GIANT REFINING CO.

March 30, 1992

Route 3, Box 7
Gallup, New Mexico
87301

505
722-3833

Barbara Driscoll
U.S. Environmental Protection Agency
Region VI
1445 Ross Avenue
Suite 1200
Dallas, Texas 75202-2733

Re: RFI Phase II Report - Giant Refining Company
NMD048918817

Dear Ms. Driscoll:

Pursuant to the recent letter from Bill Honker, Giant Refining Company - Ciniza, submits the following information:

1. An extended explanation of the statistical methods used by Dr. Mark Wilson is attached.

2. Research into the original analytical data from the 1986 land demonstration indicated that the four (4) background valves for beryllium were indeed 1.0 ppm. There is no explanation for this other than coincidence or laboratory error. Beryllium values for other depths did vary from below 1.0 ppm to above 1.0 ppm, therefore we do not suspect laboratory error.

The background plot was resampled in 1991 and the new data was submitted to Rich Mayer. That data showed variance, so in the Phase II Draft Report both sets of data were used.

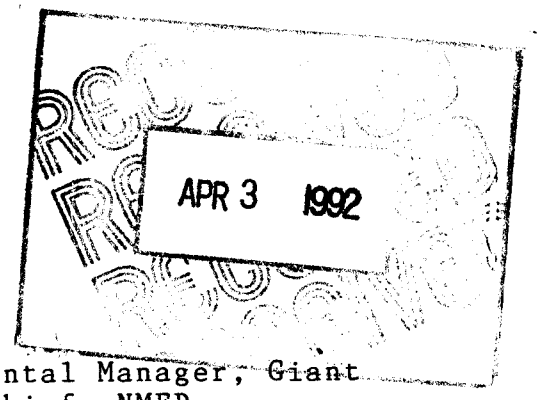
Hopefully, this information will answer your questions about the statistical analysis and background values.

If you require additional information, please contact me at (505) 722-0227.

Sincerely,

Lynn Shelton

Lynn Shelton
Environmental Assistant
Giant Refining Company - Ciniza



cc w/attachments: Zeke Sherman, Environmental Manager, Giant
Benito Garcia, Bureau Chief, NMED

GIANT REFINERY

Analysis of Variance for Beryllium (0 — 5 feet)

The analysis of variance carried out below is the standard one-way classification procedure. Data from four groups are used to calculate the *between-groups sum of squares*, SS_cores and the *within-groups sum of squares*, SS_error. From these values the corresponding *mean-squares* are calculated. An F value is then calculated from the mean squares:

(a) $$SS\_cores = \sum_{\text{all groups}} \left(\frac{\left(\sum X_t\right)^2}{N_t} \right) - \frac{X_{tot}^2}{N}$$

(b) $$SS\_error = \sum_{\text{all groups}} \left(\sum X_t^2 - \frac{\left(\sum X_t\right)^2}{N_t} \right)$$

In these two formulas the X_t's represent values in the *ith* group, N_t is the number of values in each group, X_{tot} is the grand total of all data values, and N is the total number of data values in all groups.

(c) $$MS\_cores = \frac{SS\_cores}{K - 1}$$ where K is the number of cores.

(d) $$MS\_error = \frac{SS\_error}{N - K}$$

(e) $$F = \frac{MS\_cores}{MS\_error}$$

Background Data:

1.0 1.0 1.0 1.0

$\sum X_1 = 4; \quad \sum X_1^2 = 4; \quad N_1 = 4$

Data from Evaporation Lagoons:

5.8	6.0	2.2	1.4	1.0	1.1	1.6	2.1	1.1	0.6	0.9	7.7	7.8
18.2	9.3	9.5	7.8	8.3	5.6	3.6	3.0	3.2	2.2	2.9	2.8	2.1
1.9	3.7	4.0	1.5	4.5	3.1	2.9	4.3	4.6				

$\sum X_2 = 148.3; \quad \sum X_2^2 = 1048.4; \quad N_2 = 35; \; \bar{x} = 4.2; \; s = 3.5$

Data from Aeration Lagoons:

1.8 2.8 2.6 2.7 3.5 3.4

$$\boxed{\sum X_3 = 16.8;\ \sum X_3^2 = 48.9;\ N_3 = 6;\ \bar{x} = 2.8;\ s = 0.56}$$

Data from Diversion Ditch:

2.4 3.6 4.3 3.2 4.1 4.3 4.3 4.6 4.9

$$\boxed{\sum X_4 = 35.7;\ \sum X_4^2 = 146.4;\ N_4 = 9;\ \bar{x} = 4;\ s = 0.73}$$

Totals:

$$\boxed{X_{tot} = 4 + 148.3 + 16.8 + 35.7 = 204.8: \quad N = 54}$$

Calculations for the analysis of variance are given below:

1. $SS\_cores = \left[\frac{4^2}{4} + \frac{148.3^2}{35} + \frac{16.8^2}{6} + \frac{35.7^2}{9}\right] - \frac{204.8^2}{54} = 44.3$

2. $SS\_error = \left[4 - \frac{4^2}{4}\right] + \left[1048.4 - \frac{148.3^2}{35}\right] + \left[48.9 - \frac{16.8^2}{6}\right] +$

$$+ \left[146.4 - \frac{35.7^2}{9}\right] = 426.7$$

3. $MS\_cores = \frac{44.3}{3} = 14.8.$

4. $MS\_error = \frac{426.7}{50} = 8.53$

5. $\boxed{F = \frac{14.8}{8.53} = 1.74}$

The critical value of F, at the 5% level of significance with (3, 50) degrees of freedom, is $\boxed{F_{crit} = 2.79.}$

CONCLUSION: Since the calculated F value of 1.74 does not exceed the critical value of 2.79, we conclude that there is no statistical significant difference in the concentrations of beryllium in the four areas where sampling occurred.

In other compliance cores, levels of beryllium comparable to those measured in the compliance cores in the analysis given above were found, using the tolerance interval method, to exceed the tolerance limit. Because the four values obtained as background data at the 0-5 foot level exhibit no variation, the tolerance interval approach does not apply to measurements of beryllium made in compliance cores at that depth.

The analysis of variance test performed above produced a somewhat counter-intuitive result, since the average of beryllium values from the three compliance cores were 4.23, 2.8, and 3.97. These numbers would appear to indicate significantly higher concentrations of beryllium in the compliance cores as compared to the background data average of 1. A "Student's t" test of the hypotheses $H_0 : \mu = 1$ versus $H_1: \mu > 1$ leads to a rejection of the null hypothesis in each case of the compliance cores.

The reason for the failure of ANOVA to detect a statistically significant difference in the level of beryllium in the compliance cores could be attributed to the following factors:

1) There are many more values from the compliance cores as compared to the background well (50 vs 4). So, the weighting given to the data from the compliance cores is much greater than that given to the background data.
2) The values obtained from the compliance cores are quite similar.

If there had been, say, six more values of 1.0, or thereabouts, recorded in the background well, then the resulting F value in the analysis of variance would have exceeded the critical value and the conclusion would then be that the compliance cores have significantly higher levels of beryllium than the background well.

Another way to put the background data on a more even footing with the compliance cores would be to take a random sample of four values from the data collected for the compliance cores. Using a table of random digits, I did this and obtained the following sets of data:

Background Data:

1.0 1.0 1.0 1.0

Data from Evaporation Lagoons:

2.2 7.7 2.8 3.2

Data from Aeration Lag

2.8 2.6 2.7 3.5

Data from Diversion Ditch:

2.4 3.6 4.3 4.6

1. SS_cores = 21.8
2. SS_error = 22.4
3. MS_cores = 7.6
4. MS_error = 1.9
5. $F = \frac{7.6}{1.9} = 4.0$

The F critical value is $F_{0.5,\,(3,12)} = 3.49$.

Since 4.0 > 3.49, we conclude that there is a statistically significant difference in the levels of beryllium among the four cores.

In the document Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities - Interim Final Guidance, Section 5, *Background Well to Compliance Well Comparisons*, Neither a test of hypotheses (Student's t) nor an ANOVA on a 'reduced' data set were recommended. I therefore request the EPA's guidance regarding an acceptable statistical test. It is somewhat difficult to see what to do when the background data exhibits no variation and consists of just 4 values.

Mark Wilson
207 E Green
Gallup, NM 87301 (505) 722-2312

Mark Wilson