CS215 Assignment 1 Solutions

Satyam Sinoliya, 23B0958
Vaibhav Singh, 23B1068
Shaik Awez Mehtab, 23B1080

August 2024

Question 1. There are two friends playing a dice-roll game. Friend A has (n + 1) fair dice and Friend B has n fair dice (a fair die has equal probability of every face). On every roll, a win is achieved if we get a prime number on the top. What is the probability that A will have more wins than B if both roll all of their dice?

[5 marks]

Solution 1. So A or B wins when they get either $2, 3$ or 5, whose probability is $\frac{1}{2}$. So the probability that A wins i number of times when all of the $(n+1)$ dice are rolled is

$$p(A_i) = \binom{n+1}{i}\left(\frac{1}{2}\right)^{i}\left(\frac{1}{2}\right)^{n+1-i} \tag{1}$$

$$= \binom{n+1}{i}\left(\frac{1}{2}\right)^{n+1} \tag{2}$$

Similarly, probability that B gets i wins when all of the n are rolled is

$$p(B_i) = \binom{n}{i}\left(\frac{1}{2}\right)^{i}\left(\frac{1}{2}\right)^{n-i} \tag{3}$$

$$= \binom{n}{i}\left(\frac{1}{2}\right)^{n} \tag{4}$$

Now the probability that A has more wins than B is

$$p(i > j) = \sum_{i=0}^{n+1}\sum_{\substack{j=0\\ i>j}}^{n} p(A_i)p(B_j) \tag{5}$$

Since number of wins of A is independent from that of B. By writing each possible term $\binom{n+1}{i}$ and $\binom{n}{j}$ in a grid manner. We observe that our required terms cover exactly half of the grid. Thus equation (5) simplifies to

$$p(i > j) = \frac{\sum_{i=0}^{n+1}\binom{n+1}{i}\left(\frac{1}{2}\right)^{n+1}\cdot\sum_{j=0}^{n}\binom{n}{j}\left(\frac{1}{2}\right)^{n}}{2} = \frac{1}{2} \tag{6}$$

Thus, the probability of A winning more times than B is $\frac{1}{2}$.

Question 2. Read about the following plots:

1. Violin Plot
2. Pareto Chart
3. Coxcomb Chart
4. Waterfall Plot

Describe the uses of these plots. Take some sample data and generate one example plot for each of them.

[8 marks]

Solution 2. Here are the descriptions and usages of the given plots along with their examples:

1. **Violin plot:** It's a hybrid of box plot and kernel density plot. A box plot represents data in a linear fashion. It's made of a straight line from lowest value to highest value along with a box from first to third quartiles, marking all the quartiles of the dataset. Here's an example:

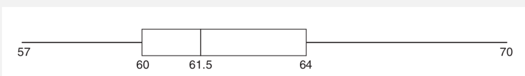

Figure 1: Box plot

A kernel density plot represents the density/frequency of data points. It's similar to a histogram, but smooth. In a violin graph, this is kept vertical, with two mirror images of it reflected along y-axis. This is useful in the sense that we can look into both centrel tendencies of the data (like mean, median etc.) but also how the data is distributed. Both at once. We can visualise the following data which I've taking from (*) containing the average number of hours a person studies given the number of courses taken

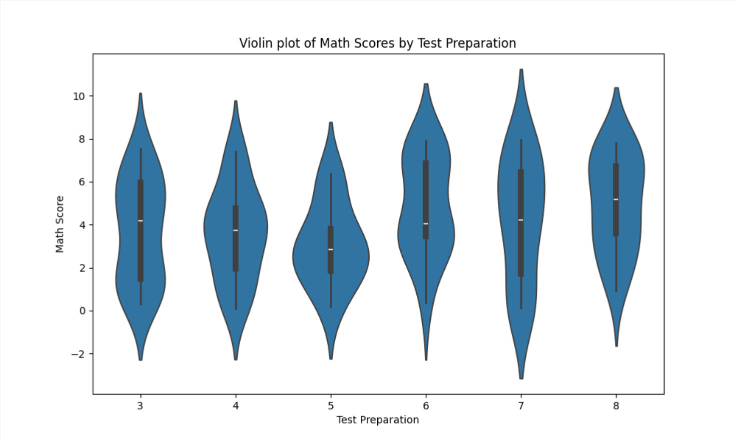

Figure 2: Violin Plot

As mentioned before, a violing plot can show both statistical summary along with distribution, which a normal plot can't. Here, the gray line represents the box plot component of it. And the plot you get by rotating it by 90° is the distribution plot.