UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

MAR - 2 1989

MEMORANDUM

OFFICE OF WATER

SUBJECT: Statistical Analysis of Sampling Data from Hebdon Electronics, Inc.

FROM: Marla Dombroski, Statistician mtd
Statistics Section, ESAB, AED (WH-586)
Office of Water Regulations and Standards

TO: Jeremy Johnstone, Environmental Engineer
Compliance Branch
Water Management Division, (W-4) Region 9

THRU: Henry D. Kahn, Chief HDK
Statistics Section, ESAB, AED (WH-586)
Office of Water Regulations and Standards

This memo is in response to your request for a statistical analysis of the sampling data from Hebdon Electronics, Inc. (HEI) transmitted to Henry Kahn in your memorandum of January 27, 1989. The data from HEI (see Appendix) indicate that two types of sampling methods were used: grab and composite. These data include a substantial number of values that exceed the Federal Categorical Pretreatment Standards for lead, copper, total metals, and pH. The Standards limit the concentrations of given parameters that may be discharged over a 24-hour period. The Standards are ≤0.6 mg/l for lead, ≤4.5 mg/l for copper, ≤10.5 mg/l for total metals, and ≥5.0 for pH. The relevant regulations for these Standards specifically state that sampling for compliance with these Standards should be based upon 24-hour composite sampling. In the data from HEI, clear patterns of violation of the Standards exist for both grab and composite samples.

The critical question is whether the patterns of violations from the composite and grab samples lead to different conclusions about discharge from the HEI plant. Based on the analyses described in this memorandum, the conclusion is that both grab and composite samples display similar patterns of violation of the Standards for lead, copper, total metals and pH. The analyses did not find any statistically significant difference in the violation rates between the grab and composite data. Statistical tests of pH using the local limits show no significant difference between the grab and composite samples in the local violation rates. Furthermore, additional statistical tests of the two data sets generated by the two sampling methods indicate that the means and variances for each pollutant are similar. Each analysis is described in this memo.

The discussion below is divided into four sections. The first section gives a general overview of grab and composite sampling methods and their relevance to the HEI data. The second section describes the HEI data. The third section gives the statistical analysis of the HEI data. This third section is divided into four parts: graphical displays, descriptive statistics, direct tests of violation rates, and tests of comparability of the data sets. Conclusions based on the analyses are summarized in the fourth section of the memo.

GRAB VERSUS COMPOSITE SAMPLES

The terms "grab" and "composite" refer to the manner in which samples of wastewater are obtained. This section describes grab and composite samples in four parts. The first two parts give general discussions of grab samples and composite samples. The third part compares the two sampling methods. The fourth part discusses the relevance of the sampling methods to the HEI data.

Grab Samples

A grab sample is a single sample taken at one point during the time period being sampled. The concentration measured in the sample is then taken to represent the concentration over the entire sampling period. In this case, the grab samples were evidently taken once during a daily, 24 hour period. The measured concentration in the grab sample is then associated with the entire 24 hour period.

Composite Samples

A composite sample is a number of grab samples taken during a sampling period that are then mixed, i.e., composited. The concentration measured in the mixture is then associated with the entire sampling period. In determining the concentration of a composite sample, correction is made for the different flow rates when each grab sample was taken.

Comparison of Sampling Methods

In order to assess the comparability of grabs and composites taken on the same day, a number of lab analyses need to be performed. The composite sample and each grab sample that comprises the composite must be separately analyzed chemically. The results from the grab samples may then be compared to the results of the composite sample. These analyses were not done in the case of the HEI data. Grab samples do, however, provide information about the process that would be included in a composite sample taken on the same day. Thus, it is conceivable that if a grab sample indicates a violation then a composite sample taken on the same day would also indicate a violation.

HEI Data

The HEI data under review here can be used to address the question of whether the chemical measurements made on grab and composite samples display comparable patterns of violation of the Standards. The conclusion based on the analyses described here is that both grab and composite samples display similar violation rates of the Standards. In addition, the data is used to assess the similarity of the means and variances between the grab and composite sample data.

DESCRIPTION OF THE HEI DATA

This section gives a general description of the HEI data. The sampling from the HEI plant took place over a four year time period: June 1, 1984 through June 8, 1988. HEI used two types of sampling methods: grab and composite. The grab samples were all collected prior to October 1987. The composite samples were collected from October 1987 through June 1988. Both types of data display patterns of violation of the Standards as described below.

HEI Grab Samples

Forty-three grab samples were collected at the HEI plant between June 1, 1984 and September 16, 1987. Thirty of these forty-three samples were in violation of at least one of the Standards. Twenty-five lead samples, twenty-three copper samples, and twelve pH samples violated their respective Standard. Only eleven samples were analyzed for the concentration of total metals. All eleven of these samples were in violation of the Standard.

HEI Composite Samples

After September 1987, composite sampling was used at the HEI plant. Nine composite samples were collected between October 14, 1987 and June 8, 1988. Of these nine samples, six were in violation of at least one of the Standards. Five lead samples, four copper samples, and one pH sample violated their respective Standard. Only two samples were analyzed for total metals. Both were in violation of the Standard for total metals.

STATISTICAL ANALYSIS OF THE HEI DATA

The approach used in the analysis is to examine the data using a variety of statistical procedures. If a number of different statistical procedures show similar results, the hypotheses regarding the comparability of data generated by two sampling methods are strengthened. Lead, copper, and pH are analyzed by examining graphical displays, descriptive statistics, direct tests of violation rates, and tests of comparability of the data sets. Since the values reported for total metals are in most cases lower bounds on the actual concentrations, only the graphical display and the descriptive statistics for total metals are included in the discussion below. Since all of the values for total metals, regardless of the sampling method, exceed the Standard, the violation rates are the same for both sampling methods.

The analyses of the direct tests of violation rates and tests of comparability of the data sets use a significance level of 0.05. The significance level is the probability of accepting the hypothesis that the two data sets are the same when they are actually different. The significance level is set at 0.05 by convention for statistical analyses. If the test statistic calculated from the data has a significance level less than 0.05, one of two outcomes are possible. One is that the two data sets are different; the other is that the data sets are not different and a rare event been observed. The probability of the rare event occurring is 0.05. Since the probability of the event is small, an observed test statistic significant at the 0.05 or less level is considered evidence that the data sets are different. When the test statistic has a significance level greater than 0.05, the conclusion is that the test failed to find evidence that the two data sets are different.

The analyses of the graphical displays, descriptive statistics, direct violation rates, and tests of comparability of the data sets are each presented separately.

GRAPHICAL DISPLAYS

Plots of the data across time provide a visual indication of variability of grab and composite sample data and whether the variability is time related. These plots also show the similarities in violation rates between the data generated by the two sampling methods. General similarities between the overall distribution of the two data sets from the two sampling methods can also be observed in the plots. The results of the graphical summaries are given below for lead, copper, total metals, and pH.

<u>Results</u>

Lead

The lead data show a similar scattering of values above and below the lead Standard for both grab and composite samples. The lead Standard is indicated by the horizontal line at 0.6 mg/l in Figure 1. The points in red indicate violations of the Standard.

Copper

The copper data also show a pattern of measurements above and below the Standard for both grabs and composites. The copper data, however, display a pattern of decreased variability over time that may indicate a change in processes at the plant. The change in variability seems to have occurred in 1986. This decreased variability over time is evident in looking at the pattern of the grab data. One composite observation in 1987 is extremely high. Because of this one large composite observation, the composite data are actually more variable than the grab data although this difference is not statistically significant (see section on Comparison of variances). Aside from the one large composite value, the decreased variability seems to hold for the composite data. The copper Standard is indicated by the horizontal line at 4.5 mg/l in Figure 2. The points in red indicate violations of the Standard.

Total Metals

All of the total metals data exceed the Standard. Due to one large grab value in 1985, the variability of the grab samples is higher than the composite samples. The difference in variability is not statistically significant. The data are sparse: only eleven grab samples and two composite samples. The data are plotted in Figure 3; the horizontal line indicates the Standard of 10.5 mg/l. All of the points are in red indicating that all violate the Standard.

pH

The pH data show a similar scattering of values above and below the pH Federal Standard and the local limits for both grab and composite samples. The pH Federal Standard is indicated by the solid horizontal line at 5.0 in Figure 4. The points in red indicate violations of this Standard. The broken lines indicate the local limits of 6 and 9.

DESCRIPTIVE STATISTICS

The descriptive statistics indicate comparability of data sets. The descriptive statistics listed below include the mean and standard deviation of the natural logarithms of the data (listed as the logmean and logS.D. respectively). The logmean and logS.D. are used in comparing the variances and means where the tests assume that the data are normally distributed. The logarithmic transformation of the data gives an approximately normal distribution. By definition, pH is already logarithmic. The descriptive statistics for both sampling methods: grab and composite, are discussed below for lead, copper, total metals, and pH.

Results

Lead:

The sample sizes (N) are different for grab and composite samples: 43 versus 9 respectively. The number of violations (N_v) of the Federal Standard of ≤0.6 mg/l is 25 for grab samples versus 5 for composite samples. The violation rates (N_v/N) are about the same: 0.581 for grabs, 0.556 for composites. The means and standard deviations (S.D.) are similar for both sampling methods. The range of data is larger for the grabs than the range for the composites. The logmean and log standard deviation are included below.

Type	N	N_v	Mean	S.D.	Min	Max	logmean	logS.D.
Grab	43	25	1.53	1.89	0.04	9.0	-0.55	1.62
Composite	9	5	1.36	1.80	0.29	6.0	-0.20	1.00

Copper:

The sample sizes (N) are different for grab and composite samples: 43 versus 9 respectively. The number of violations (N_v) of the Federal Standard of ≤4.5 mg/l is 23 for grab samples versus 4 for composite samples. The violation rates (N_v/N) are close: 0.535 for grabs, 0.444 for composites. The means and standard deviations (S.D.) appear somewhat different for both sampling methods. However, comparison of variances and means do not show statistically significant differences. The results from those analyses are presented in a later section (Tests of Comparability of the Data Sets). The range of data is larger for the composites than the range for the grabs. The logmean and log standard deviation are included below.

Type	N	N_v	Mean	S.D.	Min	Max	logmean	logS.D.
Grab	43	23	7.26	7.74	0.16	36.3	1.42	1.20
Composite	9	4	11.10	20.51	1.17	65	1.46	1.34

Total Metals:

The sample sizes (N) are different for grab and composite samples: 11 versus 2 respectively. Most of the data values for total metals are reported as lower bounds of the concentrations in the samples. The calculations for the statistics listed below use these lower bounds as the data values. All of the data are in violation of the Federal Standard of ≤10.5 mg/l. The means and standard deviations (S.D.) appear somewhat different for both sampling methods. However, the means and standard deviations are not exact values; they are calculated from the reported lower bounds of the true values of the data. The range of data is larger for the grabs than the range for the composites. The logmean and log standard deviation are included below.

Type	N	N_v	Mean	S.D.	Min	Max	logmean	logS.D.
Grab	11	11	48.21	91.47	11	322.8	3.21	0.94
Composite	2	2	38.00	38.18	11	65	3.29	1.26

pH:

The sample sizes (N) are different for grab and composite samples: 43 versus 9 respectively. The number of violations (N_V) of the Federal Standard of ≥5.0 is 12 for grab samples versus 1 for composite samples. The federal violation rates (N_V/N) are 0.279 for grabs and 0.111 for composites. These violation rates are not statistically significantly different (see section on Direct Tests of Violation Rates). The number of violations (N_L) of the local limit of 6 to 9 is 18 for grab samples versus 4 for composite samples. The local violation rates (N_L/N) are 0.419 for grabs and 0.444 for composites. The means and standard deviations (S.D.) appear somewhat different for both sampling methods. However, the comparisons of variances and means are not statistically significant (see section on Tests of Comparability of the Data Sets). The range of data is larger for the grabs than the range for the composites.

Type	N	N_V	N_L	Mean	S.D.	Min	Max
Grab	43	12	18	5.99	1.90	2.3	12.4
Composite	9	1	4	7.27	2.56	2.1	11.2

DIRECT TESTS OF VIOLATION RATES

Direct tests of violation rates are explicit tests of whether data sets differ with respect to violation rates. Both the binomial test and Fisher's exact test provide direct comparisons of the rates of violation versus compliance in the data generated by the two sampling methods. The tests require that the data be classified into two outcomes: either violation or compliance. These tests consider only the number of observations in each of the classifications. (These tests do not consider the magnitude of the deviation of the individual values from the Standard.) The numbers of violations for each sampling method are given above in the discussion of the descriptive statistics. The results from the binomial test and Fisher's test are described below.

Binomial Test

The binomial test provides a direct comparison of the rates of compliance versus violation between the two sampling methods. The results given below are computed from the normal approximation to the binomial test as described in Fienberg (1980). The value given by this test is compared to the critical value determined by the significance level. If the value exceeds the critical value, then evidence of an association exists between the sampling methods and the outcomes (violation or compliance).

Results

The rates of violation versus compliance between grab samples and composite samples are not statistically different for lead, copper, and pH at the 0.05 significance level. Two different limits are used in assigning violations to pH values. The Federal Standard requires that the values be greater than 5.0 to comply. The local limits are between 6 and 9. The results for pH are not significantly different regardless of which limit is chosen. In order to be significant, the absolute value would need to exceed the critical value of 1.96 for lead, copper, and pH. The results of the binomial test are listed below.

Metal	Test Absolute Value	Critical Value	Significant Difference?
Lead	0.141	1.96	No
Copper	0.497	1.96	No
pH (federal)	1.057	1.96	No
pH (local)	0.140	1.96	No

Fisher's Exact Test

Fisher's exact test also compares the rates of compliance versus violation in the data generated by the two sampling methods. The purpose of the test is to determine whether the observed violation rates in the data generated by the two sampling methods are significantly different. The test assumes that the violation rates are the same for the two data sets generated by the two sampling methods. The test uses this assumption in calculating the probability of obtaining the observed arrangement of the data into classifications of compliance and violation for the two data sets. A small probability (i.e., less than the significance level of 0.05) for the observed results indicates a small likelihood that the rates are the same for both data sets. A large probability (i.e., greater than the significance level of 0.05) supports the hypothesis that the rates are the same for both data sets.

Results

The observed results from Fisher's exact test all have probabilities substantially greater than the significance level of 0.05; thus, the data do not provide evidence of differences in violation rates between the data generated by the two sampling methods for lead, copper and pH. Two different limits are used in assigning violations to pH values. The Federal Standard is that the values must be greater than 5.0. The local limits require pH to be between 6 and 9. The results for pH are not significant regardless of which limit is chosen. The results of Fisher's exact test are listed below.

Metal	Probability	Significant Difference?
Lead	>0.999	No
Copper	0.722	No
pH (federal)	0.420	No
pH (local)	>0.999	No

TESTS OF COMPARABILITY OF THE DATA SETS

The tests of comparability are explicit tests of indicators of the similarity between the two data sets generated by the two sampling methods. The indicators tested in these analyses are the means and variances. The results from the comparison of variances and from the comparison of means are discussed below.

Comparison of variances between grab and composite samples

The variances for grab and composite samples are compared by using the F-test. The variance is the square of the standard deviation and is a standard measure of the variability in a data set. The standard deviations are listed in the section on Descriptive Statistics.

F-test

The F-test compares the ratio between the variance of the composite samples and the variance of the grab samples. The F-test assumes that the data are normally distributed. The variances of the natural logarithms of the data (logS.D. squared) are used for lead and copper since effluent pollutant data of this kind are usually lognormally distributed. This ratio is then compared to critical values corresponding to the degrees of freedom associated with each group. The degrees of freedom for a group is the number of observations in the group minus one. The variances of the two sampling methods are different if the value of the ratio is outside the range given by the two critical values.

Results

The F-test did not find a significant difference between the grab and composite data for lead, copper, total metals, and pH. In order to be significant at the 0.05 level, the ratio would need to be outside the range of (0.261,2.512) for lead, copper and pH. The results of the comparison are listed below.

Metal	Ratio	Critical Values	Degrees of Freedom	Significant Difference?
Lead	0.380	0.261,2.512	8,42	No
Copper	1.243	0.261,2.512	8,42	No
pH	1.815	0.261,2.512	8,42	No

Comparison of means between grab and composite samples

The means for grab and composite samples are compared by using two methods. The means are listed in the section on Descriptive Statistics. The first method described below is the t-test. The t-test assumes that the data are normally distributed. The means and standard deviations of the natural logarithms of the data (logmean and logS.D.) are used for lead and copper since effluent pollutant data of this kind are usually lognormally distributed. The second method is the Wilcoxon two-sample test. The Wilcoxon two-sample test is a non-parametric (i.e., distribution-free) test. Non-parametric tests do not assume that the data follow particular distributional forms. The Wilcoxon two-sample test is therefore more general than the t-test which assumes that the data are normally distributed. Both tests are presented below.

t-test

The t-test compares the means of the two sampling methods by taking the difference between the two means and dividing by a pooled standard deviation. This number is then compared to a critical value corresponding to the degrees of freedom. The degrees of freedom are the total number of samples in both groups minus two. The pooled standard deviation accounts for the number of observations for each sampling method. A pooled standard deviation is used since the F-test failed to find differences between variances for the two sampling methods (described above in the section on Comparison of variances). If the t-value exceeds the critical value, then the means of the two sampling methods are statistically significantly different.

Results

None of these comparisons using the t-test show a statistically significant difference between the means. In order to be significant at the 0.05 level, the absolute value of t would need to exceed 2.009 for lead, copper and pH. The results of the comparison are listed below.

Metal	Absolute Value of t	Critical Value	Degrees of Freedom	Significant Difference?
Lead	0.607	2.009	50	No
Copper	0.068	2.009	50	No
pH	1.731	2.009	50	No

Wilcoxon Two-Sample Test

The Wilcoxon two-sample test is also a test of the difference in the means for the two sampling methods. The Wilcoxon test assigns ranks in ascending order to each of the observed values. The smallest value is assigned the rank of 1;

the largest value is assigned the rank equal to the total number of observations from the combined sampling methods. The ranks are then summed within each sampling method, and adjusted for the number of observations. The values obtained from these calculations are compared, and the test value computed. The absolute value of the test value is compared to a critical value which is determined by the significance level. If the critical value is exceeded, then evidence exists that the means for the two sampling methods are different.

Results

The comparisons for lead and copper show no evidence for differences between the means of the grab samples and the composite samples at the 0.05 significance level. The critical value for pH is slightly exceeded at a significance level of 0.05. However the value for pH (2.020) at a slightly smaller significance level of 0.04 does not exceed the corresponding critical value of 2.056.

Metal	Test Absolute Value	Critical Value	Significant Difference?
Lead	0.351	1.96	No
Copper	0.314	1.96	No
pH	2.020	1.96	Yes

CONCLUSION

Based on the analyses described in this memorandum, both grab and composite samples display similar patterns of violations of the Standards. The analyses are: graphical displays, descriptive statistics, direct tests of violation rates, and tests of comparability of the data sets.

The graphical displays give a general visual impression of the data. In general, the graphical displays show similar scattering of values around the Federal Standard for both sampling methods. The graphical display of copper indicates a decreased trend of variability after 1986.

The descriptive statistics also give a general impression of the data. The descriptive statistics indicate comparability of the data sets. The sample sizes of the two data sets are dissimilar. The means, variances, and violation rates look different in some cases; however, the sections on direct tests of violation rates and tests of comparability of the data sets did not show significant differences between the sampling methods.

The direct tests of the violation rates did not find any statistically significant difference in the violation rates of the Federal Standards between the grab and composite data at a significance level of 0.05. Statistical tests of pH using the local limits show no significant difference between the grab and composite samples in the local violation rates.

Tests of comparability of the two data sets generated by the two sampling methods indicate that the means and variances are similar for lead and copper. At a significance level of 0.05, all of the analyses for pH, except one, support the hypothesis that the variances and means are similar for both sampling methods. The Wilcoxon two-sample test for pH shows a significant difference in the means between the two sampling methods. However, the test value is close to the critical value; at a slightly lower significance level of 0.04, the result is not significant.

In conclusion, the statistical tests support the hypotheses that the two sampling methods do not have different violation rates; and that the two sampling methods have the same variances and means for lead, copper, and pH.

REFERENCES

Armitage, P., Statistical Methods in Medical Research, 1971, p.131-138.

Daniel, Wayne, Applied Nonparametric Statistics, 1978, p.110-115.

Duncan, Acheson, Quality Control and Industrial Statistics, 1965, p.512-513.

Fienberg, Stephen, The Analysis of Cross-Classified Categorical Data, 1980, p.8-10.

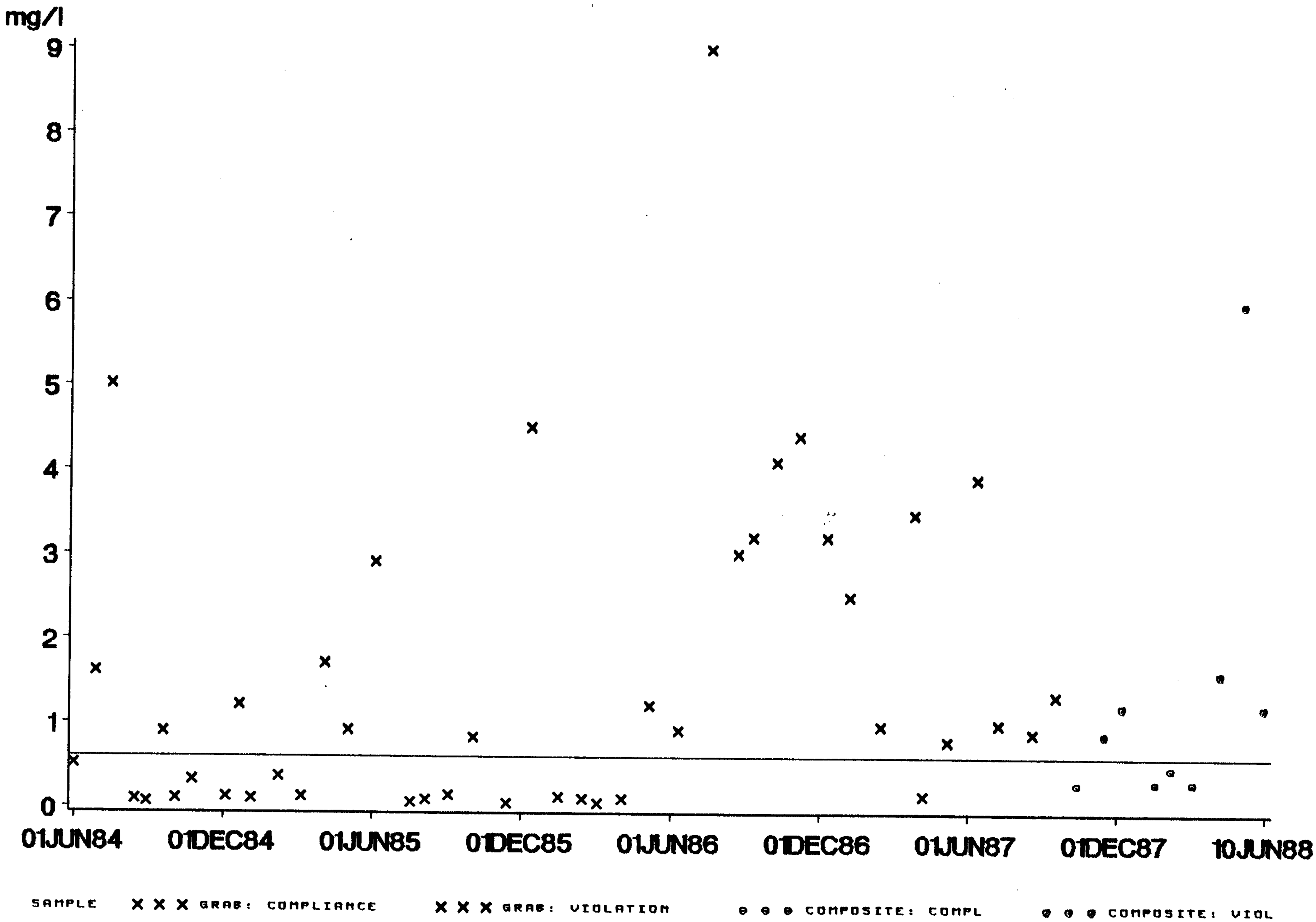
Walpole, Ronald & Myers, Raymond, Probability and Statistics for Engineers and Scientists, 1978, p.474-480.

APPENDIX: DATA FROM SAMPLING AT HEBDON ELECTRONICS, INC

Date	Sample Type	Lead	Copper	Total Metals	pH
06-01-84	GRAB	0.51	9.3		3.9
06-26-84	GRAB	1.6	7.1		4.0
07-13-84	GRAB	5.0	36.3	36.3	2.3
08-15-84	GRAB	0.08	0.25		6.92
08-29-84	GRAB	0.05	5.0		6.77
09-17-84	GRAB	0.88	1.02		4.34
10-03-84	GRAB	0.09	1.31		6.06
10-24-84	GRAB	0.32	21.4	>21.4	5.58
12-04-84	GRAB	0.12	2.9		5.13
12-19-84	GRAB	1.2	12	>12	3.89
01-04-85	GRAB	0.09	0.55		8.15
02-07-85	GRAB	0.36	7.8	322.8	3.2
03-07-85	GRAB	0.11	1.9		6.96
04-03-85	GRAB	1.7	14	>14	3.32
05-02-85	GRAB	0.91	29.5	>29.5	3.34
06-03-85	GRAB	2.9	14.6	>14.6	10.25
07-18-85	GRAB	0.05	2.0		6.7
08-06-85	GRAB	0.08	1.8		7.21
09-03-85	GRAB	0.14	1.0		7.26
10-02-85	GRAB	0.82	19	>19	3.42
11-13-85	GRAB	0.04	0.16		6.65
12-10-85	GRAB	4.5	5.6		3.58
01-16-86	GRAB	0.11	1.9		6.35
02-14-86	GRAB	0.10	3.9		5.74
03-04-86	GRAB	0.04	0.52		6.88
04-03-86	GRAB	0.09	2.4		6.93
05-06-86	GRAB	1.2	5.6		6.00
06-10-86	GRAB	0.9	9.8		6.48
07-15-86	GRAB	9.0	11	>11	12.40
08-21-86	GRAB	3.0	2.8		6.55
09-09-86	GRAB	3.2	17	>17	6.36
10-07-86	GRAB	4.1	3.3		6.95
11-04-86	GRAB	4.4	4.2		6.85
12-09-86	GRAB	3.2	5.7		5.66
01-06-87	GRAB	2.5	2.6		6.76
02-13-87	GRAB	0.96	8.6	32.67	3.51
03-25-87	GRAB	3.47	7.1		4.91
04-08-87	GRAB	0.14	6.0		6.67
05-07-87	GRAB	0.78	8.5		6.29
06-10-87	GRAB	3.9	4.9		6.53
07-08-87	GRAB	0.98	3.2		7.04
08-20-87	GRAB	0.87	5.6		6.99
09-16-87	GRAB	1.32	3.08		6.97
10-14-87	COMP	0.29	1.48		7.13
11-16-87	COMP	0.87	4.4		7.08
12-08-87	COMP	1.2	65	>65	9.19
01-21-88	COMP	0.31	1.18		5.56
02-10-88	COMP	0.48	1.56		2.10
03-09-88	COMP	0.31	1.17		7.56
04-12-88	COMP	1.6	6.1		9.00
05-10-88	COMP	6.0	11.0	11.0	11.2
06-08-88	COMP	1.2	8.0		6.64

FIGURE 1

HEI LEAD SAMPLING RESULTS VS TIME



Missing Figure 2

FIGURE 3

HEI TOTAL METALS SAMPLING RESULTS VS TIME

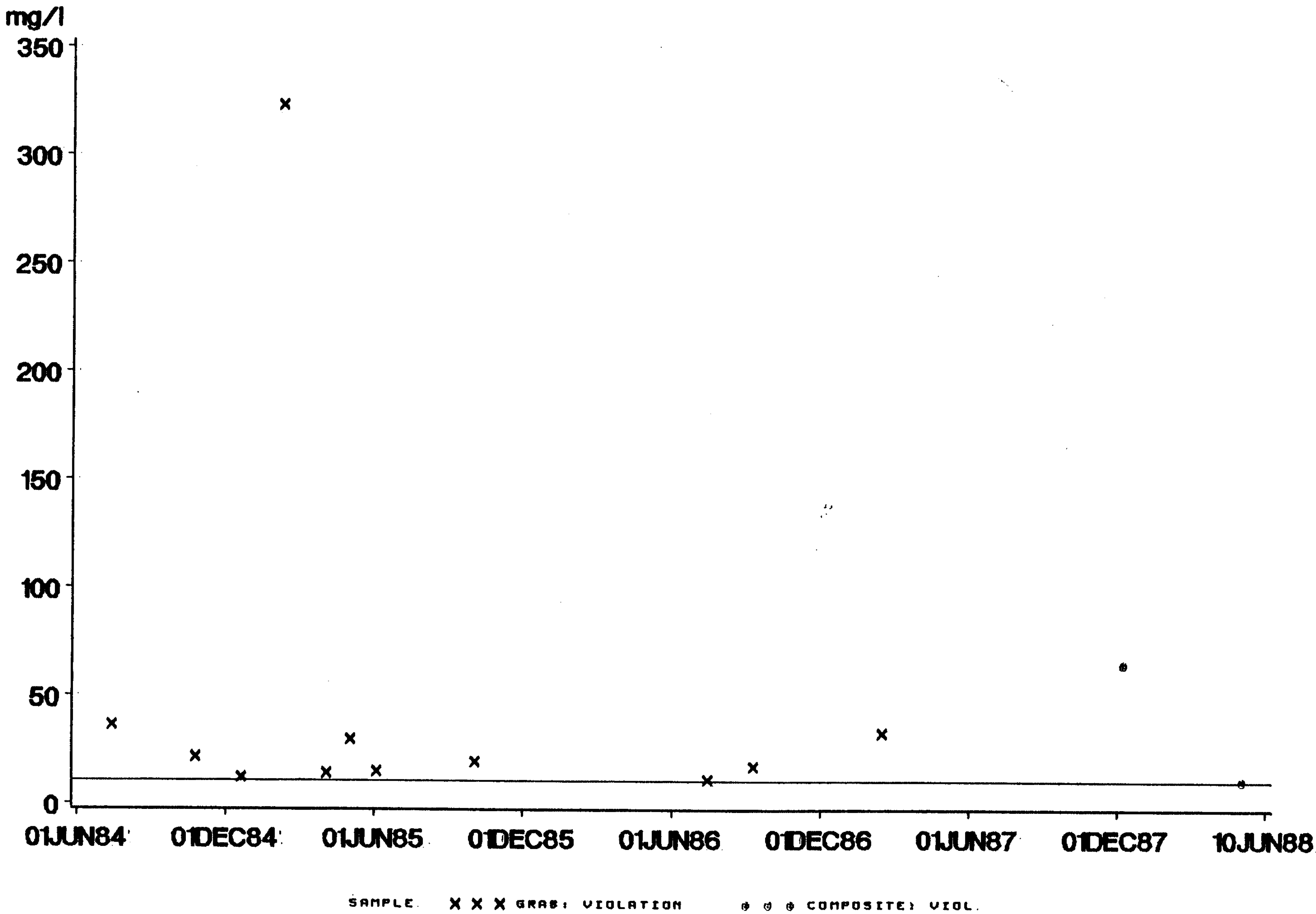
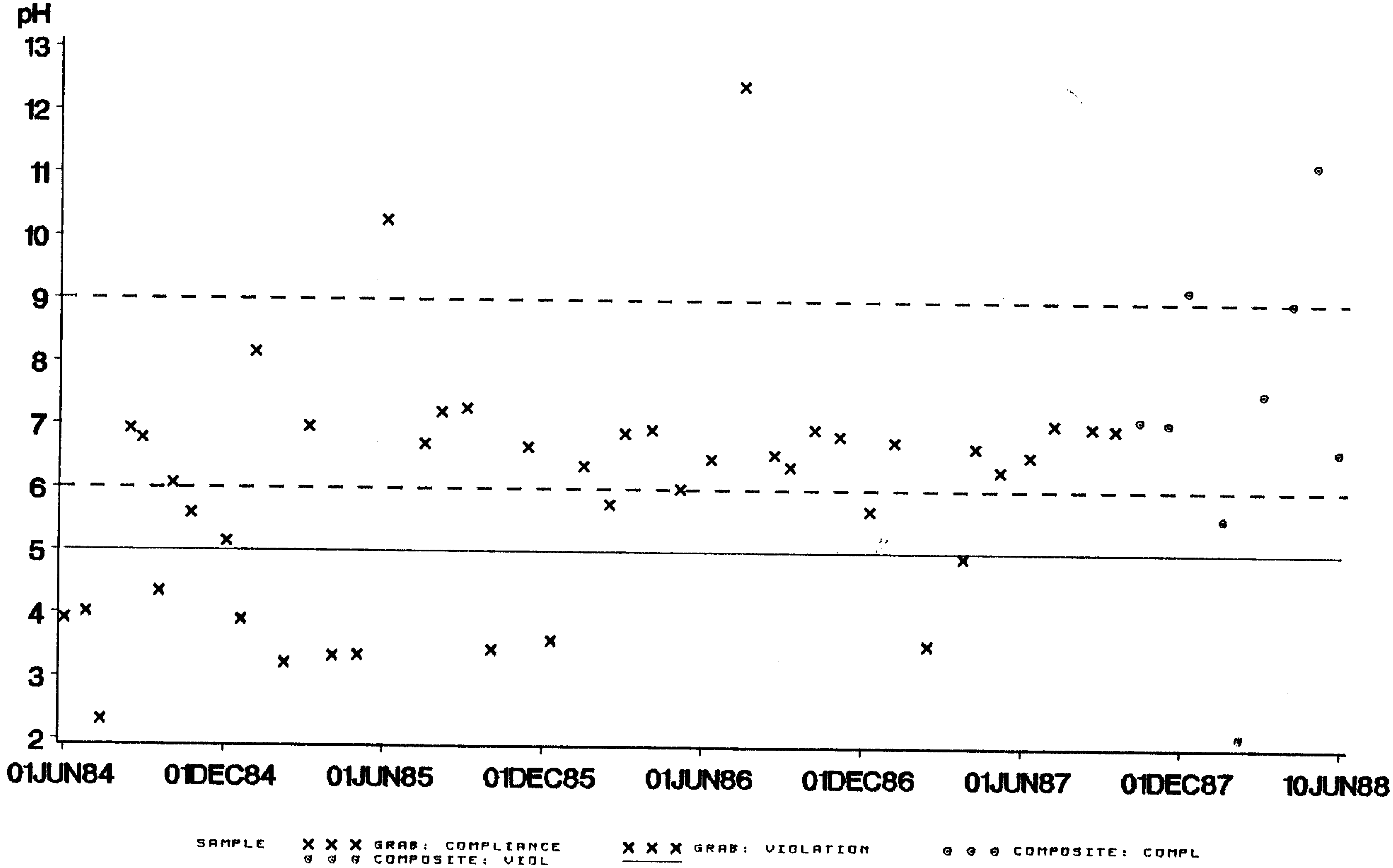


FIGURE 4

HEI pH SAMPLING RESULTS VS TIME



SOLID LINE INDICATES FEDERAL LIMIT
BROKEN LINES INDICATE LOCAL LIMITS