

Introduction:

The fishing data given in <http://edshare.soton.ac.uk/19466/> represents the data of a fisherman catching fishes of various sizes throughout the day using 3 different type of baits (A,B, and C). The dataset has 3 features or columns which are the Catch Time, Catch Size, and the Bait Type. I have performed a statistical analysis on the dataset and provided relevant findings in the report.

Abstract:

The statistical analysis is used to understand the inter-relation between the 3 columns provided in the dataset. In this short report, I have presented the effectiveness of different types of bait, and it's relationship with time, size, and frequency of the catch. Measures of centrality and spread are also presented. Along with that, I have also analyzed the the total size of catch on an hourly basis and for each bait type. For understanding purpose I have termed the sum of sizes as "mass" in the report.

The fig.1 below represents a histogram of Time(X-Axis) vs Catch Frequency(Y-axis). We can see from the plot that hours 11-12 am yield the most number of catches. So fishing at those hours will yield more catches, where as fishing in hours 9-10pm are shown to be the most unproductive. The hours past 10pm yield no catches at all.

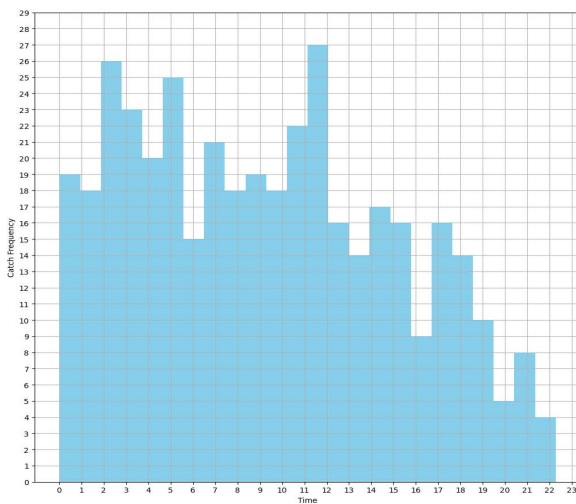


fig: 1 Histogram of Time vs Catch Frequency

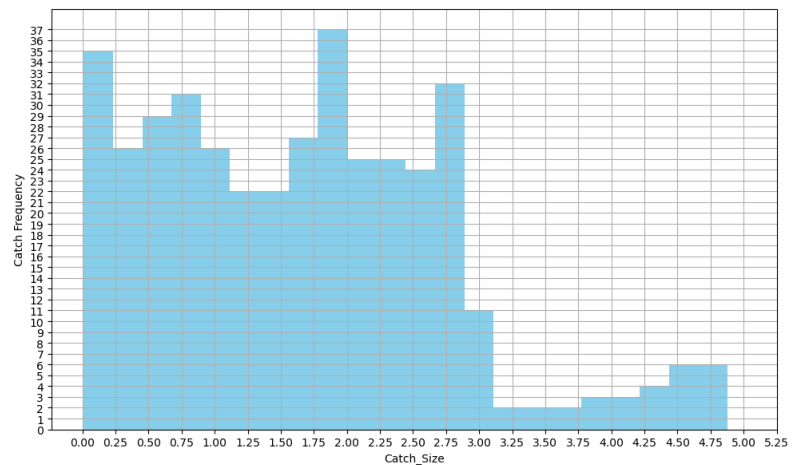


fig: 2 Histogram of Catch Size vs Catch Frequency

fig.2 represents a histogram of Catch size(X-axis) vs Catch Frequency(Y-axis). We can see from the plot that catches of size 1.75 to 2 are the most common. Catches of size 3.125 to 3.75 are the least common. Bin size being used is 24 considering the nature of the time sensitive data. Each bin corresponds to hourly data.

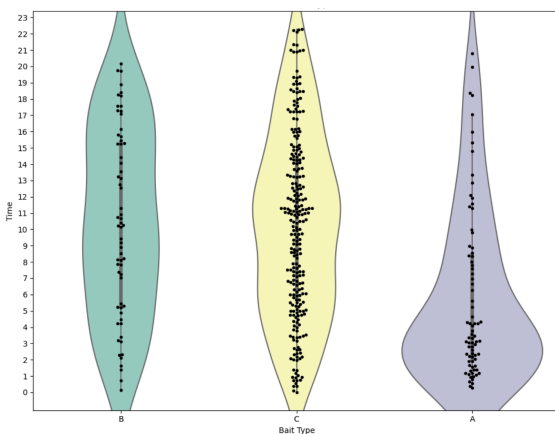


fig: 3 Violin and Swarm plot for Time of Catch for B,C,A

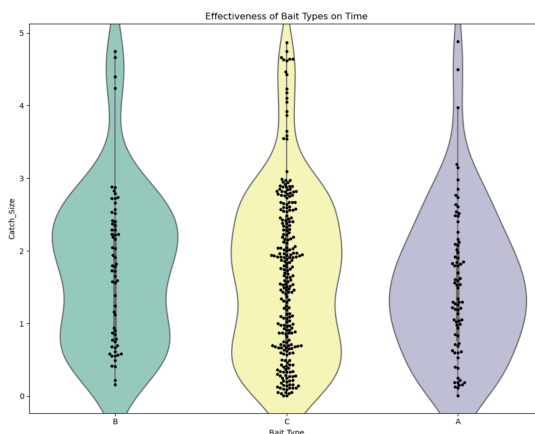


fig: 4 Violin and Swarm plot for Size of Catch for B,C,A

The Violin Plot, in fig.3 represents a Bait Type vs Time plot where we have also a swarm plot on top of it representing the density of that bait-type used every hour, also indirectly representing the density of catches made with a bait type for the hour.

Fig 3 shows that the Bait type 'C'(yellow) yields the most number of catches. Also, the hour 11-12 are the widest as indicated by the breadth of the violin as well as the swarm plot. So we can conclude that the hours 11-12 yield the most number of catches for Bait 'C'.

Fig 3 also shows that in the late hours of 21-22 Bait 'B' and 'A' yield no catches, only bait 'C' yields catches in those hours.

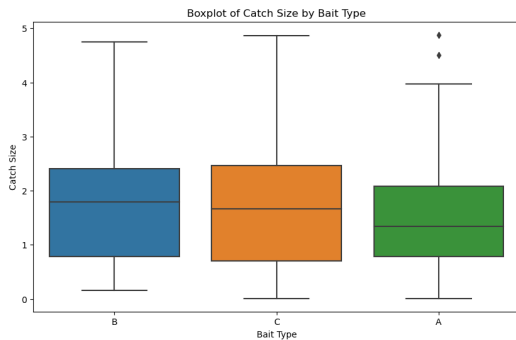


Fig 4 has a swarmplot which clearly shows that Bait 'C' yields most catches in the size range 2-3.

Fig 5 shows that the Bait type 'A' is the one with the largest catch size with the 2 extending dots. This seems un-intuitive as Bait Type 'C' is used the most as shown in figure 3. Also, even though bait 'C' is the most frequently used, bait 'B' yields the bigger catches on average, this is also shown by the mean line of Bait B which is a little above C's.

fig: 5 - Box plot for Size of Catch

Bait_Type	Mean	Median	Standard Deviation	Minimum	Maximum	Q1 \	Q3
A	5.849747	3.57	5.247670	0.26	20.79	1.975	8.465
B	10.384219	10.24	5.702955	0.14	20.16	5.280	15.330
C	10.200350	10.49	5.585533	0.01	22.27	5.690	14.290

fig: 6 - Centrality measures for the Time of Catch

Figure 6 represents the Centrality measures for the time. As shown, bait 'A' is, on average, mostly used at 5 am, where as bait B and C are, on average, used mostly during 10am. From the Maximum column we can figure out that bait C is used in the late hours of up-to 22(10pm) as well, where as A and B are only used till 20(8pm). This can also be related to figure 3 where the swarm-plot indicates this very observation, as there are only data points for Bait 'C' after 22(10pm).

Bait_Type	Mean	Median	Standard Deviation	Minimum	Maximum	Q1 \	Q3
A	1.528987	1.34	1.012441	0.01	4.88	0.780	2.085
B	1.783750	1.79	1.078417	0.16	4.75	0.785	2.405
C	1.680973	1.66	1.142795	0.01	4.87	0.700	2.460

fig: 7 - Centrality measures for the Size of Catch

Figure 7 represents the Centrality measures for the size of catch. As shown, bait 'B', on average, gives the biggest catch sizes. This can also be seen on the Box-Plot in figure 5, where B has the highest mean line. And Bait A has the highest upper limit of 4.88, also shown in the figure 5 with the 2 upper dots. Bait C has the largest Q3 meaning that 75% of the catches for it lie below 2.46.

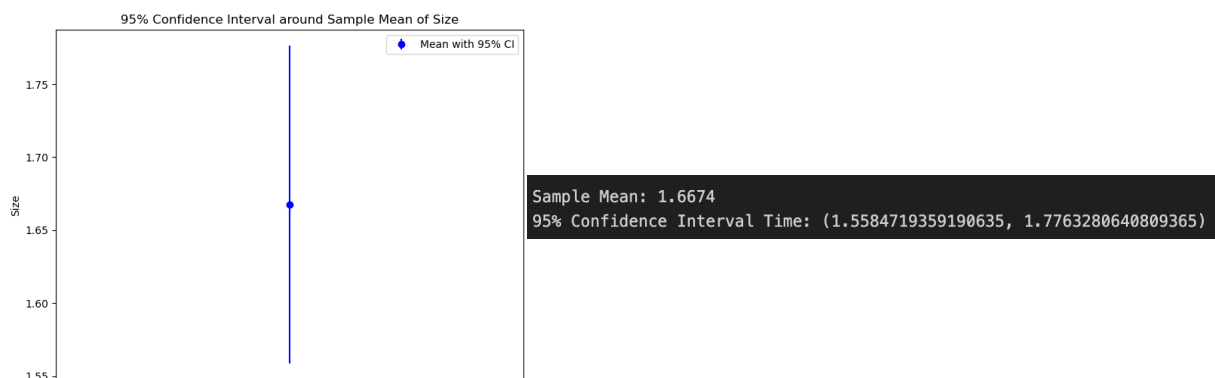


fig: 8 - 95% Confidence Interval for Size of Catch

Figure 8 shows a plot where a dot in the middle represents the mean value for size and the ends represent the edges of the 95% confidence value. The 95% confidence interval line extends from 1.558 to 1.776. Therefore, we can say with 95% confidence that the mean value for the population will lie within the line.



fig: 9 - 95% Confidence Interval for Time of Catch

Figure 9 shows a plot where the dot in the middle represents the mean value of time, which is around 9.3 as seen in the figure. The 95% confidence value lies on the line that extends from 8.8 to 9.8. Therefore, we can say with 95% confidence that the mean value for the population will lie within the line.

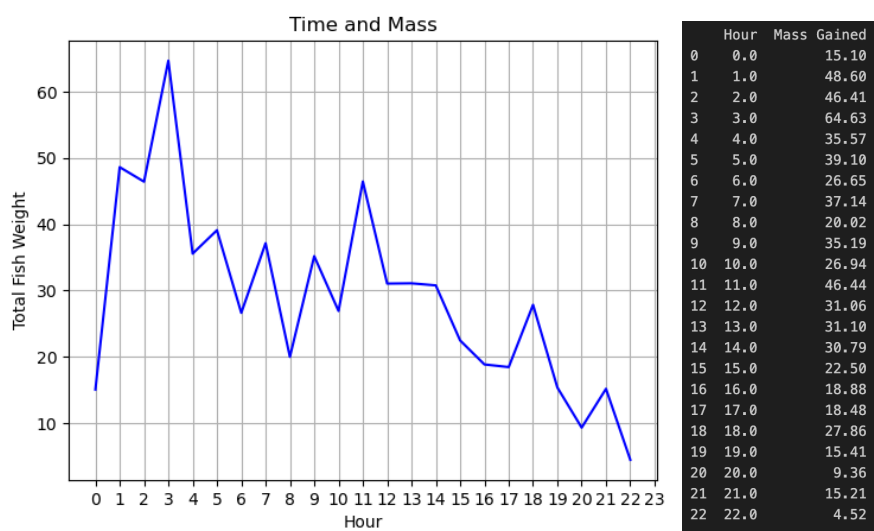


fig: 10 Hour vs Total Weight

This graph shows the sum of size of catches (Total Weight) every hour. The mass was calculated as the sum of the catch size on an hourly basis. We can see that 3am yields the maximum weight of 64.63. This is the sum of catches in that hour, and should not be confused with the frequency of catches in that hour. In terms of frequency, the hours 11-12 am yield the most catches as indicated in figure 1.

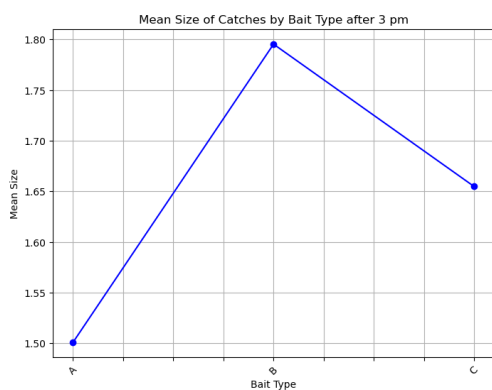


fig 11: Bait type vs Mean Size of Catch for 3 pm

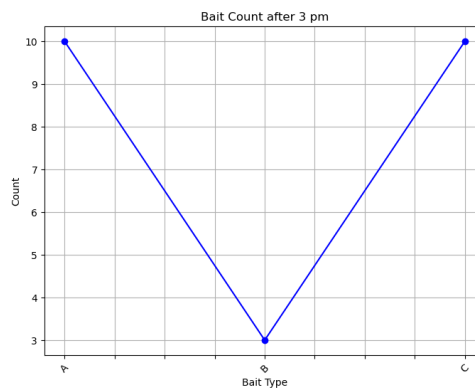


fig 12: Bait type vs Catch Count for 3 pm

As observed on the line graph on figure 11 The bait 'B' yields the heaviest catches on average at 3pm. On figure 12, we can observe that bait B yields the least number of catches at the 3pm hour, whereas both 'A' and 'C' yield equal number of catches, 10 catch each.

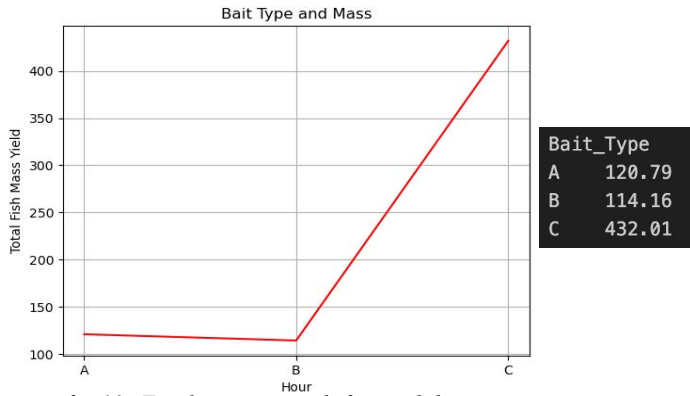


fig 13: Total mass gained for each bait type

In conclusion, the best type of bait to use overall is 'C' as it yields most frequent catches, also as seen in figure 13, 'C' yields the most total mass(calculated as the sum of size). The best time to go for fishing is 11-12 in the morning, as supported by figure1. We also observe that at 3 pm, the best type of bait to use for single largest catch size is the Bait 'B'. As seen in the swarm-plot(figure 4), bait 'A' and 'C' yield most catches at 3pm. 'A' and 'C' both yield 10 catches each whereas bait 'B' only yields 3 catches at 3pm, also observed in figure 12. So if we were to aim for more frequent catches at 3pm we would go with bait 'A' or 'C', whereas if the aim was to maximize catch size at 3pm we would go with bait 'B'.

We also see that although bait 'A' doesn't yield many catches on average, the single largest size is gained using bait A.