1.

``` python

import pandas as pd

import numpy as np

df = pd.DataFrame({

'Year' : range(1985, 2007),

'Accident' : [4,2,4,3,11,6,4,4,1,4,2,3,3,1,2,2,6,0,2,1,3,2]

})

mean = df.Accident.sum()/df.shape[0]

median = df.Accident[df.shape[0]//2]

std = np.sqrt(((df.Accident-mean)\*\*2).sum()/(df.shape[0]-1)) # df.shape[0]-1 is the degree of freedom, unbiased estimator

print("sample mean of the accident rate: ", mean)

print("sample median of the accident rate: ", median)

print("sample standard deviation of the accident rate: ", std)

``` output

sample mean of the accident rate: 3.1818181818181817

sample median of the accident rate: 3

sample standard deviation of the accident rate: 2.3224856068314814

2.

We cannot conclude that the sample mean of the weights of the adults of town A is larger than the sample mean of the weights of the adults of town B based on the given information.

The reason is that the differences in the sample means between men and women within each town suggest that there may be gender differences in weight that are not accounted for. It is possible that the mean weight of women in town A is higher than the mean weight of men in town B, which could skew the overall mean weight of adults in each town.

In order to draw a valid conclusion about the difference in mean weights between the two towns, we would need to perform a statistical test that takes into account both gender and town as factors. This could involve conducting a two-way ANOVA or a regression analysis with both town and gender as independent variables, and weight as the dependent variable.

Furthermore, the difference in sample sizes between the adult women and men in each town could also have an impact on the ability to draw conclusions about the mean weights of adults in each town. If, for example, the sample of adult women in town A was much larger than the sample of adult men in town A, then the overall sample mean for town A could be influenced more by the weights of the women than the weights of the men. In such a case, we would need to carefully consider the sample sizes and their representativeness before making any conclusions about the mean weights of adults in each town.

3.

``` python

import pandas as pd

import numpy as np

df = pd.DataFrame({

'Height' : [64,65,66,67,69,70,72,72,74,74,75,76],

'Salary' : [91,94,88,103,77,96,105,88,122,102,90,114]

})

corr = df['Height'].corr(df['Salary'])

print("correlation with series.corr()", corr)

numerator = (df['Height'] - df['Height'].mean()) \* (df['Salary'] - df['Salary'].mean())

denominator = np.sqrt(((df['Height'] - df['Height'].mean())\*\*2).sum() \

\* ((df['Salary'] - df['Salary'].mean())\*\*2).sum())

corr = numerator.sum() / denominator

print("correlation with definition", corr)

```output

correlation with series.corr() 0.48384587052918976

correlation with definition 0.4838458705291898

4.

let Di be the event that ratio i is defective (D1 and D2 are independent)

P(D2 | D1)

= P(D1D2) / P(D1) (D2 when D1 is true)

= P(D1D2|A)P(A) + P(D1D2|B)P(B) / P(D1|A)P(A) + P(D1|B)P(B)

= .05^2(½) + .01^2(½) / .05(½) + .01(½)

5.

A = A is executed, P(A) = ⅓

B = B is set free, P(B) = ⅓

C = C is set free, P(C) = ⅓

P(A|B)

= P(B|A) \* P(A) / P(B)

= ½ \* ⅓ / ⅓ = ½

6.

E = p\* [1 - (1-p)^2] + (1-p\*) [1-p^2]

when dE/dp = 0, the expected score will be the maximum.

p = p\*

7.

(a)

E[(2+4X)^2]

= E[4+16X+16X^2]

= 4 + 16\*2 + 16\*8

= 164

(b)

E[X^2+(X+1)^2]

=E[2X^2+2X+1]

= 21

8.

If adapt to Cov(a,b) = E[ab) - E(a)E(b),

Cov(X1-X2, X1+X2) = 0

= E[(X1-X2)(X1+X2)] - E[(X1-X2)]E[(X1+X2)]

= E[X1^2-X2^2] - E[X1^2-X2^2]

= 0

9.

P(X<=x)

= P(Y <= (x-a)/b)

= P((a+by) <= x)

(a)

E[X] = a + bE[Y]

(b)

Var(X) = b^2Var(Y)

10.

(a)

P(X1-10 / σ > 15 - 10 / σ)

= P(Z > 5/σ)

P(X1+X2 > 25)

= P(X1+X2-20 / σ > 10 / σ)

= P(Z > 5 / σ)

(b)

same as (a)

(c)

P(x-20 / σ) = P(5 / σ) → x = 5+20

11.

X is normal distribution with mean a, variance b^2