# Application of Statistical Methods to Analyze Groundwater Quality

**Yashbir Singh[1] and Manish Kumar[2]**

## Abstract

Multivariate Hotelling $T^2$ and Univariate student's t-test are applied to compare the water quality of two monitoring sites. Four water quality parameters viz. Chloride, Alkalinity, Fluoride and Total Dissolve Solids (TDS) were examined and statistical analysis showed that both monitoring sites are not identical in water quality parameters (except fluoride) and both sites are contaminated because contamination level of Fluoride and TDS in water are above the WHO and Indian standard. This paper also demonstrates that multivariate test (Hotelling $T^2$) is more powerful (in the sense of rejecting null hypothesis when in fact it is false) than student's t-test. A graphical comparison of the power of both tests is also presented.

**Keywords:** Multivariate analysis, Hotelling $T^2$, student's t-test, water quality.

---

[1] School of Mathematics, Statistics and Computational Sciences, Central University of
  Rajasthan, Kishangarh (Ajmer), India, e-mail : singh.yshbir@gmail.com
[2] Department of Statistics, University of Rajasthan, Jaipur (Rajasthan), India,
  e-mail:  manishkr29@gmail.com

# 1 Introduction

Statistical techniques play a major role in evaluation of environmental data. These techniques help us to know whether monitoring sites are contaminated or not and also to check the similarity in water quality data among monitoring sites. In past decades lots of rules and regulations and standards were defined by WHO, EPA based on statistical procedures. Indian government has also established concern department Central Pollution Control Board (CPCB) dedicated to evaluate water quality in India and Bureau of Indian Standard (BIS) to set the standards for safe drinking water. In the present study we demonstrate an application of some useful statistical techniques to examine the water quality of two monitoring sites.

# 2 Preliminary Notes

There are various statistical techniques like student's t-test, ANOVA, Multivariate Hotelling $T^2$ and MNOVA to compare two or more monitoring sites including one or more water quality parameters. Before applying these techniques we make certain assumptions for these parametric tests:

    (i)      Samples are selected randomly from normally distributed populations having common variance.

    (ii)      Samples from monitoring sites are independent.

Generally the environmental data are not normal so we used normal probability plot (Graphic method) and Shapiro-Wilk test to check the normality of data. Test result shows that data is not normal. We applied log transformation and inverse transformation to data to get normality of data. Log transformation applied on Chloride, Alkalinity and TDS whereas inverse transformation applied on Fluoride. Finally we used student's t-test and Hotelling $T^2$ to examine water quality of two monitoring sites.

## 2.1 Student's t-test

Student's t-test used for two purposes (I) to determine whether water quality parameters of particular site is contaminated or not and (II) to compare the water quality of two monitoring sites. The statistical test for student's t is given by

(I)    To compare each parameter of a particular site with defined standard, we have

$$t = \frac{\overline{x} - \mu}{s / \sqrt{n}}$$

(1)

where $\overline{x}$ is the mean value of examined parameter, s is the standard deviation and n is the sample size.

(II) To compare water quality parameter of both monitoring sites, we have

$$t = \frac{\overline{x}_1 - \overline{x}_2}{S_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

(2)

where $S_p = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$ and $\overline{x}_1$ and $\overline{x}_2$ are the mean values of examined parameter for site1 and site2 respectively, $s_1$ and $s_2$ are the standard deviations and $n_1$ & $n_2$ are the sample size for both sites respectively.

## 2.2 Hotelling $T^2$ test

Hotelling $T^2$ test is also similar test as student's t-test except it considers all water quality parameters at same time rather compare one by one. The $T^2$ is given by

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} \left(\overline{x}_1 - \overline{x}_2\right)' S^{-1} \left(\overline{x}_1 - \overline{x}_2\right)$$

(3)

where $\overline{x}_1$ and $\overline{x}_2$ are the mean vectors of site1 and site2 respectively and S is the pooled sample covariance matrix from both sites. S is given by

$$S = \frac{s_1 + s_2}{n_1 + n_2 - 2}$$

(4)

where $s_1$ and $s_2$ are the covariance matrices of site1 and site2.

# 3 Main Results

We applied individual student's t-test to check whether water quality parameters of both sites are under permissible limit or differ significantly. We defined null hypothesis (one tail test) with 5% level of α that parameters of water quality are less than or equal to defined standard by WHO or Indian government. As per Bureau of Indian Standards (BIS) for drinking water permissible limit of alkalinity is 200 mg/l. We setup hypothesis that alkalinity is below or equal to 200 mg/l for a particular site. P-value of the test is more than defined level of α (i.e. 0.05) and we accept the null hypothesis concluding that alkalinity is under permissible limit for the examined site. T-test result and the corresponding p-value for each water quality parameter of both monitoring sites are given below in Table 2.

As result given in Table 2 clearly shows that two water quality parameters viz. fluoride and TDS are contaminated for both sites; whereas other two (alkalinity and chloride) are under permissible limit.

Further we applied student's t-test to check the similarity in water quality data of both sites. We setup the null hypothesis that there is no significance difference in water quality of both sites i.e. $\mu_1 = \mu_2$. To apply the test we also test the assumptions of equal variance for each parameter and test result showed that both sites have the same variability because the p-value of test is above 0.05. Result shown in the Table 3 indicates that there is significance difference in water quality of both monitoring sites (except fluoride). Further a multivariate Hotelling $T^2$ test applied to compare the water quality of both sites. Null hypothesis defined assumed no difference in water quality for both the sites. i.e. $H_0$: $\underset{\sim}{\mu}^{(1)} = \underset{\sim}{\mu}^{(2)}$, where $\underset{\sim}{\mu}^{(1)}$ and $\underset{\sim}{\mu}^{(2)}$ are the

population mean vector of site1 and site2 respectively. P-value (given in Table 4) is more than 0.05 which shows that water quality data differ significantly for both sites. As both student's t-test and Hotelling $T^2$ test provided the same result but if we compare the power of both test (Figure 1) we found that Hotelling $T^2$ test gives more power than student's t-test for small number of sample observations.

Table 1: Descriptive statistics of water quality parameters

| Site# | Sample Size | Parameters | Min. | Max. | Mean | Std. Deviation |
|-------|-------------|------------|------|------|------|----------------|
| 1 | 30 | Alkalinity | 100 | 260 | 168.77 | 46.66 |
| | | Chloride | 60 | 290 | 116.37 | 48.56 |
| | | Fluoride | 0.40 | 1.60 | 0.64 | 0.22 |
| | | TDS | 594 | 1904 | 909.61 | 269.66 |
| 2 | 30 | Alkalinity | 70 | 270 | 128.39 | 51.88 |
| | | Chloride | 50 | 190 | 92.25 | 32.07 |
| | | Fluoride | 0.40 | 1.60 | 0.69 | 0.30 |
| | | TDS | 406 | 1154 | 765.40 | 225.92 |

Table 2: Hypothesis testing result of water quality parameters

| Site# | Parameters | Null Hypothesis | p-value | Result |
|-------|------------|-----------------|---------|--------|
| 1 | Alkalinity | $\mu \leq 200$ | 0.99 | Not contaminated |
| | Chloride | $\mu \leq 250$ | 0.99 | Not contaminated |
| | Fluoride | $\mu \leq 1.5$ | 0.01 | Contaminated |
| | TDS | $\mu \leq 500$ | 0.01 | Contaminated |
| 2 | Alkalinity | $\mu \leq 200$ | 0.99 | Not contaminated |
| | Chloride | $\mu \leq 250$ | 0.99 | Not contaminated |
| | Fluoride | $\mu \leq 1.5$ | 0.01 | Contaminated |
| | TDS | $\mu \leq 500$ | 0.01 | Contaminated |

Table 3: Water quality comparison of both sites using student's t-test

| Parameters | Null Hypothesis | Degree of freedom | t-value | p-value |
|---|---|---|---|---|
| Alkalinity | $\mu_1 = \mu_2$ | 58 | 3.58 | 0.0007 |
| Chloride | $\mu_1 = \mu_2$ | 58 | 2.30 | 0.0249 |
| Fluoride | $\mu_1 = \mu_2$ | 58 | 0.50 | 0.6212 |
| TDS | $\mu_1 = \mu_2$ | 58 | 2.44 | 0.0178 |

Table 4: Hotelling $T^2$ test for water quality comparison of both sites

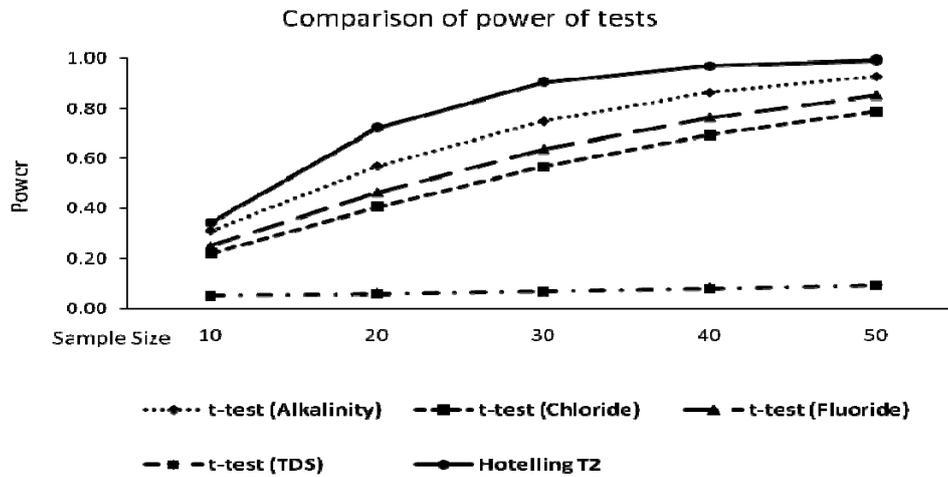| Hotelling $T^2$ Statistic | F | DF1 | DF2 | p-value |
|---|---|---|---|---|
| 15.066 | 3.5716 | 4 | 55 | 0.0116 |



Figure 1:  Comparison of power of tests

# 5 Conclusions

Statistical procedures are very useful for environmental data to determine whether monitoring sites are contaminated or not and also widely used to compare the sites quality. In the present study both univariate (t-test) and multivariate (Hotelling $T^2$ test) procedure used to determine the water quality of two monitoring sites. Both tests provided the same result but multivariate tests gives more power (probability of correctly rejecting null hypothesis when it's actually false) and need smaller number of samples.

# References

[1]   N. Johnson and D. Wichem, *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1982.

[2]   G.W. Snedecor and W. G. Cochran, *Statistical Methods*, Iowa State University Press, Ames, 1989.

[3]   D.R. Helsel and R.M. Hirsch, *Statistical Methods in Water Resources*, Elsevier Science Publishing Company Inc., New York, 1992.

[4]   U.S. Environmental Protection Agency (EPA), *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities*, *Addendum to Interim Final Guidance*, Washington, DC, 1992.

[5]   A. Genizi, Some power comparisons between bivariate tests and a twofold univariate test, *Journal of the Royal Statistical Society*, **35** (Series B), (1993), 86-96.

[6]   R.D. Gibbons, *Statistical Methods for Groundwater Monitoring*, John Wiley & Sons, Inc., New York, 1994.

[7]   Ross, D. Lauren, Multivariate Statistical Analysis of Environmental Monitoring Data, *Ground Water*, **35**(6), (1997), 1050-1057.