CP2403 - Project – Part 2 - ANOVA

First Name: Kantapong
Last Name: Wongsanguan

Step 1: hypothesis
Investigative question: Does the Silicate concentration in water vary significantly across different water depth categories?
Null hypothesis (H$_o$): There is no significant difference in the mean Silicate concentration levels among the different water depth categories (Silicate concentration and depth are unrelated). H0: $\mu1 = \mu2 = \mu3 = \mu4$ (where $\mu1$, $\mu2$, $\mu3$, and $\mu4$ represent the mean Silicate concentrations for the "shallow," "kinda shallow," "kinda deep," and "deep" water depth categories, respectively).
Alternative (H$_a$) hypothesis: There is a significant difference in the mean Silicate concentration levels among at least two of the different water depth categories (Silicate concentration and depth are related). Ha: Not all μ are equal; indicating that at least one pair of μ values is significantly different.
Step 2: Data Selection
- CalCOFI bottle dataset   - Water depths between 0-200m (shallow), 200-400m (kinda shallow), 400-600m (kinda deep), and 600-1000m (deep)   - Silicate concentration of water at these depths   - Null values dropped
Step 3: Assess the evidence (ANOVA)
R-squared (R^2): 0.799 This indicates that approximately 79.9% of the variance in Silicate concentration can be explained by the depth categories.
F-statistics: 4.479e+05 The F-statistic is very high, suggesting that there is a significant overall effect of depth categories on Silicate concentration.
Prob(F-statistics): 0.00 The p-value is less than 0.05 (assuming a significance level of 0.05), indicating that there is strong evidence to reject the null hypothesis. This means that there is a significant difference in the mean Silicate concentration levels among the different depth categories.

Mean values:
means for SiO3µM by depth category:

DEPTH_CAT	SiO3uM
shallow	12.460418
kinda shallow	48.673586
kinda deep	76.594739
deep	102.190119

STD values:
std for SiO3µM by depth category:

DEPTH_CAT	SiO3uM
shallow	11.771988
kinda shallow	11.207970
kinda deep	8.657754
deep	14.099362

Step 4: Draw Conclusion

The statistical analysis strongly suggests that there is a significant relationship between water depth categories (DEPTH_CAT) and Silicate concentration (SiO3uM). The mean Silicate concentration levels increase as we move from "shallow" to "deep" categories. Additionally, the variance in Silicate concentration is relatively lower in the "kinda deep" category compared to the other depth categories.

In practical terms, this means that there is an association between water depth and Silicate concentration, and this relationship is statistically significant.

Implications: the increase in Silicate concentration of the water with depth may indicate that more sardines are thriving in the shallower water, as sardines are high in Silicate and the bodies of dead sardines release Silicate into the water. However, Silicate, although essential for all life, can be a pollutant that causes eutrophication. Eutrophication leads to algae growth and oxygen depletion, which in turn causes the depletion of fish species.
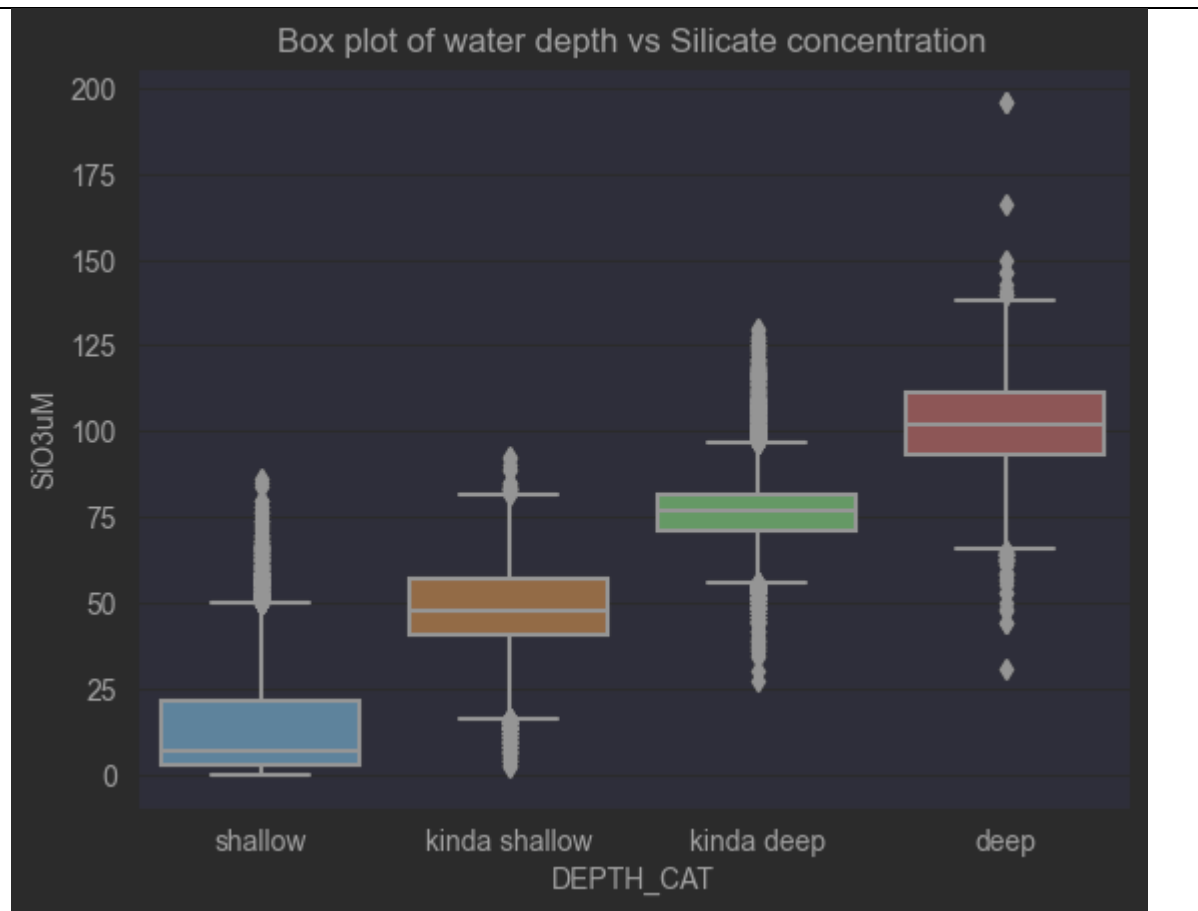
Box Plot

Figure 1: Box plot of water depth category vs silicate concentration

It is clear from the box plot that there is an increasing trend between depth categories and silicate concentration. As the water depth increase, the silicate increases gradually as shown in the figure 1. The incline in silicate level seems to be on a positive level with no stopping as the four boxes of the box plot do not overlap each other in any level of water depth.

Post Hoc Test

```
        Multiple Comparison of Means - Tukey HSD, FWER=0.05
    ================================================================
      group1      group2   meandiff p-adj  lower   upper   reject
    ----------------------------------------------------------------
        deep    kinda deep -25.5954  0.0 -26.1349 -25.0559  True
        deep kinda shallow -53.5165  0.0 -54.0413 -52.9917  True
        deep      shallow -89.7297  0.0 -90.246 -89.2134  True
   kinda deep kinda shallow -27.9212  0.0 -28.123 -27.7193  True
   kinda deep      shallow -64.1343  0.0 -64.3129 -63.9558  True
  kinda shallow     shallow -36.2132  0.0 -36.3405 -36.0858  True
    ----------------------------------------------------------------
```

New p-value: 0.05 / 6 = 0.0083

Mean Differences:

	Shallow	Kinda Shallow	Kinda Deep	Deep
Shallow				

Kinda Shallow	-36.2132			
Kinda Deep	-64.1343	-27.9212		
Deep	-89.7297	-53.5165	-25.5954	

P-values:

	Shallow	Kinda Shallow	Kinda Deep	Deep
Shallow				
Kinda Shallow	0.00			
Kinda Deep	0.00	0.00		
Deep	0.00	0.00	0.00	

Conclusion from post hoc test: True for difference between all depth categories, so we can say that there is a significant difference in silicate concentration between all depth categories. This aligns with the evidence from others; ANOVA and boxplot previously.