The Process of Statistical Tests

Hypothesis Testing for Means

We now know how to test hypotheses and compute p-values when we are talking about a probability or a proportion. But what if we wanted to do inference about mean values? Well, it turns out we can still do hypothesis tests and it's really not much harder.

Example 1

Recall Example 3 from Video 24, where we obtained data from 60 patients who had undergone plastic surgery. We were interested in whether the perceived age changes, so for each patient we have the difference in their perceived age from before to after the surgery. For our sample in particular, patients seemed to look younger post-surgery. Does that mean that for patients everywhere, on average surgery is going to make them look younger? Or were these patients just lucky.

For our sample of 60 patients, we can calculate the sample mean $\bar{X} = 7.1777$ and standard deviation $s = 2.948$. We know in the theoretical world, we have the true mean number of years that this plastic surgery makes you look younger, μ, which is unknown. We also have the estimated number of years that it makes you look younger, and that is our $\bar{X} = 7.177$.

We would like to test whether or not μ is greater than zero, so we have the following null and alternative hypothesis

$$H_0 : \mu = 0 \text{ vs } H_a : \mu > 0$$

In the theoretical world, assuming $H_0 : \mu = 0$ is true, we have that

$$\frac{\bar{X} - \mu}{\sqrt{s^2/n}} \approx t_{n-1},$$

where t_{n-1} is the t distribution with $n - 1$ degrees of freedom. What we have observed is that

$$\bar{X} - \mu = 7.177 - 0 = 7.177$$

Thus, the p-value should be the probability of observing such an extreme value under the null hypothesis.

$$\begin{aligned}
p\text{-value} &= P(\text{observing such an extreme value}|H_0) \\
&= P(\bar{X} - \mu \geq 7.177) \\
&= P\left(\frac{\bar{X} - \mu}{\sqrt{s^2/n}} \geq \frac{7.177}{\sqrt{2.948^2/60}}\right) \\
&\approx P\left(t_{59} \geq 18.86\right) \\
&\doteq \frac{1}{100,000,000,000,000,000,000,000,000}
\end{aligned}$$

Since this p-value is *extremely* small, we can reject the null hypothesis and conclude that this plastic surgery did indeed make people look younger on average.

Example 2

Recall Example 4 in Video 24, where there was a long held belief that the average human normal body temperature is equal to 37° Celsius, based on Wunderlich's results. A follow-up study in 1992 obtained the temperature of 148[1] healthy men and women and the researchers wondered if their data confirmed or denied Wunderlich's claim. Their summary statistics were $\bar{X} = 36.80513$ and $s = 0.407324$.

To set this up as an experiment, we can say that the null hypothesis is that the true normal human body temperature is equal to 37°, and the alternative hypothesis is that it is something else besides 37°. In this case we do not have any prior knowledge of whether it is more or less, so we will do a two-sided test.

$$H_0 : \mu = 37 \text{ vs } H_a : \mu \neq 37$$

In the theoretical world, assuming $H_0 : \mu = 37$ is true, we can state that

$$\frac{\bar{X} - \mu}{\sqrt{s^2/n}} \approx t_{129}$$

Thus, we can calculate the p-value to be:

$$\begin{aligned}
p\text{-value} &= P(|\bar{X} - \mu| \geq 0.19487) \\
&= P\left(\left|\frac{\bar{X} - \mu}{\sqrt{s^2/n}} \geq \frac{0.19487}{\sqrt{0.407324^2/130}}\right|\right) \\
&\approx P\left(|t_{129}| \geq 5.455\right) \\
&= 2 \times P(t_{129} \leq -5.455) \\
&\doteq \frac{1}{4,000,000}
\end{aligned}$$

[1]Since the original data is not available for this study, we will be using a simulated dataset created by Shoemaker which only contains n=130 subjects. Shoemaker, A. (1996) What's Normal? – Temperature, Gender, and Heart Rate. Journal of Statistics Education4(2)

The p-value calculated is about one chance in 4 million which is really small. We can reject the null hypothesis and say that the average normal human body temperature is not equal to 37° Celsius.

From the examples above, we see that we can do hypothesis tests and compute p-values even when we are talking about a mean or an average value. And once again, we can decide when to reject the null hypothesis, whether or not the mean values are bigger than or smaller than or equal to certain values.