The Effective Use of Statistical Tests

Connection Between Confidence Intervals and Testing

We now know about confidence intervals and hypothesis testing. In this lecture, we are going to consider the connection between them.

At first it seems like there really is not much of a connection because their purpose is pretty different. Hypothesis testing is when we want a yes or no answer, whether we can reject the null hypothesis or fail to reject it. Confidence intervals tell us a range of plausible values for the quantity. We have seen examples where we have done one-sided hypotheses tests and two-sided hypotheses tests, but we have only done two-sided confidence intervals. But despite all that, there is one nice clear connection between two-sided hypothesis tests and two-sided confidence intervals that we are now going to explore.

Example 1

Let us consider again flipping that bottle cap and observing either red or silver. We wanted to know the probability that it will come up red.

Recall that we wanted to test the null hypothesis that the probability of getting red was 0.5 against the alternative hypotheses that it was some other value besides 0.5. We obtained a very small p-value which lead us to reject the null hypothesis and conclude that the proportion of red outcomes is not 50%. Previously we constructed a 95% confidence interval for the true proportion of red outcomes and that gave us a confidence interval from 0.532 to 0.620, and we notice that that confidence interval misses the value 0.5. We seemed to have gotten a similar conclusion in two different ways and this is not just a coincidence. We are going to see that this equivalence happens all the time.

Suppose were testing a two-sided hypotheses

$$H_0 : \mu = \mu_0 \text{ vs } H_a : \mu \neq \mu_0.$$

Our data is size n, the sample mean is $\bar{x}$, the sample standard deviation is s, and the observed difference $d = |\bar{x} - \mu_0|$. We would reject the null hypothesis H_0 if

$$\begin{aligned}
&P(|\bar{X} - \mu_0| \geq d) \leq \alpha \\
&\Rightarrow P\left(\left|\frac{\bar{X} - \mu_0}{\sqrt{s^2/n}}\right| \geq \frac{d}{\sqrt{s^2/n}}\right) \leq \alpha \\
&\Rightarrow P\left(|t_{n-1}| \geq \frac{d}{\sqrt{s^2/n}}\right) \leq \alpha
\end{aligned}$$

This happens only if $\frac{d}{\sqrt{s^2/n}} \geq T_{\alpha/2,n-1}$ where $T_{\alpha/2,n-1}$ is a **critical value** such that $P(|t_{n-1}| \geq T_{\alpha/2,n-1}) = \alpha$. You can think of a critical value as being a value where we would be right on the edge of rejecting the null hypothesis or not; if the observed difference is more than this critical value then we will reject it, and if the observed difference is less than this critical value then we will not reject it.

On the other hand, if we consider a $(1 - \alpha) \times 100\%$ confidence interval for μ given by $\bar{x} \pm T_{\alpha/2,n-1}\sqrt{s^2/n}$, this will miss μ_0 when

$$
\begin{aligned}
&|\bar{x} - \mu_0| \geq T_{\alpha/2,n-1}\sqrt{s^2/n} \\
&\Rightarrow \frac{d}{\sqrt{s^2/n}} \geq T_{\alpha/2,n-1}.
\end{aligned}
$$

Thus, it turns out that two-sided hypothesis tests are equivalent to confidence intervals, keeping our significance level α fixed. That is, if we reject the null hypothesis or if we say that the corresponding confidence interval does not include the null hypothesized value, we are making the same statement.

There is still a different way of thinking. A hypothesis test is deciding yes or no, whether or not we are going to reject the value. The confidence interval is giving a range of plausible values. But they do have this connection, that the range of plausible values will include certain particular values if and only if that particular value would not have been rejected by the corresponding hypothesis test. This gives us a way to think about these two different subjects in a similar way.