Appendix D

Statistical Analysis for Tank Farm Closure

Appendix D
Statistical Analysis for Tank Farm Closure

D-1. INTRODUCTION

Several different statistical methods will be applied to the Tank Farm Facility (TFF) closure data. There are two primary objectives with regard to the statistical analysis that will be performed on the data. The first objective is to determine if the constituents of interest are present in levels greater than the specified action level. Confidence intervals will be used for this analysis. The second objective is to determine if the contents of the tanks and the vault sumps came from the same population. This will be done by performing Analysis of Variance (ANOVA) on the data from the samples collected in the vault sumps at Tanks WM-184, WM-185, and WM-186 and the data from samples collected within the two tanks. ANOVA also will be used when more data are obtained from other tanks. Five samples will be taken from each tank and one sample from each of the two vault sumps for each tank (a total of six samples from the vault sumps). This provides a total of 21 samples from Tanks WM-184, WM-185, and WM-186.

D-2. CONFIDENCE INTERVALS

Confidence intervals will be used to determine if any of the constituents of concern in the tanks or the vaults exceed the specified action levels. This is done by constructing a 90% confidence interval for the concentration of each constituent in each tank and comparing the upper confidence limit with the specified action level. If the upper confidence limit is less than the action level, then the constituent is considered to be present in levels less than the action level. If the upper confidence limit is greater than the action level, then it is assumed that the constituent is present in concentrations that are greater than the action level and appropriate action will be taken.

D-2.1 Construction of a Confidence Interval

A confidence interval is constructed using the sample mean and standard deviation of the data. For each constituent, the mean concentration, $\overline{X}$, is calculated using the equation

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{D-1}$$

where

n = the number of observations in the data set

X_i = the i^{th} observation in the data set.

The standard deviation, s, is calculated using the equation

$$s^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1} \tag{D-2}$$

The confidence interval is calculated using the expression

$$\overline{X} \pm t_{1-\alpha,n-1}\sqrt{\frac{s^2}{n}}$$

where

$t_{1-\alpha,n-1}$ = the *t*-statistic at *1-α* with *n-1* degrees of freedom

So

$$UCL = \overline{X} + t_{1-\alpha,n-1}\sqrt{\frac{s^2}{n}} \qquad \text{(D-3)}$$

where

UCL = upper confidence limit.

The *t*-statistic can be found on a *t*-table or from a statistical software package. In the case of the analysis for the TFF closure, *α*=0.05 since the 95% upper confidence limit is being used. This is the significance level of a statistical hypothesis test. Essentially comparing the upper limit of a confidence interval to the action level is comparable to performing a one-sample *t*-test of the sample mean against the action level at the *α*=0.05 level. (The 95% upper confidence limit is the upper limit of a 90% confidence interval. Since it is only the upper confidence limit that is being compared to the action level, setting α=0.05 gives the test an overall significance level of 0.05.)

D-2.2 Use of the Confidence Interval

Once the confidence interval has been calculated for a given constituent concentration, a comparison can be made against the action level for that constituent. The general rule is if

$$\overline{X} + t_{1-\alpha,n-1}s < AL \qquad \text{(D-4)}$$

where

AL = action level

then it can be confidently concluded that the constituent concentration is less than the action level. However, if

$$\overline{X} + t_{1-\alpha,n-1}s \geq AL \qquad \text{(D-5)}$$

then it cannot be concluded that the constituent concentration is less than the action level. In this situation, it is assumed that the constituent concentration exceeds the action level and the appropriate action should be taken.

A confidence interval will be constructed for every constituent of concern in each tank and in the vault sumps for each tank. This means if there are 10 constituents of interest, 40 confidence limits will be calculated and compared to the appropriate action levels.

Let's work through an example calculation to determine the 95% upper confidence limit. If the sample data are $\overline{X} = 0.87$, $s^2 = 0.073$, $t_{0.05,9} = 1.833$, and $UCL = 0.87 \pm 0.1565$, which corresponds to an upper confidence limit at 1.03 mg/L, then the calculation yields the following:

Liquid Arsenic Sample Data (Example)

Sample No.	Concentration (mg/L)	Sample No.	Concentration (mg/L)
1	0.79	6	0.98
2	0.85	7	0.87
3	0.92	8	0.78
4	0.75	9	0.88
5	0.80	10	1.06

Since the action level for liquid arsenic has been set at 1.05, it can be determined that for these 10 samples, there is 95% confidence that the true mean is less than 1.03 mg/L. This method is adapted from *Test Methods for Evaluating Solid Wastes*, *Physical/Chemical Methods* EPA SW-846 (1998).

D-2.3 Assumptions of Confidence Intervals

When constructing a confidence interval, the data must be approximately normally distributed to meet the assumptions of the confidence interval. Since the *t*-statistic is used to generate the confidence interval, the interval is robust against certain variations from the normal distribution. However, the data still need to be symmetric about the mean and free of outliers. Since the *t*-statistic is robust against slight variations from the normal distribution, performing a hypothesis test to verify the normality of the data is not appropriate. Statistical tests that are used to determine if a data set follows a certain distribution are highly sensitive to variations of the data from the distribution in question. Because of this, data that fail to meet the requirements of the statistical test for normality may still produce a reliable confidence interval. In fact, if a statistical test for determining the normality of the data does show that the data are normal (i.e., the null hypothesis is not rejected), then the *z*-statistic should be used in the confidence interval instead of the *t*-statistic. The normality of the data can be better assessed by examining the summary statistics of the data and through graphical methods such as histograms.

Another assumption that is made when constructing a confidence interval is that the sample mean and the standard deviation are independent. This is always the case if the data are truly normally distributed. Because of this, it is assumed that this assumption is met if the data appear to be approximately normally distributed.

D-2.4 Using the Lognormal Transformation

Since the type of data that will be obtained from the TFF tanks is non-negative, it is likely that the data will be log normally distributed rather than normally distributed. This means that the natural log of the data points have a normal distribution. The traditional method for analyzing lognormal data is to take the log of all of the data points and perform the statistical analysis on the transformed data. Any methods that are appropriate for the normal distribution can be applied to the transformed data. However, this can pose some complications with some analytical methods. For example, a confidence interval that is generated using the transformed data is accurate for estimating the mean of the transformed data, but the interval cannot be transformed back to the scale of the raw data to estimate the mean of the raw data. However, the *t*-test can be accurately performed on the transformed data against a cutoff value such as the action level of a constituent. The test is performed by taking the log of the raw data and calculating the mean and standard deviation using the transformed data. These values are then used to perform a *t*-test against the log of the action level. Because the confidence interval is only being used to conduct a *t*-test for the data from the TFF, the results obtained by comparing the 95% upper confidence limit of log transformed data against the log of the action level is as accurate a test as comparing the 95% upper confidence limit against the action level if the raw data were truly normally distributed.

It is possible that the data that will be obtained from the TFF will be neither normal nor log normally distributed. If this is the case, other transformations will be attempted on the data to see if normality can be achieved with some transformation. The methods described above will be applied to the transformed data. As with the natural log transformation of the data, confidence intervals can be used to perform a *t*-test on the transformed data.

D-3. ANALYSIS OF VARIANCE (ANOVA)

The second type of analysis of interest is the use of one-way ANOVA to determine if the contents of the tanks and vault sumps came from the same population. A separate ANOVA will be performed for each constituent of concern. One-way ANOVA is similar to the *t*-test. In fact, the *t*-test is a special case of one-way ANOVA. ANOVA is a statistical hypothesis test for determining if the means of several groups are different from each other. In the situation of the tanks and vault sumps in the TFF, each tank or vault sump is considered a group. ANOVA is used instead of a *t*-test because many different *t*-tests would need to be performed to make all of the desired comparisons. This will increase the significance level, α. Since multiple tests would be run on the same set of data, the significance level would no longer be 0.05. This is because the significance level applies to the chance of achieving significance in the analysis, not just one test. Although the chance of making a Type I error (rejecting the null hypothesis when it is in fact true) on a single test is only 0.05, the chance of making a Type I error somewhere in at least one of several tests is much greater than 0.05. ANOVA is a more appropriate way to deal with this type of situation.

D-3.1 Use of ANOVA

As stated above, ANOVA is a test of the means between several different groups. The null hypothesis is that there is no difference in analyte concentrations between all of the tanks and vault sumps. This means that the contents of the tanks and vault sumps came from the same population. The alternative hypothesis is that there is a difference in analyte concentration levels between the tanks and vault sumps. This means that the contents of the tanks and vault sumps do not come from the same population. Note that the alternative hypothesis does not specify which tanks or sump vaults are different from each other. It could be that all the tanks and vault sumps have significantly different constituent concentrations or it could be that only one of the tanks or vault sumps has a different mean concentration than one, or all, of the other tanks or vault sumps. If the P-value associated with the ANOVA test indicates that there is a significant difference in concentration levels between the tanks and vault sumps (i.e., $P < 0.05$), then multiple means comparison testing will be used to determine which tanks and/or vault sumps are different from each other. Just because significance is achieved using ANOVA, it does not necessarily mean that there is significant contamination in the tanks or vault sumps. It could be that two of the post-decontamination residuals in the tanks have different mean concentrations from each other, but that none of the tanks or vault sumps have constituent concentrations that are significantly greater than the action level.

The results of the ANOVA test are presented in a table that looks like this:

Model	DF	SS	MS	F	P
Group	DFG	SSG	MSG	F	P
Error	DFE	SSE	MSE		
Total	DFT	SST			

In the table,

DFG = number of tanks and sump vaults –1

DFT = total number of samples –1

DFE = DFT – DFG

$$SSG = n\sum_{groups}(\bar{x}_i - \bar{x})^2$$

$$SST = \sum_{obs}(x_{ij} - \bar{x})^2$$

SSE = SST – SSG

MSG = SSG/DFG

MSE = SSE/DFE

F = MSG/MSE

where

n = the total number of samples taken from each tank

DFX = the degrees of freedom for term X

SSX = the sum of squares for the term X

MSX = the mean square for the term X

F = the F-statistic

P = P-value.

The P-value can be found from an *F*-table. The degrees of freedom in the numerator are DFG and the degrees of freedom in the denominator are DFE (this is only pertinent if you are in fact going to look up the P-value on a table).

The P-value is the number that is of primary interest. If P is less than 0.05, then the null hypothesis is rejected and there is some difference between the analyte concentrations in the tanks and/or sump vaults. If P is greater than or equal to 0.05, then there is not sufficient evidence to reject the null hypothesis and it can be concluded that the contents of the tanks and vault sumps come from the same population.

ANOVA can be used to analyze the data from Tanks WM-184, WM-185, and WM-186 and the corresponding vault sumps, and can also be used to analyze the data as more data are obtained. A separate ANOVA needs to be generated for each constituent of concern.

One issue with this particular data set is that the data are unbalanced. This means that each group does not have the same number of observations in it. Each of the tanks will consist of 5 observations per tank. Each vault sump group will contain 2 observations. There are two different ways to handle this situation. One way is to analyze the tanks separately from the vault sumps. The benefit of doing this is that the design will be balanced and the mathematics will be simpler. The disadvantage is that a direct comparison between the tanks and the vault sumps cannot be made. The other method is to use type III sums of squares to generate the F-statistics instead of the type I sums of squares. The advantage of this method is that all of the tanks and vault sumps can be analyzed in the same design and therefore they all can be compared against each other. The disadvantage is that the equations for the sums of squares for ANOVA that are listed above are no longer applicable, so the mathematics become very complex in generating the sums of squares. However, since a computer will be used to perform all of the calculations, the mathematical complexity does not present a problem. It is recommended that all of the data be analyzed in the same model and that type III sums of squares are used to generate the F-statistics.

D-3.2 Assumptions of ANOVA

Several assumptions are made when performing ANOVA on the data. They are as follows:

- The data are approximately normally distributed
- The groups have approximately equal variance
- The group mean and standard deviation are independent.

These assumptions need to be verified before the results of ANOVA can be considered reliable. Since ANOVA is based on the F-statistic, the test is robust against small variations from the normal distribution. However, the data do need to be symmetric and free of outliers. As with the confidence interval, the use of a statistical test to determine the normality of the data is not appropriate because it is far more conservative than is necessary for ANOVA (see Section D-2.3).

The normality assumptions can be verified through examining residual plots. Residual plots are generated by plotting the residuals against the predicted values generated from ANOVA and by plotting the residuals against the groups. A residual is calculated by subtracting the value predicted from the ANOVA model from the corresponding observed data value. Residual plots also are the standard method for determining if the groups have approximately equal variance. Normal-quantile plots and symmetry plots also can be used to assess symmetry, the presence of outliers in the data, and how close the data follow a normal distribution. A histogram of the residuals can also be examined to determine the normality of the data. These methods are sufficient for establishing that the normality assumption has been met. As with the confidence intervals, if the data look to be sufficiently normally distributed then it is assumed that the group mean and standard deviation are independent. This is because for data that are truly normal, the sample mean and standard deviation are always independent.

D-4. REFERENCES

EPA, 1998, *Test Methods for Evaluating Solid Wastes, Physical/Chemical Methods*, EPA SW-846, Revision 5, April.