EPA 530/R-09-007a

ERRATA SHEET —MARCH 2009 UNIFIED GUIDANCE
August 9, 2010

The following are corrections made to a number of Example calculations and equations in the March 2009 Unified Guidance [*EPA 530/R-09-007*]:

Chapter 10, Example 10-1: In the initial table of Nickel concentrations, 'Years' has been changed to 'Wells' to maintain consistency with succeeding examples. At the bottom of page 10-11, the following sentence was added (to allow for pooling): "Assume that the individual well data sets can be shown to arise from a single common population."

Chapter 10, Example 10-4 Calculations for the Multiple Group Shapiro-Wilk Test (full revised example text in parentheses):

"The previous examples in this chapter pooled the data of **Example 10-1** into a single group before testing for normality. This time, treat each well separately and compute the Shapiro-Wilk multiple group test of normality at the $\alpha = .05$ level.

SOLUTION

Step 1. The nickel data in **Example 10-1** come from $K = 4$ wells with $n_i = 5$ observations per well. Using equation [10.10], the SW_i individual well test statistics are calculated as:

Well 1: $SW_1 = 0.7577$
Well 2: $SW_2 = 0.7396$
Well 3: $SW_3 = 0.7065$
Well 4: $SW_4 = 0.8149$

Step 2. Since $n_i = 5$ for each well, use **Table 10-7** of **Appendix D** to find $\varepsilon = .5521$. First calculating u_1 with equation [10.20]:

$$u_1 = \ln\left(\frac{.7577 - .5521}{1 - .7577}\right) = -.1641$$

Then performing this step for each well group and using linear interpolation on u in **Table 10-7**, the approximate G_i statistics are:

Well 1: $u_1 = -.1641$ $G_1 = -1.783$
Well 2: $u_2 = -.3280$ $G_2 = -1.932$
Well 3: $u_3 = -.6425$ $G_3 = -2.200$
Well 4: $u_4 = .3502$ $G_4 = -1.254$

Step 3.   Compute the multiple group test statistic using equation [10.21]:

$$G = \frac{1}{\sqrt{4}}[(-1.783)+(-1.932)+(-2.200)+(-1.254)] = -3.585$$

Step 4.   Since $\alpha = 0.05$, the lower $\alpha \times 100th$ critical point from the standard normal distribution in **Table 10-1** of **Appendix D** is $z_{.05} = -1.645$. Clearly, $G < z_{.05}$; in fact G is equivalent to a Z-value probability of .0002. Thus, there is significant evidence of non-normality in at least one of these wells (and perhaps all of them). ◄ "

Chapter 12, Example 12-1 Calculations for Screening with Probability Plots

In Figures 12-1 through 12-4 and the accompanying text, normality correlation coefficients have been adjusted (using the method in UG Section 10.6) as follows:

Figure 12-1 Raw Correlation Coefficient (N =20) -- .502

Figure 12-2 Log "   "   (N =20)   -- .973

Figure 12-3 Raw "   ", 1 outlier removed (N = 19) -- .854

Figure 12-4 Log "   ", 1 outlier removed (N = 19) -- .987

Chapter 13, Example 13-1 Tables of Iron Concentrations

p.13-3. The median should be 50.06 ppm for Well 1; for Well 2, the mean is 55.74 ppm.

For comparative purposes, well sample standard deviations have been added at the bottom of this table: Well 1—12.40; Well 2—20.34; Well 3—59.35; Well 4—25.95; Well 5—92.16; and Well 6—51.20 ppm

p.13-4 (log iron concentration table) The median for Well 1 is 3.91 log(ppm). Sample well log standard deviations are already found in the table on page 13-7.

Chapter 13, Example 13-2 ANOVA Calculations for Log Iron Concentrations (full revised example text in parentheses):

" SOLUTION

Step 1. With 6 wells and 4 observations per well, $n_i = 4$ for all the wells. The total sample size is $n = 24$ and $p = 6$. Compute the (overall) grand mean and the sample mean concentrations in each of the well groups using equations [17.1] and [17.2]. These values are listed (along with each group's standard deviation) in the above table.

Step 2. Compute the sum of squares due to well-to-well differences using equation [17.3]:

$$SS_{wells} = \left[4(3.820)^2 + 4(3.965)^2 + \ldots + 4(5.000)^2\right] - 24(4.354)^2 = 4.331$$

This quantity has $(6 - 1) = 5$ degrees of freedom.

Step 3. Compute the corrected total sum of squares using equation [17.4] with $(n - 1) = 23$ df:

$$SS_{total} = \left[(4.06)^2 + \ldots + (5.08)^2\right] - 24(4.354)^2 = 8.935$$

Step 4. Obtain the within-well or error sum of squares by subtraction using equation [17.5]:

$$SS_{error} = 8.935 - 4.331 = 4.604$$

This quantity has $(n - p) = 24 - 6 = 18$ degrees of freedom.

Step 5. Compute the well and error mean sum of squares using equations [17.6] and [17.7]:

$$MS_{wells} = 4.331/5 = .866$$

$$MS_{error} = 4.604/18 = .256$$

Step 6. Construct the F-statistic and the one-way ANOVA table, using **Figure 13-3** as a guide:

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares	F-Statistic
Between Wells	4.331	5	0.866	F = 0.866/0.256=3.38
Error (within wells)	4.604	18	0.256	
Total	8.935	23		

Step 7.   Compare the observed *F*-statistic of 3.38 against the critical point taken as the upper 95th percentage point from the *F*-distribution with 5 and 18 degrees of freedom. Using **Table 17-1** of **Appendix D**, this gives a value of $F_{.95,5,18} = 2.77$. Since the *F*-statistic exceeds the critical point, the null hypothesis of equal well means can be rejected, suggesting the presence of significant spatial variation.   ◄ "

Chapter 13, Example 13-3 ANOVA Pooled Variance Used for Prediction Limit Calculations for Log Iron Concentrations

With the slight modification in $\sqrt{MS_{error}} = \sqrt{.256} = .506$ from Example 13.2, as shown in Step 2 p.13-11, the resulting table with well-specific prediction limits is changed to:

	Adjusted 99% Prediction Limits for Iron (ppm)					
	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6
Log-mean	3.820	3.965	4.348	4.188	4.802	5.000
RMSE	0.5079	0.5079	0.5079	0.5079	0.5079	0.5079
df	18	18	18	18	18	18
$t_{.99,18}$	2.552	2.552	2.552	2.552	2.552	2.552
99% PL	193.2	223.3	327.5	279.1	515.8	628.7

Chapter 14, Section 14.2.2 Procedure for Estimating Sample Size for a prediction limit with significant temporal variation, Step 9 (modifies equations for total events):

" Step 9.   If there is no spatial variability but a significant temporal effect exists among a set of background wells, compute an appropriate *interwell* prediction or control chart limit as follows. First replace the background sample standard deviation (*s*) with the following estimate built from the one-way ANOVA:

$$\hat{\sigma} = \sqrt{\frac{1}{W}\left[MS_T + (W-1)\,MS_E \right]} \qquad [14.12]$$

Then calculate the effective sample size for the prediction limit as:

$$n* = 1 + \left\{\left[TK \cdot (TK-1) \cdot (F_T + W - 1)^2 \right] \Big/ \left[TK \cdot F_T^2 + (TK-1)\cdot(W-1) \right]\right\} \quad [14.13] \text{ "}$$

Chapter 14, Example 14-2, Steps 5, 7 and 8:

Although the final calculation is correct, the value for (W-1) in equation [14.7] is 3, not 7:

" Step 5.   Compute the mean error sum of squares term using equation [14.7]:

$$MS_E = \left[(-1.150)^2 + (-.780)^2 + \ldots + (1.338)^2 + (-.765)^2\right]\Big/(4\cdot 2)(3) = 1.87 \text{ "}$$

The corrected degrees of freedom language in Step 7 is:

" Step 7. Test for a significant temporal effect, computing the *F*-statistic in equation [14.11]:

$$F_T = 7.55/1.87 = 4.04$$

The degrees of freedom associated with the numerator and denominator respectively are (*TK*–1) = 7 and *TK*(*W*–1) = 24. Just as with Levene's test run earlier, the 5% level critical point for the test is $F_{.95,7,24}$ = 2.42. Since F_T exceeds this value, there is evidence of a significant temporal effect in the manganese background data. "

In Step 8, the estimated adjusted standard deviation is:

$$\hat{\sigma} = \sqrt{\frac{1}{4}\left[7.55 + 3 \cdot (1.87)\right]} = 1.814 \; ppm$$

Chapter 14, Example 14-7, Step 2:

Although the final calculation is correct, the value for N_e should be a decimal:

" Step 2. To estimate the minimum time interval between sampling events enabling the collection of physically independent samples of ground water, calculate the horizontal component of the average linear groundwater velocity (V_h) using Darcy's equation [14.17]. With K_h = 15 ft/day, *Ne* = .15 (15%), and *i* = 0.003 ft/ft, the velocity becomes:

$$V_h = (15 \; ft/day \times .003 \; ft/ft)/.15 = .3 \; ft/day \; or \; 3.6 \; in/day \quad "$$

Chapter 16, Example 16-3, Shapiro-Wilk Calculations in Steps 1 and 2:

The G multiple group value in Step 1 is -6.671 using original benzene data. The corresponding value for the log transformed data is G = -.512 in Step 2. Other results are correct.

Chapter 17, Example 17-6, Mann-Kendall Test Calculations, Steps 2 & 3:

In Step 2, the Mann-Kendall statistic S =194, not 196.   The standard deviation calculation of 37.79 in Step 3 is correct.   The modified S changes the Z-approximation to 5.11 in Step 3:

$$Z = (|194| - 1)/37.79 = 5.11$$

Chapter 18, Example 18-2, Statistics in Table, and Step 2:

The joint well log mean for chrysene is 2.553, not 2.533 ppb. This changes the future mean prediction limit calculation in Step 2:

$$PL = 2.553 + (2.998)(.706)\sqrt{\frac{1}{4}+\frac{1}{8}} = 3.85 \log((ppb))$$

Chapter 19, Example 19-2, Step 4:

The median value of the initial three data for well CW-2 is .36, not .41 ppb. Other results are not affected.

Chapter 21, Example 21-7, Step 4:

The correct value for (n-1) in the denominator should be 9, not 8 as applied with equation [21.25]. This changes the UCL calculation results as follows:

" Step 4. Since the comparison to the GWPS of 20 ppb is to be made at the $\alpha = 0.05$ significance level, the confidence limit is $(1-\alpha) = 95\%$ confidence. Since the remediation effort aims to demonstrate that the true mean TCE level has dropped below 20 ppb, a one-way UCL needs to be determined using equation [21.25]. A logical point along the trend to examine is the last sampling event at $t_0 = 30$. Using the estimated regression value at $t_0 = 30$, and the fact that $F_{.90,2,8} = 3.1131$, the UCL on the mean TCE concentration at this point becomes:

$$UCL_{95} = 6.861 + \sqrt{2 \times 15.60 \times 3.1131 \times \left[\frac{1}{10}+\frac{(30-15.3)^2}{9 \times 88.2333}\right]} = 12.87 \; ppb$$

Since this upper limit is less than the GWPS for TCE, conclude that the remediation goal has been achieved by $t_0 = 30$. In fact, other times can also be tested using the same equation. At the next to last sampling event ($t_0 = 26$), the UCL is:

$$UCL_{95} = 13.272 + \sqrt{2 \times 15.60 \times 3.1131 \times \left[\frac{1}{10}+\frac{(26-15.3)^2}{9 \times 88.2333}\right]} = 18.14 \; ppb \; "$$