of the necessary steps, and make your answer culminate in a statement similar to 'There is sufficient/insufficient evidence to conclude that (state a conclusion in the context of the problem).'" The major mistake here is that students believe that constructing side-by-side boxplots, or a histogram or other plot, or comparing percentages is sufficient to show that a new drug, for example, is more effective than the old, established drug.

The confidence interval seems like such a simple idea, but it can cause a surprising amount of potential trouble for students. Confidence interval questions on the exam typically have been in two parts. The first part asks students to use information provided to construct a confidence interval. The second part is to interpret the confidence interval. Typically, students do well on the first part, and this is not surprising. After all, most of them enroll in AP Statistics because they have performed well in previous math classes, many of which involve lots of computation. Where students frequently crash and burn is on the interpretation part. Interpretation requires a sentence or two that explains what the confidence interval is all about.

Interpretations of confidence intervals are sufficiently important that they can raise the points on that question from, say, a 3.5 to a 4, or if the interpretation is botched, lower the points to a 3 when holistic grading is applied to a student response. Under certain circumstances, a student can employ a confidence interval approach for a question specifically asking for a significance test, and if done correctly, this strategy will receive full credit.

The Inference Steps

Let's begin with a test of significance: inference for a mean or a proportion, comparing two means or proportions, a chi-square test, or inference for regression.

1. **State hypotheses.** Every inference procedure should begin with a clear statement of the null and alternative hypotheses. The best response here is to state the hypotheses using symbols and then write out statements of the hypotheses in words. If the symbols approach is used alone, and commonly used symbols for the population parameters are used, then this would be considered acceptable. If nonroutine symbols are used, then they must be defined clearly and correctly in order to receive full credit.

2. **Identify the inference procedure, and state and check assumptions and conditions.** Inference procedures produce valid results only if certain conditions are satisfied (like hypotheses for a theorem in geometry). The failure to check

assumptions was a glaring omission for many students in the early days of the AP Exam. When this was identified as a failing, students began stating the assumptions, such as "$np \geq 10, n(1-p) \geq 10$," and left it at that. While this was an improvement over cavalierly ignoring assumptions, such statements alone did not receive credit! Such constructions were regarded as stating the assumptions, not checking them. In order to get credit for this step, students must state the assumptions correctly and then substitute numbers for that problem to actually verify their validity so that the inference procedure could be legitimately performed. Floyd Bullard's article in these materials gives additional insight about the role of assumptions in inference.

3. **Mechanics.** The next step in the inference procedure is to calculate the test statistic, using the appropriate formula. (Formulas are provided with the exam.) It is best to write the appropriate formula for the test statistic and then show substitution into that formula. The computation can then be performed on a calculator and the test statistic reported. The test statistic value should be used to determine the *p*-value (the probability of obtaining a result as extreme or more extreme than the value we obtained, by chance alone, if the null hypothesis were true). The *p*-value is then used to make the decision to reject or fail to reject the null hypothesis.

4. **Conclusion.** If the *p*-value is sufficiently small to cause rejection of H_0, then a conclusion needs to be made *in the context of the problem*. In the early days of the AP Statistics Exam, some students believed that a statement of "reject" or "fail to reject" was the moral equivalent of QED, and that nothing else needed to be said. Not so! In statistics, context is critical.

Summarizing, the four required parts in an inference setting are:

1. State hypotheses in words and/or symbols.

2. Identify the correct inference procedure and verify conditions for using it.

3. Calculate the test statistic and the *p*-value (or rejection region).

4. Draw a conclusion in context that is directly linked to the *p*-value or rejection region.