Another Example of Hypothesis Testing

To make things a little faster (you won't have to write down anything I put on the board, but you should still fill in what I say) I'm providing these notes. This example is similar to the coin thumping experiment, but forgoes the data collection phase in favor of "canned" observations.

1 The Problem

In a discussion of SAT scores, someone comments: "Because only a minority of high school students take the test, the scores overestimate the ability of the typical high school seniors. The mean SAT mathematics score is about 475 with a standard deviation of 100, but I think that if all seniors took the test, the mean score would be no more than 450." In an attempt to show the "sceptic" that he is wrong, the "experts" give the test to an SRS of 500 seniors from California. These students have a mean score of $\bar{x} = 461$. Does this support the experts' claim and refute the sceptic's hypothesis?

2 Hypotheses

The null and alternative hypotheses may be stated in words.

- The *alternative* hypothesis is what we wish to show; in this case the claim of the experts that the mean SAT-M score is at least 450.
- The *null* hypothesis is the opposite of what we wish to show; that the mean SAT-M score is less than 450.

The "mathese" equivalents of these statements are:

$$\begin{aligned} H_0 : \quad \mu &\leq 450 \\ H_a : \quad \mu &> 450 \end{aligned}$$

3 P-value

We now give the "sceptic" the best possible chance and assume that $\mu = 450$ (*i.e.* the null hypothesis) is true. The probability of seeing what we saw or something more extreme is the probability of seeing a mean which is greater than 461. In "mathese" this is

$$\mathrm{P}(\bar{X} > 461).$$

Now, by the information given above, under the null hypothesis $X \sim N(450, 100^2)$. The CLT tells us that $\bar{X} \sim N(450, 100^2/500)$.

Unfortunately, we don't have a $N(450, 100^2/500)$ table, so we need to standardize.

$$\begin{aligned} \mathrm{P}(\bar{X} > 461) &= \mathrm{P}\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{461 - 450}{100/\sqrt{500}}\right) \\ &= \mathrm{P}(Z > 2.46) \end{aligned}$$

Using Table A we see that

$$\begin{aligned} \mathrm{P}(Z > 2.46) &= 1 - \mathrm{P}(Z \leq 2.46) \\ &= 1 - 0.9931 \\ &= 0.0069 \end{aligned}$$

4 Testing

Now, personally, I think that seeing something that should only happen about 7 times out of 1000 is a little strange. So, something must be wrong. We have two possibilities:

- We were unlucky and got a bad sample in our SRS.
- Assuming the null hypothesis was a mistake and the alternative must be true.

SRS's usually don't lie. Thus, the null hypothesis must have been wrong. Hence, I reject the null hypothesis and conclude that the mean SAT-M score must be greater than 450. I do this knowing that there is a 7 in 1000 chance of having made a mistake.

In many cases a mistake in 5 of 100 trials is acceptable. In sociology experiments I might even accept 1 in 10 if nobody is getting hurt. However, if I'm dealing with human hearts and mistakes mean lives, my value would have to be closer to 1 in 100 or maybe 1 in 1000. The value that I choose to compare my p-value against is called the test's *level*, α. Since $0.001 < 0.0069 < 0.01$, we would reject the null hypothesis if $\alpha = 0.01$ (a mistake in 1 out of 100 trials) but not if $\alpha = 0.001$ (a mistake in 1 out of 1000 trials).