CO3 – Skilling questions(skill 7)

1. Questions on variance of sum of two random variables.

1.A. Consider the following data of 10 people of two different countries USA and India. In this data Bodytemp\_usa represents body temperature, Height\_usa represents height of a person in feet.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sl. No | Name | SSN NO | Country | Bodytemp\_usa | Height\_usa(feet) |
| 1 | x | 12us01 | USA | 97.2 | 6.4 |
| 2 | y | 12us02 | USA | 97.4 | 6.5 |
| 3 | z | 12us03 | USA | 98.5 | 6.2 |
| 4 | p | 12us04 | USA | 98.6 | 5.9 |
| 5 | q | 12us05 | USA | 97.8 | 5.8 |
| 6 | r | 12us06 | USA | 97.6 | 5.11 |
| 7 | s | 12us07 | USA | 98.6 | 6.3 |
| 8 | t | 12us08 | USA | 98.9 | 6.5 |
| 9 | u | 12us09 | USA | 99.2 | 6.7 |
| 10 | v | 12us10 | USA | 99.2 | 6.6 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sl. No | Name | SSN NO | Country | Bodytemp\_india | Height\_ind(Feet) |
| 1 | A | 15ind01 | IND | 98.4 | 5.8 |
| 2 | B | 15ind02 | IND | 98.4 | 5.7 |
| 3 | C | 15ind03 | IND | 97.8 | 5.5 |
| 4 | D | 15ind04 | IND | 98.6 | 5.4 |
| 5 | E | 15ind05 | IND | 97.8 | 5.6 |
| 6 | F | 15ind06 | IND | 98.6 | 6.2 |
| 7 | G | 15ind07 | IND | 98.6 | 6.1 |
| 8 | H | 15ind08 | IND | 98.8 | 5.10 |
| 9 | I | 15ind09 | IND | 99.6 | 5.6 |
| 10 | J | 15ind10 | IND | 98.8 | 5.4 |

Q1. Write a program in python to find the variance of body temperature of both the countries. Var(Bodytemp\_india), var(Bodytemp\_usa) ?

Q2. Write a program in python to find the variance of sum of body temperature of both the countries. Find var( Bodytemp\_india + Bodytemp\_usa).

Q3. Write a program in python to find the variance of height of both the countries. Var(Height\_usa), var(Height\_int) ?

Q4. Write a program in python to find the variance of sum of height of both the countries. Find var( Height\_ind + Height\_usa).

1.B.

Data-set

The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by the British statistician, eugenicist, and biologist Ronald Fisher in his 1936. The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimetres. Based on the combination of these four features, Fisher developed a linear dis-criminant model to distinguish the species from each other. The dataset is available in following link.

<https://www.kaggle.com/arshid/iris-flower-dataset>

Q1. Write a program in python to find the variance of ‘SepalLengthCm’ and ‘SepalWidthCm’ . Var(SepalLengthCm), var(SepalWidthCm) ?

Q2. Write a program in python to find the variance of sum of ‘SepalLengthCm’ and ‘SepalLengthCm’. Find variance(SepalLengthCm + SepalWidthCm).