Problem 1: Sample Size for the Estimation of Mean of Sampled Population

Muddy Brook will be sampled to develop a best estimate of mean phosphorus concentration to help in the development of a comprehensive watershed management plan. Based on the existing monitoring data from Muddy Brook, calculate how many more samples need to be collected to be within 10 and 20 percent of the estimated true mean TP concentration.

Existing data from Muddy Brook:
Mean = 0.89 mg/L
Std Dev. = 0.62 mg/L
n = 165

Results:
The ½ confidence interval (d) for 10% and 20% would be:
d = 0.10 x 0.89 = 0.09 mg/L
d = 0.20 x 0.89 = 0.18 mg/L

The two-tailed t value for 164 d.f. at *P* = 0.05 is 1.97

For a 10% difference:

$$n = \frac{(1.97)^2 \times (0.62)^2}{(0.09)^2} = 184$$

Because the t value for n=184 does not change, additional steps are not necessary. Note, if using this formula to estimate the total number of samples, 184 samples a year would be required to obtain an estimate of mean TP concentration that was $\pm$10% of the true mean of the sampled conditions. Taking 184 samples in a single year may introduce excessive autocorrelation (which reduces the effective sample size - see section 3.4.2). Weekly samples will probably still exhibit some autocorrelation, but would be an improvement over biweekly sampling. Dividing by 52 yields about 3.5 years which means this is probably unrealistic for most problem assessment efforts.

For a 20% ½ confidence interval:

$$n = \frac{(1.97)^2 \times (0.62)^2}{(0.18)^2} = 46$$

Because the value of t at 45 d.f. is 2.01, a second iteration is necessary:

$$n = \frac{(2.01)^2 \times (0.62)^2}{(0.18)^2} = 48$$

Because the t value for n=48 does not change, additional steps are not necessary. Therefore, 48 samples should be taken to estimate the annual mean TP concentration within 20%. This number of samples is approximately equal to biweekly samples over 2 years.