# CP2403 - Assignment – Part 2 – Task 2: Chi-Squared Analysis

First Name:

Last Name:

|  |
| --- |
| **Step 1: hypothesis** |
| Null hypothesis (Ho):  There is no significant association between the two selected categorical variables. |
| Alternative (Ha) hypothesis:  There is a significant association between the two selected categorical variables. |
| **Step 2: Data Selection** |
| For this analysis, I selected   1. categorical variable "T\_qual" (Temperature Quality) 2. categorical variable "S\_qual" (Salinity Quality) |
| **Step 3: Assess the evidence (Chi Squared)** |
| Cross Tab (observed numbers)  Cross Tab (observed numbers):  S\_qual 4.0 5.0 6.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 \  T\_qual  0.0 1 1 1 0 0 2 1 3 10 7  1.0 0 0 0 0 0 0 0 0 0 0  2.0 0 0 0 0 0 0 0 0 0 0  3.0 0 0 0 0 0 0 0 0 0 0  4.0 0 0 0 0 0 0 0 0 0 0  5.0 0 0 0 0 0 0 0 0 0 0  6.0 0 0 11585 0 0 0 0 0 0 0  7.0 0 0 0 0 0 0 0 0 0 0  8.0 0 0 0 190 18 0 0 0 0 0  9.0 0 0 0 37 7702 0 0 0 0 0  S\_qual ... 285.0 287.0 290.0 291.0 298.0 307.0 311.0 312.0 323.0 \  T\_qual ...  0.0 ... 0 0 0 0 0 0 0 0 0  1.0 ... 0 0 0 0 0 0 0 0 0  2.0 ... 0 0 0 0 0 0 0 0 0  3.0 ... 0 0 0 0 0 0 0 0 0  4.0 ... 0 0 0 0 0 0 0 0 0  5.0 ... 0 0 0 0 0 0 0 0 0  6.0 ... 1 2 2 1 2 0 0 0 0  7.0 ... 0 0 0 0 0 2 1 2 1  8.0 ... 0 0 0 0 0 0 0 0 0  9.0 ... 0 0 0 0 0 0 0 0 0  S\_qual 344.0  T\_qual  0.0 0  1.0 0  2.0 0  3.0 0  4.0 0  5.0 0  6.0 0  7.0 1  8.0 0  9.0 0  [10 rows x 253 columns] |
| Cross Tab (Percentages)  S\_qual 4.0 5.0 6.0 8.0 9.0 10.0 \  T\_qual  0.0 0.588235 0.588235 0.588235 0.000000 0.000000 1.176471  1.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  2.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  3.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  4.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  5.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  6.0 0.000000 0.000000 98.578965 0.000000 0.000000 0.000000  7.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  8.0 0.000000 0.000000 0.000000 91.346154 8.653846 0.000000  9.0 0.000000 0.000000 0.000000 0.478098 99.521902 0.000000  S\_qual 11.0 12.0 13.0 14.0 ... 285.0 287.0 \  T\_qual ...  0.0 0.588235 1.764706 5.882353 4.117647 ... 0.000000 0.000000  1.0 0.000000 0.000000 0.000000 0.000000 ... 0.000000 0.000000  2.0 0.000000 0.000000 0.000000 0.000000 ... 0.000000 0.000000  3.0 0.000000 0.000000 0.000000 0.000000 ... 0.000000 0.000000  4.0 0.000000 0.000000 0.000000 0.000000 ... 0.000000 0.000000  5.0 0.000000 0.000000 0.000000 0.000000 ... 0.000000 0.000000  6.0 0.000000 0.000000 0.000000 0.000000 ... 0.008509 0.017018  7.0 0.000000 0.000000 0.000000 0.000000 ... 0.000000 0.000000  8.0 0.000000 0.000000 0.000000 0.000000 ... 0.000000 0.000000  9.0 0.000000 0.000000 0.000000 0.000000 ... 0.000000 0.000000  S\_qual 290.0 291.0 298.0 307.0 311.0 312.0 \  T\_qual  0.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  1.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  2.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  3.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  4.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  5.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  6.0 0.017018 0.008509 0.017018 0.000000 0.000000 0.000000  7.0 0.000000 0.000000 0.000000 28.571429 14.285714 28.571429  8.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  9.0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000  S\_qual 323.0 344.0  T\_qual  0.0 0.000000 0.000000  1.0 0.000000 0.000000  2.0 0.000000 0.000000  3.0 0.000000 0.000000  4.0 0.000000 0.000000  5.0 0.000000 0.000000  6.0 0.000000 0.000000  7.0 14.285714 14.285714  8.0 0.000000 0.000000  9.0 0.000000 0.000000  [10 rows x 253 columns] |
| Cross Tab (expected numbers)  S\_qual 4.0 5.0 6.0 8.0 9.0 10.0 \  T\_qual  0.0 0.008327 0.008327 96.479060 1.890277 64.286064 0.016654  1.0 0.005780 0.005780 66.967818 1.312074 44.622092 0.011560  2.0 0.006613 0.006613 76.615724 1.501102 51.050698 0.013226  3.0 0.005094 0.005094 59.022483 1.156405 39.327945 0.010189  4.0 0.002596 0.002596 30.078766 0.589322 20.042126 0.005192  5.0 0.006319 0.006319 73.210580 1.434386 48.781778 0.012638  6.0 0.575655 0.575655 6669.540632 130.673720 4444.057801 1.151310  7.0 0.000343 0.000343 3.972667 0.077835 2.647073 0.000686  8.0 0.010189 0.010189 118.044967 2.312809 78.655890 0.020377  9.0 0.379084 0.379084 4392.067303 86.052070 2926.528533 0.758168  S\_qual 11.0 12.0 13.0 14.0 ... 285.0 287.0 \  T\_qual ...  0.0 0.008327 0.024982 0.083272 0.058290 ... 0.008327 0.016654  1.0 0.005780 0.017340 0.057801 0.040460 ... 0.005780 0.011560  2.0 0.006613 0.019838 0.066128 0.046289 ... 0.006613 0.013226  3.0 0.005094 0.015283 0.050943 0.035660 ... 0.005094 0.010189  4.0 0.002596 0.007788 0.025961 0.018173 ... 0.002596 0.005192  5.0 0.006319 0.018957 0.063189 0.044232 ... 0.006319 0.012638  6.0 0.575655 1.726965 5.756552 4.029586 ... 0.575655 1.151310  7.0 0.000343 0.001029 0.003429 0.002400 ... 0.000343 0.000686  8.0 0.010189 0.030566 0.101886 0.071320 ... 0.010189 0.020377  9.0 0.379084 1.137252 3.790840 2.653588 ... 0.379084 0.758168  S\_qual 290.0 291.0 298.0 307.0 311.0 312.0 323.0 \  T\_qual  0.0 0.016654 0.008327 0.016654 0.016654 0.008327 0.016654 0.008327  1.0 0.011560 0.005780 0.011560 0.011560 0.005780 0.011560 0.005780  2.0 0.013226 0.006613 0.013226 0.013226 0.006613 0.013226 0.006613  3.0 0.010189 0.005094 0.010189 0.010189 0.005094 0.010189 0.005094  4.0 0.005192 0.002596 0.005192 0.005192 0.002596 0.005192 0.002596  5.0 0.012638 0.006319 0.012638 0.012638 0.006319 0.012638 0.006319  6.0 1.151310 0.575655 1.151310 1.151310 0.575655 1.151310 0.575655  7.0 0.000686 0.000343 0.000686 0.000686 0.000343 0.000686 0.000343  8.0 0.020377 0.010189 0.020377 0.020377 0.010189 0.020377 0.010189  9.0 0.758168 0.379084 0.758168 0.758168 0.379084 0.758168 0.379084  S\_qual 344.0  T\_qual  0.0 0.008327  1.0 0.005780  2.0 0.006613  3.0 0.005094  4.0 0.002596  5.0 0.006319  6.0 0.575655  7.0 0.000343  8.0 0.010189  9.0 0.379084  [10 rows x 253 columns] |
| * Chi Squared value: 177900.07488813816 * p-value: 0.0 |
| **Step 4: Draw Conclusion** |
| The Chi-Squared Test resulted in a highly significant Chi Squared value (Chi Squared = 177900.07, p < 0.05), indicating a significant association between the quality of temperature (T\_qual) and the quality of salinity (S\_qual). The observed and expected numbers in the cross-tabulation table demonstrate a significant deviation from what would be expected under the assumption of independence.  Conclusion: We reject the null hypothesis. There is a significant association between the quality of temperature and the quality of salinity. |
| **Post-hoc test (if any)** |
| Hint : Compute the new p-value  Since the p-value is already extremely low (p < 0.05), indicating a highly significant result, there is no need for a post-hoc test in this case. |
| **Plot/Chart(s)** |
|  |