Lennart Qryo
Intelligent data analysis
DV1597

Exercise 2

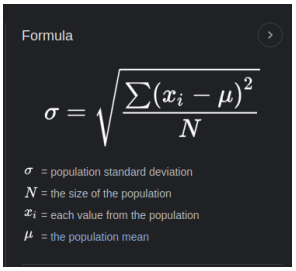
1. Suppose that P(A) = 0.3 and P(B) = 0.7.
    a. Can you compute P(A and B) if you only know P(A) and P(B)?
    No you can't compute P(A and B) if you only know P(A) and P(B) because we don't know if they are dependent or independent.

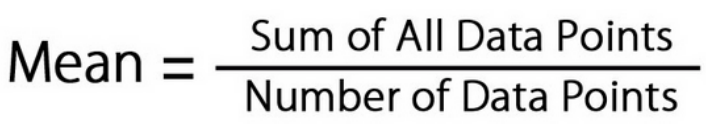
    b. Assuming that events A and B arise from independent random process:
        i. What is P(A and B)?
        P(A and B) = P(A)*P(B) = 0.3*0.7 = 0.21

        ii. What is P(A or B)?
        P(A or B) = P(A) + P(B) = 0.3+0.7 = 1

        iii. What is P(A I B)?
        Because they are independent then P(A I B) = P(A) = 0.3

2. Compare each pair of distributions to decide which one has greater mean and the greater standard deviation. (Links are in the pictures, click them!)



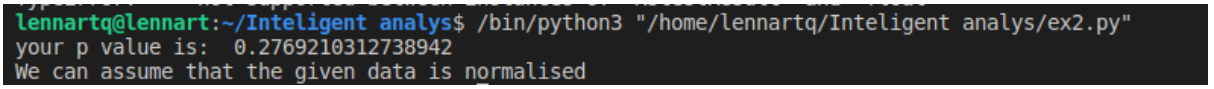


    a. First pair
        i. 1,7,7,7,9,12,12,1,14
        Mean: (1+7+7+7+9+12+12+1+14)/9 = 7,777777778
        Standard deviation: 4,34
        ii. 1,7,7,7,9,12,12,12,21
        Mean: (1+7+7+7+9+12+12+12+21)/9 = 9,777777778
        Standard deviation: 5,18
        Answear: (ii) or 1,7,7,7,9,12,12,12,21 has the biggest mean and standard deviation.

    b. Second pair
        i. -10,0,0,0,17,27,40,40
        Mean: (−10+0+0+0+17+27+40+40)/8 = 14,25
        Standard deviation: 18,35
        ii. -30,0,0,0,17,27,40,40
        Mean: (−30+0+0+0+17+27+40+40)/8 = 11,75
        Standard deviation: 22,39
        Answear: (i) or -10,0,0,0,17,27,40,40 has the greater meanwhile (ii) has the greater standard deviation

Lennart Qryo
Intelligent data analysis
DV1597

3. Is it suitable to assume normality in the given data?
   To check if the given data is normalised we could use the kolmogorov-smirnov test (Link from the lecture slides). If we use the scipy.stats.kstest module which is the kolmogorov smirnov test in python we get an p(probability) value and if it's over .05 then we can assume that the data is normalised otherwise it's not. The given data gives out an p(probability) value of .277
   ###LÄGGA IN KODEN KANSKE?###



   so we can assume that the given data is normalised.