CP2403 - Project – Part 2 – CHI Squared

First Name: Caleb
Last Name: Webster

Step 1: hypothesis

Investigative question: Is the salinity of the water affected by the temperature of the water for extremely cold water (below 5 degrees) in the CalCOFI dataset?

Null hypothesis (H_o): There is no difference in the salinity levels between different water temperature categories for temperatures below 5 degrees.
s1 = s2 = s3 = s4

Alternative (H_a) hypothesis: There is a difference in the salinity levels between different water temperature categories for temperatures below 5 degrees.
s1 != s2 != s3 != s4

Step 2: Data Selection

- CalCOFI bottle dataset
  - Water temperatures between 0 and 5 degrees, cut into categories: 0-2, 2-3, 3-4, 4-5
  - Water salinity readings taken at these temperatures, cut into two categories, <= 34.5 and > 34.5 (digitized to 0 & 1)
  - Null values dropped

Step 3: Assess the evidence (Chi Squared)

Cross Tab (observed numbers)

TEMP_CAT	1	2	3	4
SAL_CAT				
0	8	204	8363	20086
1	3548	2991	6318	3336

Cross Tab (Percentages)

TEMP_CAT	1	2	3	4
SAL_CAT				
0	0.002250	0.063850	0.569648	0.857570
1	0.997750	0.936150	0.430352	0.142430

Cross Tab (expected numbers)

TEMP_CAT	1	2	3	4
SAL_CAT				

```
0         2272.228 2041.555 9380.928 14966.29
1         1283.772 1153.445 5300.072 8455.711
```

Chi Squared Value: 15988.268

p-value: 0.0

Step 4: Draw Conclusion

The p-value (0.00) is less than 0.05, so we reject the null hypothesis and accept the alternative hypothesis; there is a difference in the salinity levels between different water temperature categories for temperatures below 5 degrees.

Implications: in extremely cold water (0 to 5 degrees), the water salinity decreases as water temperature increases. Water temperature and water salinity may not affect sardines directly, but may correlate to other factors, such as food availability, plant growth, etc. This correlation between salinity and water temperature at low temperatures may be insightful for understanding these other factors, but don't have much direct affect on the sardines themselves.

Post-hoc test (if any)

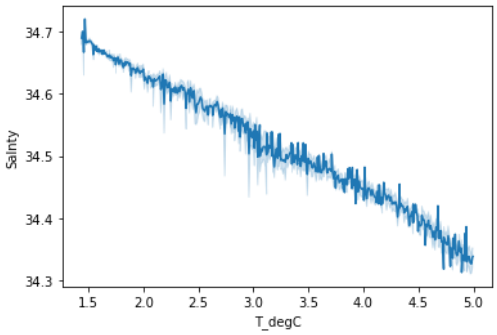
New p-value: 0.0083 (0.05 / 6 comparisons)

```
1  versus  2 Chi value:  207.9295410246399       p value:  3.886795514368588e-47
1  versus  3 Chi value:  3709.0330728229574      p value:  0.0
1  versus  4 Chi value:  11879.032761438164      p value:  0.0
2  versus  3 Chi value:  2687.7700538025792      p value:  0.0
2  versus  4 Chi value:  9770.439849917746       p value:  0.0
3  versus  4 Chi value:  3953.1829453406053      p value:  0.0
```

The results of the post hoc test show that there is a significant difference between salinity categories for all categories of water temperature, because all p-values are greater than 0.0083.

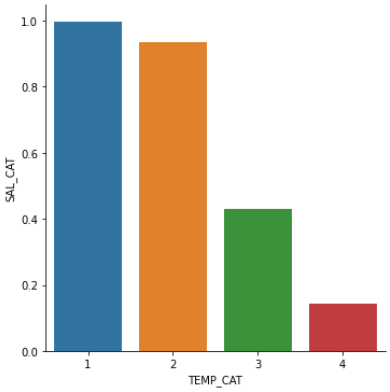
Plot/Chart(s)

Figure 1: line plot of water temperature (quantitative) vs water salinity (quantitative)



The line plot shows a strongly decreasing trend between water temperature and salinity. As temperature increases, salinity decreases.

Figure 2: Bar chart of water temperature (categorical) vs water salinity (categorical)



The bar chart also shows a decreasing trend between the temperature categories and the number of salinity readings over 34.5. As temperature categories get warmer, the number of readings above 34.5 decreases.