ILLUSTRATION 4.1

How Confident Is "Confident Enough"?

When a hypothesis is true, the level of significance expresses the probability of making the wrong decision (rejection), and the level of confidence expresses the probability of making the correct decision (fail to reject). We have emphasized that the choice of significance level in a statistical analysis depends on the judgment of the analyst and the perceived consequences of making an error. While the choice of significance level also determines the confidence level—the two probabilities must sum to one—decision makers tend to focus their attention on choosing a sufficiently high level of confidence, rather than thinking of choosing a sufficiently low level of significance. We want to tell you a fictional story to better illustrate how the consequences of a wrong decision can affect the choice of confidence or significance levels.

During the first days of the Persian Gulf war, the world was stunned by the success rate of the bombing attacks against military targets in Iraq. In his first news briefing at the outbreak of the war, General Norman Schwarzkopf reported that 80 percent of the nearly 15,000 sorties flown had successfully hit their intended targets. Many of the correspondents in attendance were highly suspicious about such an extraordinary success rate, but two reporters, Barbara Smith and Heraldo Jones, decided to test the general's assertion that the true success rate was 80 percent. They obtained a list of the locations of 100 of the 15,000 targets, then enlisted the help of a pilot of a three-seater Stealth fighter to fly them over each of the 100 targets to see how many were damaged by bombs. Smith and Jones counted 65 bomb-damaged targets, indicating only a 65 percent success rate, as opposed to the reported 80 percent.

Each reporter had to decide whether 65 percent was far enough away from 80 percent to refute the Schwarzkopf assertion. Both realized the inherent randomness of sampling and that the 100 targets may not have exactly reflected the population of 15,000 targets, known only to the general and his staff.

Smith was tempted to report to her news director that she had discovered compelling evidence that General Schwarzkopf incorrectly reported the success rate of the air war. If correct, she would probably become famous and win a Pulitzer Prize; but if wrong, she would be ruined professionally. Smith was not willing to take more than a 5 percent chance of committing a Type I error: publicly rejecting the Schwarzkopf assertion of an 80 percent success rate when the assertion was correct. In other words, she had to be 95 percent confident that, if Schwarzkopf was correct, her "test" would not reject Schwarzkopf's assertion. Smith decided that 65 percent was not far enough below 80 percent to make her feel 95 percent confident that 80 percent was an exaggeration. Therefore, she did not report her findings.

Jones, in contrast to Smith, badly wanted to be an anchor. He was willing to bet his career by taking a 75 percent risk that he was wrong and the general was right; that is, he was comfortable with only a 25 percent level of confidence. At the 25 percent level of confidence, Jones viewed a 65 percent success rate as being far enough away from the asserted 80 percent success rate that he rejected the general's assertion. Jones called his news anchor with startling evidence that General Schwarzkopf had misinformed the public about the success of the air war. Jones's network reported his story on national news, and upon hearing this news report, General Schwarzkopf decided to reveal the list of the initial 15,000 sorties. Using the entire population of 15,000 targets, the Middle East press corps verified that 80 percent of the targets were indeed damaged by bombs. As it turned out, Barbara Smith was later promoted to anchor at her network. Heraldo Jones was fired and now hosts a talk show at an obscure radio station in College Station, Texas.