The Process of Statistical Tests

Hypothesis Testing for Proportions

So far we have covered the basics about hypothesis testing and p-values. In the next few examples, we will begin to construct hypothesis tests for proportions.

Example 1

Recall an earlier poll we discussed, where the current mayor of Toronto was supported by only 42 percent of the respondents in a poll. We might be tempted to say that less than half of the people support the mayor. The question is, would that be a fair conclusion? Can we conclude from the result of that poll that indeed less than half of all the residence of the city actually support the mayor?

Let us start by remembering that the number of people surveyed was $n = 1,046$ and the percentage of them which said "yes I support the mayor" was equal to 42%. The true support of the mayor, p, is unknown, but the estimated support of the mayor from this sample is $\hat{p} = 0.42$.

We want to test as a hypothesis whether the true proportion p of people supporting the mayor is indeed less than 50%. Therefore we have

$$H_0 : p = 0.5 \text{ vs } H_a : p < 0.5$$

In words, the null hypothesis states that half the people support the mayor versus the alternative hypothesis that less than half of the people support the mayor.

We should start by remembering that in the theoretical world, assuming the null hypothesis $H_0 : p = 0.5$, we have

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx Normal(0, 1)$$

If we fill in the values for $p = 0.5$ and $n = 1,046$ we arrive at

$$\frac{\hat{p} - p}{\sqrt{0.5(1-0.5)/1046}} \approx Normal(0, 1)$$

What we observed under the null hypothesis would be

$$\hat{p} - p = 0.42 - 0.5 = -0.08$$

To compute the p-value, we want to compute the probability of observing a value that extreme or more extreme under the null hypothesis, to obtain

$$\begin{aligned}
p\text{-value} &= P(\text{observing such an extreme value}|H_0) \\
&= P(\hat{p} - p \leq -0.08) \\
&= P\left(\frac{\hat{p} - p}{\sqrt{0.5(1-0.5)/1046}} \leq \frac{-0.08}{\sqrt{0.5(1-0.5)/1046}}\right) \\
&\approx P\left(\text{Normal(0,1)} \leq -5.17\right) \\
&\doteq \frac{1}{9,000,000}
\end{aligned}$$

The test statistic of -5.17 is a pretty extreme value for the standard normal distribution so we can say the p-value is approximately 1 in 9 million, and that's very small! This p-value is so small that we can reject the null hypothesis H_0 and conclude that no, it is not true that the mayor's support is 50%, it is indeed true that it is less than 50%.

Example 2

Suppose we were interested in whether the mayor's support is less than 44%. We can proceed exactly as before to obtain our null and alternative hypothesis for this one-sided test.

$$H_0 : p = 0.44 \text{ vs } H_a : p < 0.44$$

In the theoretical world, assuming the null hypothesis $H_0 : p = 0.44$, we have

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx Normal(0, 1)$$

If we fill in the values for $p = 0.44$ and $n = 1,046$ we arrive at

$$\frac{\hat{p} - p}{\sqrt{0.44(1-0.44)/1046}} \approx Normal(0, 1)$$

What we observed under the null hypothesis would be

$$\hat{p} - p = 0.42 - 0.44 = -0.02$$

To compute the p-value, we want to compute the probability of observing a value that extreme or more extreme under the null hypothesis, to obtain

$$
\begin{aligned}
p\text{-value} &= P(\text{observing such an extreme value}|H_0)\\
&= P(\hat{p} - p \leq -0.02)\\
&= P\left(\frac{\hat{p}-p}{\sqrt{0.44(1-0.44)/1046}} \leq \frac{-0.02}{\sqrt{0.44(1-0.44)/1046}}\right)\\
&\approx P\left(\text{Normal(0,1)} \leq -1.30\right)\\
&\doteq 0.0968
\end{aligned}
$$

The test statistic calculated above of -1.30 is actually not that unlikely, as shown by the p-value of 0.0968. And depending on our level of significance, we probably cannot reject the null hypothesis. We were sure that the mayor's support is less than 50% but we cannot conclude that it is less than 44%.

Example 3

Recall the example of flipping that beer cap and trying to figure out if the proportion of the red side up was 50% or not. In this case, p is the true proportion of the red side up and it is unknown to us. In our experiment, we got 576 red flips out of 1000, making our observed estimate $\hat{p}$ equal to 0.576.

We will begin by stating our null and alternative hypothesis:

$$H_0 : p = 0.5 \text{ vs } H_a : p \neq 0.5$$

Notice that this is a two-sided test; we are not trying to figure out if it is strictly less or strictly more than 0.5, just if the proportion is different than 0.5.

Under H_0 we have observed that

$$\hat{p} - p = 0.576 - 0.5 = 0.076$$

The p-value now is the probability of observing in absolute value, because it is a two sided test, an observation equal to or more extreme under the null hypothesis.

$$
\begin{aligned}
p\text{-value} &= P(\text{observing such an extreme (absolute) value}|H_0)\\
&= P(|\hat{p} - p| \geq |0.076|)\\
&= P\left(\left|\frac{\hat{p}-p}{\sqrt{0.5(1-0.5)/1000}}\right| \geq \frac{0.076}{\sqrt{0.5(1-0.5)/1000}}\right)\\
&\approx P\left(|\ Normal(0,1)| \geq 4.81\right)\\
&= 2 \times P(Normal(0,1) \leq -4.81)\\
&\doteq \frac{1}{663,000}
\end{aligned}
$$

Since the p-value is extremely small, we can reject H_0 and conclude that when you flip that beer cap, the proportion of red outcomes is not 50%, even though the fraction was pretty close to 50%.

In the examples above, we were able to test hypotheses for proportions or probabilities by rewriting the equation and computing the p-value. In this way, we know when to reject the null hypothesis and when we cannot reject the null hypothesis. This allows us to form sound statistical conclusions about probabilities and proportions.