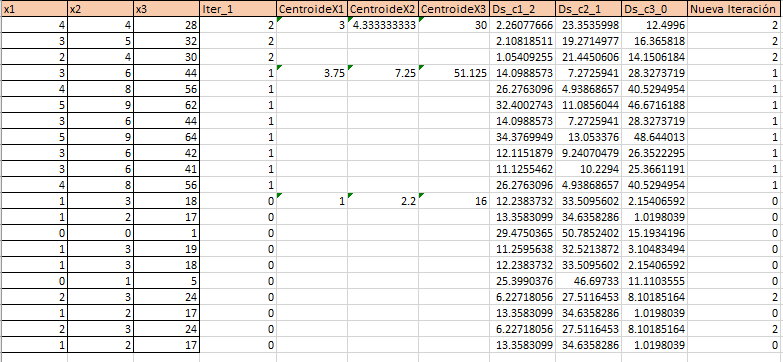
Se crean datos aleatorios entre 0 y 2 para la inicialización y se ordenan de mayor a menor.  
Se calculan los promedios de acuerdo al grupo de datos correspondientes y ya con estos y los datos de la tabla dada se calculan las distancias de acuerdo al grupo que correspondan

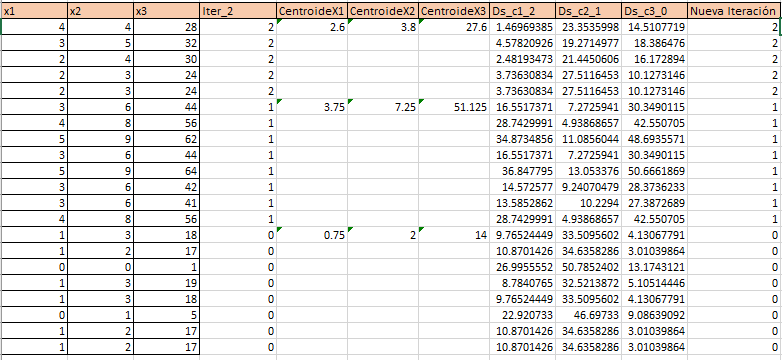
Por último, en las distancias se busca la distancia más corta de las 3 calculadas y donde encuentra la menor se pone el número del grupo en la nueva iteración, cuyos datos para la tabla siguiente serán los anteriormente llamados “inicialización”.











1. **2.1.**

De la tabla dada, se toman las variables x1 y x2

Se centra y normaliza:

Primeramente, se calculan los promedio y varianzas de las variables

= 2.33333333

= 4.42857143

VARIANZA:

,

2.22222222

6.53061224

* Al tener estos datos, dato por dato se normaliza para crear una nueva tabla:

NORMALIZACIÓN:

|  |  |
| --- | --- |
| 1.118033989 | -0.1677051 |
| 0.223606798 | -0.55901699 |
| 0.223606798 | -0.1677051 |
| 0.447213595 | 0.223606798 |
| 0.894427191 | -0.55901699 |
| 0.447213595 | 0.614918694 |
| 0.447213595 | 0.614918694 |
| 1.565247584 | -1.34164079 |
| 0.894427191 | -0.55901699 |
| 1.565247584 | -1.73295268 |
| 1.788854382 | 1.788854382 |
| 0.894427191 | -0.95032889 |
| 0.223606798 | -0.55901699 |
| 0.894427191 | -0.55901699 |
| 0.447213595 | 0.614918694 |
| 1.118033989 | 1.397542486 |
| 1.118033989 | 1.397542486 |
| 0.447213595 | 0.614918694 |
| 1.788854382 | 1.788854382 |
| 0.894427191 | -0.95032889 |
| 0.894427191 | -0.95032889 |

* Se debe hallar la matriz de covarianza y varianza de los datos normalizados:



* Para la covarianza se halla el promedio de los datos:

|  |  |  |
| --- | --- | --- |
| Promedio= | -1.16309 |  |

-9.5162

* Y se aplica la fórmula dando así:

|  |  |  |
| --- | --- | --- |
| Covarianza | 0.9375 | 0.9375 |

* Luego para la varianza se resta cada valor normalizado con el promedio al cuadrado y dividido la cantidad de datos:

|  |  |
| --- | --- |
| Calcular Varianza | |
| 0.05952381 | 0.00133929 |
| 0.00238095 | 0.01488095 |
| 0.00238095 | 0.00133929 |
| 0.00952381 | 0.00238095 |
| 0.03809524 | 0.01488095 |
| 0.00952381 | 0.01800595 |
| 0.00952381 | 0.01800595 |
| 0.11666667 | 0.08571429 |
| 0.03809524 | 0.01488095 |
| 0.11666667 | 0.14300595 |
| 0.15238095 | 0.15238095 |
| 0.03809524 | 0.04300595 |
| 0.00238095 | 0.01488095 |
| 0.03809524 | 0.01488095 |
| 0.00952381 | 0.01800595 |
| 0.05952381 | 0.09300595 |
| 0.05952381 | 0.09300595 |
| 0.00952381 | 0.01800595 |
| 0.15238095 | 0.15238095 |
| 0.03809524 | 0.04300595 |
| 0.03809524 | 0.04300595 |

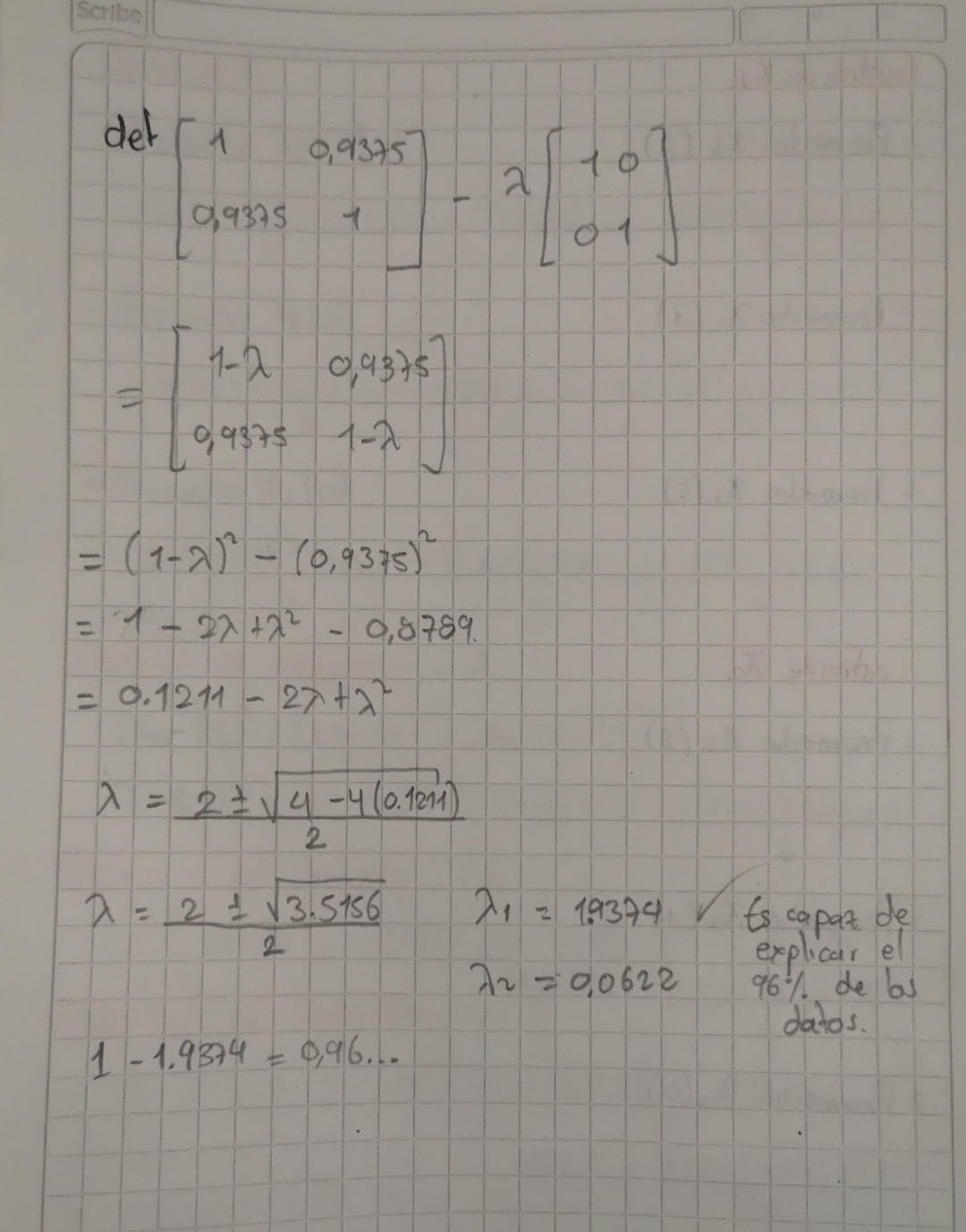
* Para hallar la varianza se hace una sumatoria de cada columna dando así:

|  |  |  |
| --- | --- | --- |
| Varianza | 1 , | 1 |

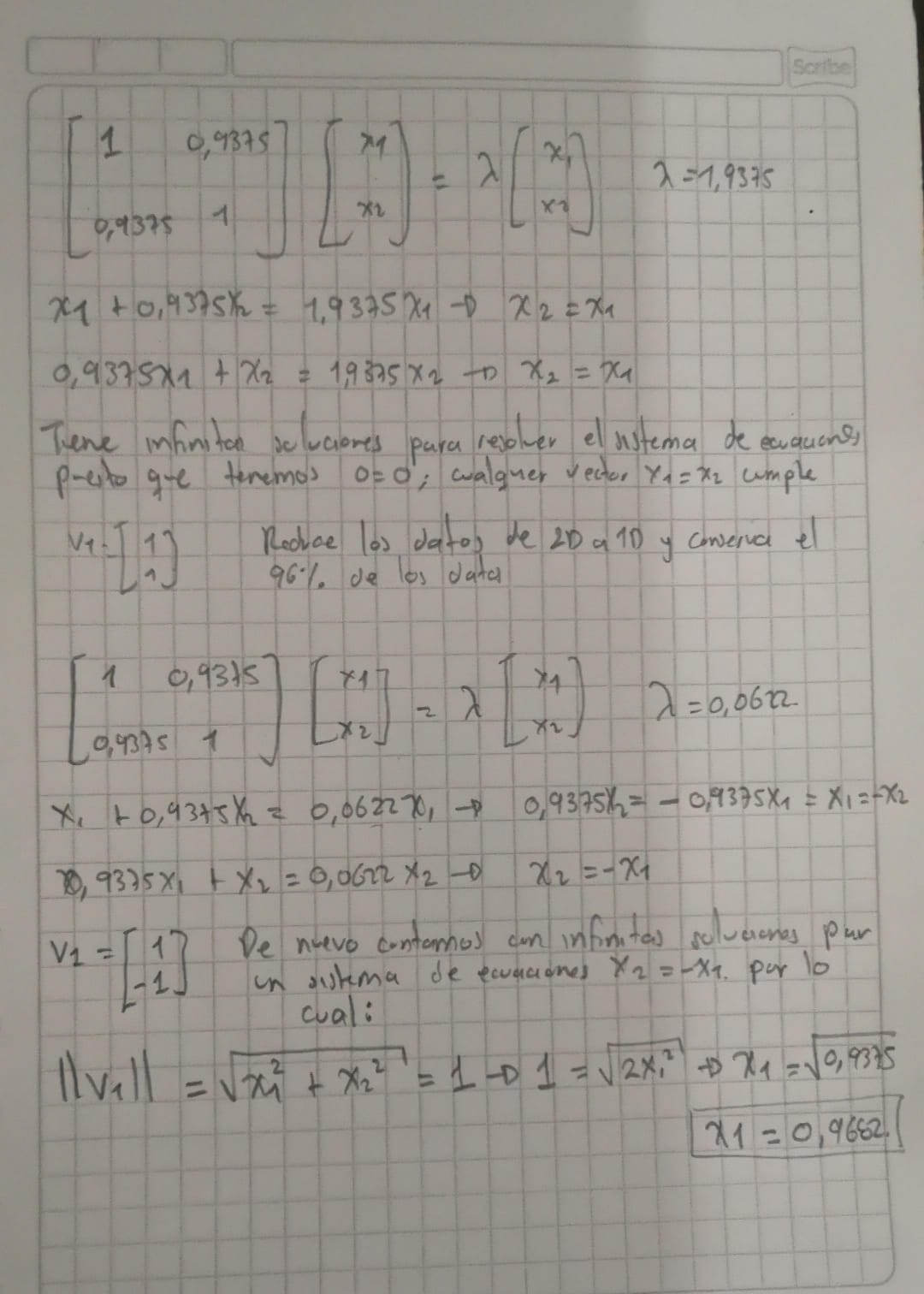
* Ya con estos valores tenemos que la matriz de covarianza y varianza es:

|  |  |
| --- | --- |
| M Varianza Covarianza | |
| 1 | 0.9375 |
| 0.9375 | 1 |

**2.1.** Eigenvalues



**2.2.** Eigenvector



**2.3.** Matriz Proyectada

* Con la tabla de las variables normalizadas y el vector encontrado se hace la multiplicación de matriz al vector [0.968 , 0.968] y generamos la proyección obteniendo:

|  |
| --- |
| Proyección |
| 0.91991837 |
| 0.75757983 |
| 0.37878992 |
| 0.64935414 |
| 1.40693397 |
| 1.02814406 |
| 1.02814406 |
| 2.81386794 |
| 1.40693397 |
| 3.19265786 |
| 3.46322208 |
| 1.78572389 |
| 0.75757983 |
| 1.40693397 |
| 1.02814406 |
| 2.43507803 |
| 2.43507803 |
| 1.02814406 |
| 3.46322208 |
| 1.78572389 |
| 1.78572389 |

**2.5.** Error

* Usando los datos de proyección y el vector se puede reconstruir la matriz y verificar junto con la normalizada la variación y pérdida de valores que se generaron. En todos los datos se ven cambios significativos, no fueron drásticos pero si variaron

|  |  |
| --- | --- |
| Reconstrucción | |
| 0.89048098 | 0.89048098 |
| 0.73333728 | 0.73333728 |
| 0.36666864 | 0.36666864 |
| 0.62857481 | 0.62857481 |
| 1.36191208 | 1.36191208 |
| 0.99524345 | 0.99524345 |
| 0.99524345 | 0.99524345 |
| 2.72382417 | 2.72382417 |
| 1.36191208 | 1.36191208 |
| 3.09049281 | 3.09049281 |
| 3.35239898 | 3.35239898 |
| 1.72858072 | 1.72858072 |
| 0.73333728 | 0.73333728 |
| 1.36191208 | 1.36191208 |
| 0.99524345 | 0.99524345 |
| 2.35715553 | 2.35715553 |
| 2.35715553 | 2.35715553 |
| 0.99524345 | 0.99524345 |
| 3.35239898 | 3.35239898 |
| 1.72858072 | 1.72858072 |
| 1.72858072 | 1.72858072 |

1. Dataset.

