How Hypothesis Testing Works

This article is meant to give a deeper understanding of how a Hypothesis Test works. Before reading this, you should already have some experience with doing Hypothesis Testing.

The pValue answers a strange question
Say you want to show that a certain population proportion is a majority, \( p > .50 \). You get a simple random sample of size \( n = 36 \). You hope that your sample proportion \( \hat{p} \) is significantly greater than .50. Say your sample proportion is \( \hat{p} = .78 \). This is higher than .50, but is it significantly higher? The answer depends on your sample size, so it’s not obvious.

We determine whether .78 is significantly higher than .50 by asking a strange question.

“What's the probability that my sample has a \( \hat{p} \) as high as .78 if the actual population proportion is \( p = .50 \)?”

This probability is easily calculated. It’s called the **pValue**. It corresponds to the shaded area under the curve of all possible samples shown in the figures. We think of the pValue as the probability that you picked the sample you did out of all the possible samples of the same size that you could have picked.

- **pValue is high**
The shaded area is large and .78 is close to .50. In this case, **.78 is not significantly higher than .50**. Even though you picked a sample with a sample proportion of \( \hat{p} = .78 \), it’s entirely likely that the actual population proportion is \( p = .50 \) and your sample "just happened by chance". We have **no evidence** that \( p > .50 \). You **cannot reject the null hypothesis** \( H_0 : p = .50 \) in favor of the alternate hypothesis \( H_1 : p > .50 \).

- **pValue is low**
The shaded area is small and .78 is far away from the .50. In this case, **.78 is significantly higher than .50**. It's very unlikely that picked a sample with \( \hat{p} = .78 \) if the real population proportion is \( p = .50 \). So, we conclude that the real population proportion is not \( p = .50 \). In fact, we conclude that the actual population proportion \( p \) must be higher than .50. (That would move the normal bump to the right making the shaded area bigger and your sample more likely to have happened.) We do **have evidence** that \( p > .50 \). You **can reject the null hypothesis** \( H_0 : p = .50 \) in favor of the alternate hypothesis \( H_1 : p > .50 \).
**Definition of the pValue**
The previous example suggests the following definition for the pValue.

The pValue can be thought of as the probability that $H_0$ is true.

- If the pValue is high, then the null hypothesis $H_0$ is probably true and there's no evidence to support the alternate hypothesis $H_1$.
- But if the pValue is low, then the null hypothesis $H_0$ is probably false and you do have evidence that supports the alternate hypothesis $H_1$.

Most often, a significance level of $\alpha = 5\%$ is used as a pass-fail cutoff for the pValue. However, the pValue can indicate evidentiary support for $H_1$ over a range of strengths.

<table>
<thead>
<tr>
<th>pValues and Evidentiary Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>under 1%</td>
</tr>
<tr>
<td>1% to 5%</td>
</tr>
<tr>
<td>5% to 10%</td>
</tr>
<tr>
<td>10% to 20%</td>
</tr>
<tr>
<td>above 20%</td>
</tr>
</tbody>
</table>

**Example**
You are the owner of the Xsort Corporation. To test your claim that the Xsort treatment improves the chances of a woman having a baby girl, you treat a sample of 100 women and they give birth to 100 babies. Consider these two possible outcomes for the number of baby girls.

- 52 girls
- 97 girls

a. As the owner of the Xsort Corporation, which outcome do you want to happen and why?
b. Which outcome has a low pValue and which outcome has a high pValue? Explain.

**Solution**
a. As the owner of the Xsort Corporation, I want 97 girls since that is common sense evidence that my Xsort treatment is really working.
b. If the outcome is 52 girls, then it's likely that Xsort doesn't work and I just got the sample by chance. So, the pValue would be high.
   If the outcome is 97 girls, then it's unlikely that Xsort doesn't work and I just got the sample by chance. So, the pValue would be low.

In the example above, the null hypothesis is $H_0: p = .50$. This says that Xsort doesn't work. So, the pValue measures the likelihood that Xsort doesn't work. Since I work for Xsort Corporation, I want that likelihood to be low. I want a low pValue.