Q-1. **A university wants to understand the relationship between the SAT scores of its applicants and their college GPA. They collect data on 500 students, including their SAT scores (out of 1600) and their college GPA (on a 4.0 scale). They find that the correlation coefficient between SAT scores and college GPA is 0.7. What does this correlation coefficient indicate about the relationship between SAT scores and college GPA?**

**Solution -** The correlation coefficient is a statistical measure of the strength of a linear relationship between two variables. Its values can range from -1 to 1. A correlation coefficient of -1 describes a perfect negative, or inverse, correlation, with values in one series rising as those in the other decline, and vice versa. A coefficient of 1 shows a perfect positive correlation, or a direct relationship. A correlation coefficient of 0 means there is no linear relationship.

As the correlation coefficient between SAT score and GPA is 0.7 which is close to 1. It thus indicates that there is a direct relationship between SAT score and GPA and hence students with higher GPA and will most likely have high SAT scores and vice versa.

**Q-2. Consider a dataset containing the heights (in centimeters) of 1000 individuals. The mean height is 170 cm with a standard deviation of 10 cm. The dataset is approximately normally distributed, and its skewness is approximately zero. Based on this information, answer the following questions:**

* **What percentage of individuals in the dataset have heights between 160 cm and 180 cm?**

**Solution –** By empirical formula 68% of individuals will lie in between 160cm and 180cm.

* **If we randomly select 100 individuals from the dataset, what is the probability that their average height is greater than 175 cm?**

**Solution –** 1 – 0.6915 = 0.3085. Thus, probability will be 0.3085

* **Assuming the dataset follows a normal distribution, what is the z-score corresponding to a height of 185 cm?**

**Solution –** Zscore = 1.5

* **We know that 5% of the dataset has heights below a certain value. What is the approximate height corresponding to this threshold?**

**Solution –** 186.5 cms

* **Calculate the coefficient of variation (CV) for the dataset.**

**Solution –** 0.0588

* **Calculate the skewness of the dataset and interpret the result.**

**Solution –** As per our assumption skewness is zero which means mean and median being the same.

**Q-3. Consider the ‘Blood Pressure Before’ and ‘Blood Pressure After’ columns from the data and calculate the following**

[**https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share\_**](https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share_)

* **Measure the dispersion in both and interpret the results.**
* **Calculate mean and 5% confidence interval and plot it in a graph**
* **Calculate the Mean absolute deviation and Standard deviation and interpret the results.**
* **Calculate the correlation coefficient and check the significance of it at 1% level of significance.**

**Q-4. A group of 20 friends decide to play a game in which they each write a number between 1 and 20 on a slip of paper and put it into a hat. They then draw one slip of paper at random. What is the probability that the number on the slip of paper is a perfect square (i.e., 1, 4, 9, or 16)?**

**Solution –** 0.2

**Q-5. A certain city has two taxi companies: Company A has 80% of the taxis and Company B has 20% of the taxis. Company A's taxis have a 95% success rate for picking up passengers on time, while Company B's taxis have a 90% success rate. If a randomly selected taxi is late, what is the probability that it belongs to Company A?**

**Solution –** 0.667

**Q-7. The equations of two lines of regression, obtained in a correlation analysis between variables X and Y are as follows: and . 2𝑋 + 3 − 8 = 0 2𝑌 + 𝑋 − 5 = 0 The variance of 𝑋 = 4 Find the**

* **Variance of Y**

**Solution –** var(Y) = 1

* **Coefficient of determination of C and Y**

**Solution –** Since there is no coefficient of Y in first equation coefficient of determination will be 0

* **Standard error of estimate of X on Y and of Y on X.**

**Solution –**

**Q-10. A factory produces light bulbs, and the probability of a bulb being defective is 0.05. The factory produces a large batch of 500 light bulbs.**

* **What is the probability that exactly 20 bulbs are defective?**

**Solution – (500)C(20)x(0.05)^20x(0.95)^480**

* **What is the probability that at least 10 bulbs are defective?**

**Solution =** 1 - [(500)C(9)x(0.05)^9x(0.95)^491 +(500)C(8)x(0.05)^8x(0.95)^492 + (500)C(7)x(0.05)^7x(0.95)^493 + (500)C(6)x(0.05)^6x(0.95)^494 + (500)C(5)x(0.05)^5x(0.95)^495 + (500)C(4)x(0.05)^4x(0.95)^496 + (500)C(3)x(0.05)^3x(0.95)^497 + (500)C(2)x(0.05)^2x(0.95)^498 + (500)C(1)x(0.05)^1x(0.95)^499 + (500)C(0)x(0.05)^0x(0.95)^500]

* **What is the probability that at max 15 bulbs are defective?**

**Solution – [**500)C(15)x(0.05)^15x(0.95)^485  **+** 500)C(14)x(0.05)^14x(0.95)^486 **+** 500)C(13)x(0.05)^13x(0.95)^487  **+** 500)C(12)x(0.05)^12x(0.95)^488 **+** 500)C(11)x(0.05)^11x(0.95)^489 **+** 500)C(10)x(0.05)^10x(0.95)^490 **+** 500)C(9)x(0.05)^9x(0.95)^491 +(500)C(8)x(0.05)^8x(0.95)^492 + (500)C(7)x(0.05)^7x(0.95)^493 + (500)C(6)x(0.05)^6x(0.95)^494 + (500)C(5)x(0.05)^5x(0.95)^495 + (500)C(4)x(0.05)^4x(0.95)^496 + (500)C(3)x(0.05)^3x(0.95)^497 + (500)C(2)x(0.05)^2x(0.95)^498 + (500)C(1)x(0.05)^1x(0.95)^499 + (500)C(0)x(0.05)^0x(0.95)^500]

* **On average, how many defective bulbs would you expect in a batch of 500?**

**Solution –** 500 x 0.05 = 25