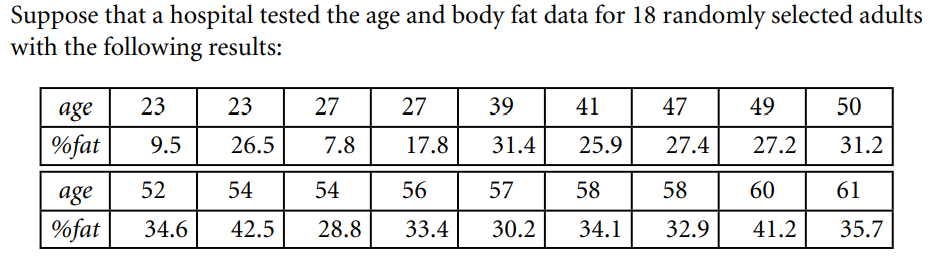
**Data Mining (CSE542)**

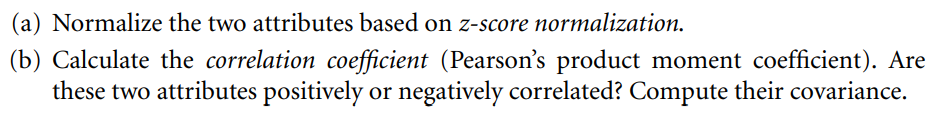
**Homework 02**

**ID: \_\_ Name: \_조원석\_ Date:\_\_2023-03-27\_\_**

**Task-1**

****

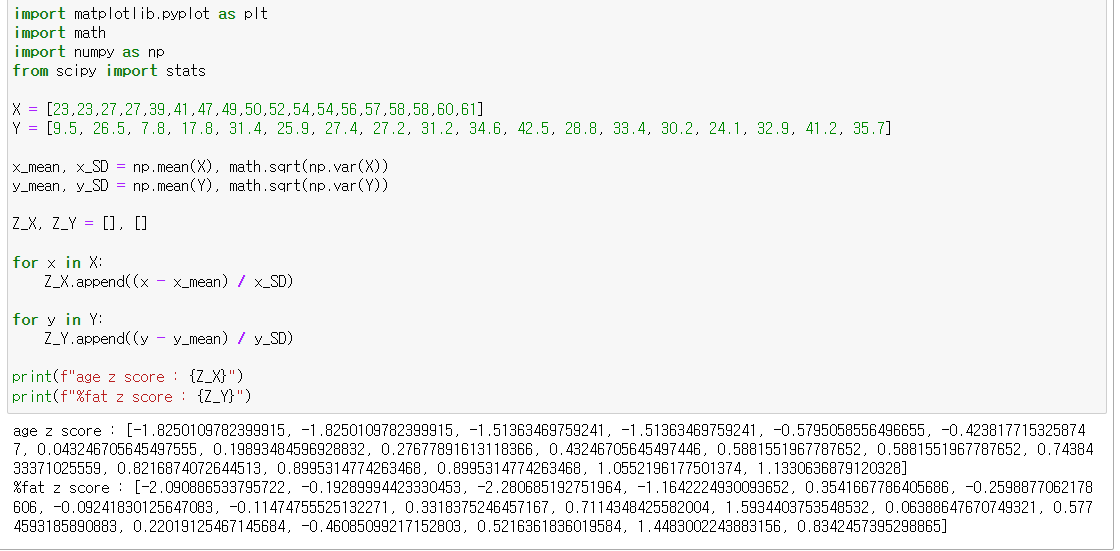
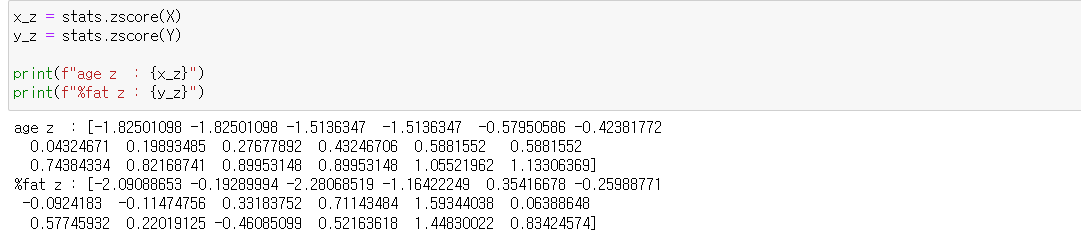
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| age | 23 | 23 | 27 | 27 | 39 | 41 | 47 | 49 | 50 |
| z-age | -1.83 | -1.83 | -1.51 | -1.51 | -0.58 | -0.42 | 0.04 | 0.20 | 0.28 |
| %fat | 9.5 | 2.65 | 7.8 | 17.8 | 31.4 | 25.9 | 27.4 | 27.2 | 31.2 |
| z-%fat | -2.09 | -0.19 | -2.28 | -1.16 | 0.35 | -0.26 | -0.09 | -0.11 | 0.33 |
| age | 52 | 54 | 54 | 56 | 57 | 58 | 58 | 60 | 61 |
| z-age | 0.43 | 0.59 | 0.59 | 0.74 | 0.82 | 0.90 | 0.90 | 1.06 | 1.13 |
| %fat | 34.6 | 42.5 | 28.8 | 33.4 | 30.2 | 34.1 | 32.9 | 41.2 | 35.7 |
| z-%fat | 0.71 | 1.59 | 0.06 | 0.58 | 0.22 | -0.46 | 0.52 | 1.45 | 0.83 |

****

**Ans:**

**Write your Answer here**

**(a) normalize the two attributes based on z-score normalization**

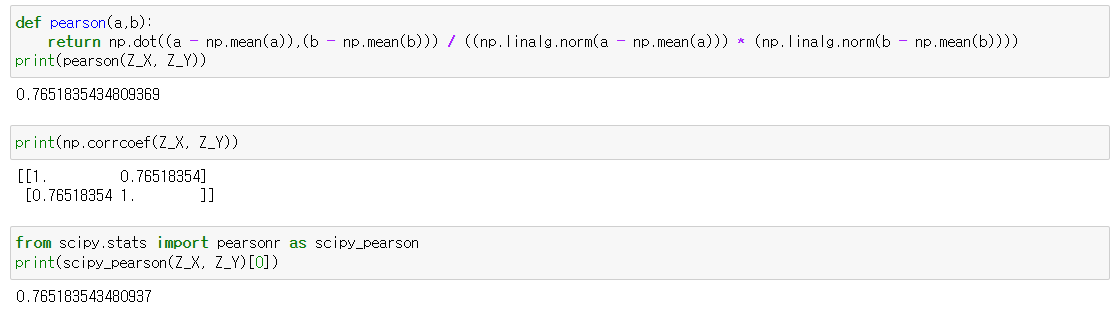
 

**age z score : [-1.8250109782399915, -1.8250109782399915, -1.51363469759241, -1.51363469759241, -0.5795058556496655, -0.4238177153258747, 0.043246705645497555, 0.19893484596928832, 0.27677891613118366, 0.43246705645497446, 0.5881551967787652, 0.5881551967787652, 0.7438433371025559, 0.8216874072644513, 0.8995314774263468, 0.8995314774263468, 1.0552196177501374, 1.1330636879120328]**

**%fat z score : [-2.090886533795722, -0.19289994423330453, -2.280685192751964, -1.1642224930093652, 0.3541667786405686, -0.2598877062178606, -0.09241830125647083, -0.11474755525132271, 0.3318375246457167, 0.7114348425582004, 1.5934403753548532, 0.06388647670749321, 0.5774593185890883, 0.22019125467145684, -0.46085099217152803, 0.5216361836019584, 1.4483002243883156, 0.8342457395298865]**

**(b)** **Calculate the correlation coefficient (Pearson's product moment coefficient). Are these two attributes positively or negatively correlated? Compute their covariance.**

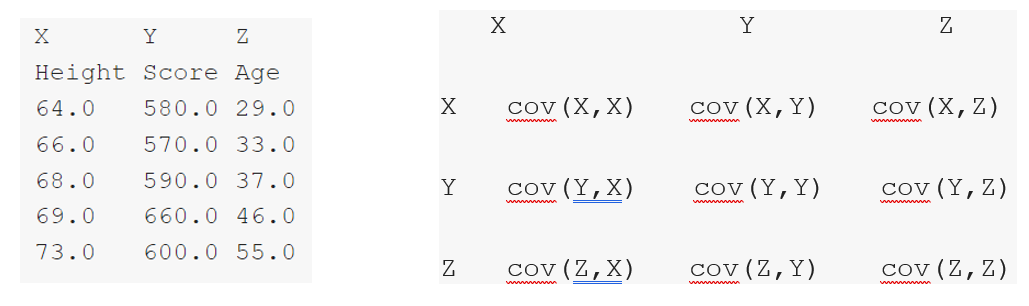
**Z\_X : age, Z\_Y : %fat**

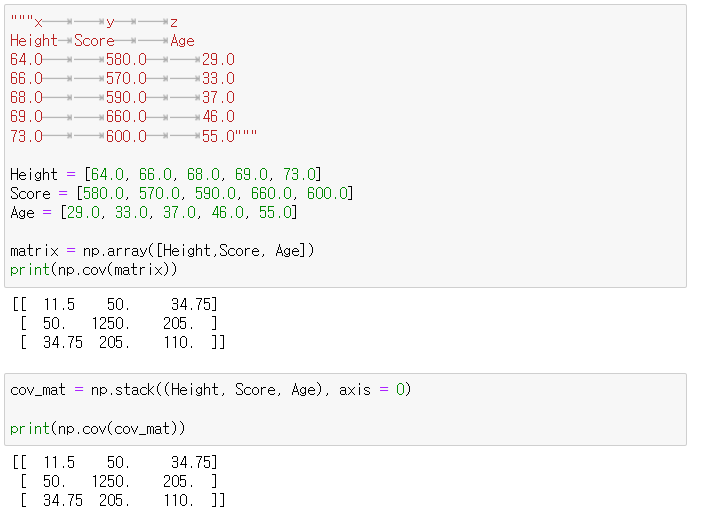
****

**Pearson’s product ≒ 0.765183543480937**

**So these two attributes have positively correlated**

**Task-2:** Values for Data item having three features Height (X), test Score (Y), and Age (Z). **Compute the covariance matrix**



****

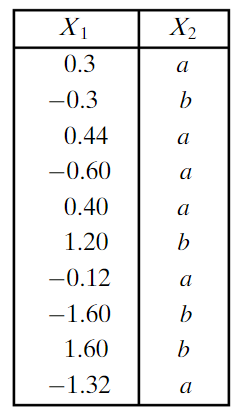
**[[ 11.5 50.0 34.75]**

**[ 50. 0 1250. 0 205.0 ]**

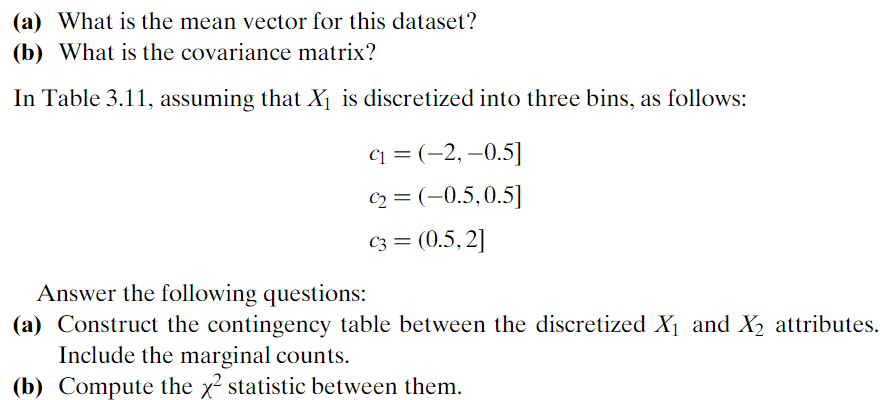
**[ 34.75 205.0 110. 0 ]]**

**Task-3: Consider the “mixed” data given in Table 3.11. Here X1 is a numeric attribute and X2 is a categorical one. Assume that the domain of X2 is given as dom(X2) = {a, b}.**

**Table 3.11.**

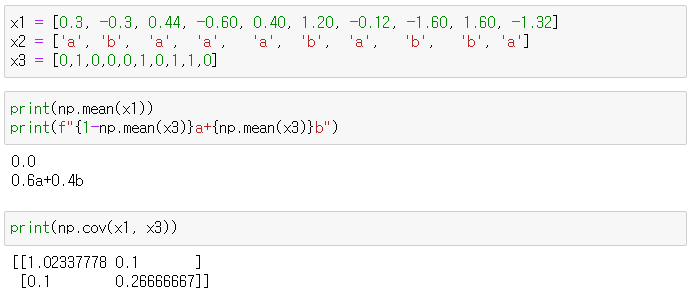
****

**Answer the following questions.**

****

**x1 = [0.3, -0.3, 0.44, -0.60, 0.40, 1.20, -0.12, -1.60, 1.60, -1.32]**

**x2 = [a,b,a,a,a,b,a,b,b,a]**

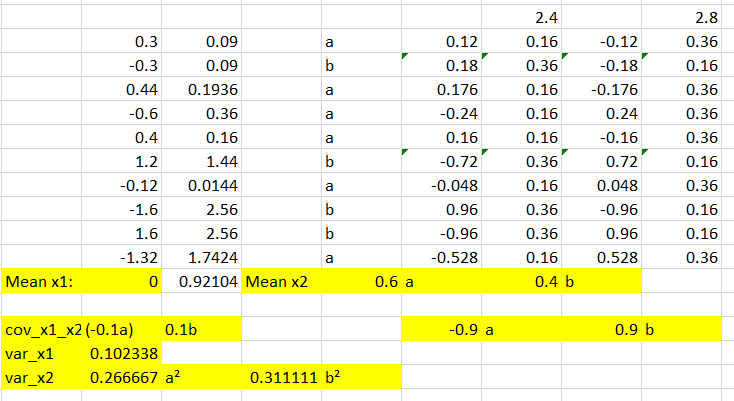
****

**(a) What is the mean vector for this dataset?**

**x1 = 0.0**

**x2 = 0.6a + 0.4b**

**(b) what is the covariance matrix?**

****

**cov\_x1\_x2 = sum((x1 - mean\_x1) \* (x2 - mean\_x2)) / (len(x1) – 1)**

**=> -0.1a+0.1b**

**var\_x1 = sum((x1 - mean\_x1)** **²) / (len(x1) - 1)**

**=>0.102338**

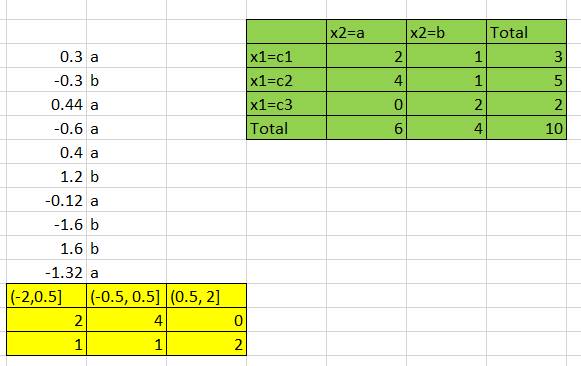
**var\_x2 = sum((x2 - mean\_x2)** **²) / (len(x2) - 1)**

**=> 0.266667a² + 0.311111b²**

**[[0.102338 -0.1a+0.1b ]**

**[-0.1a+0.1b 0.266667a² + 0.311111b²]]**

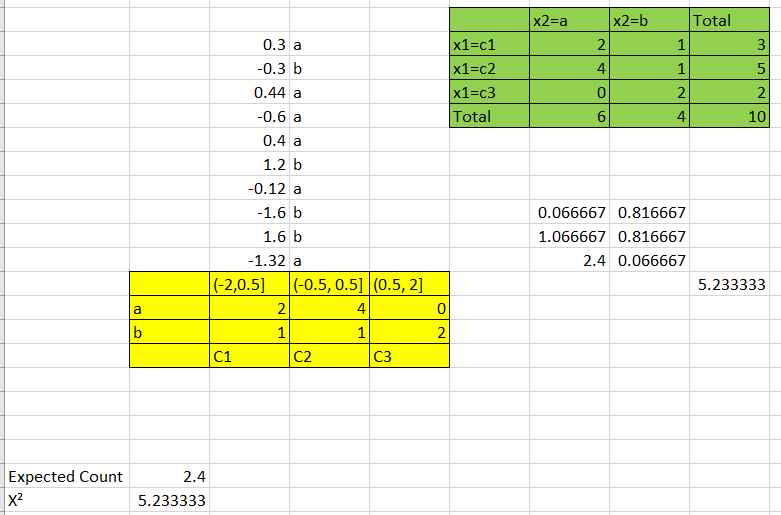
**(a) Construct the contingency table between the discretized X1 and X2 attributes. Include the marginal counts. ( Green table is answer )**

****

**(b) Compute the X² statistic between them. X² ≒5.233333**

**Expected count = (row total \* column total) / grand total**

**X² = sum((observed count - expected count)² / expected count)**

****

**Comment**

**It’s too hard to me.**

**I will study hard later..**

**Thank you for giving me a good assignment**