Hypothesis Test for Population Proportion

What it is

A hypothesis test for p determines whether observed data is consistent with a hypothesized population proportion. If you decide the data is inconsistent with the hypothesized proportion, you can reject that proportion as the true parameter. If you decide the data is consistent with the hypothesized proportion, you have not proven that it is in fact the true value. You can only conclude that *you cannot say it is not* the true value.

When to use it

The data conditions for this hypothesis test are similar to those for the confidence interval for p (a sample with a known size of independent binary outcomes). In addition, there should be a specific numeric value of interest (e.g., $p = 0.5$, or a CDC vaccination coverage goal) to use as the null hypothesis. If there is no specific value of interest, a confidence interval is probably more appropriate.

How to use it

In R, use the function `binom.test()`. Like with the confidence interval for p, we need to get the sample size and the number of successes. Suppose we want test whether the data is approximately 50% female[1]. The null hypothesis would be $H_0 : p = 0.5$, so we set the inputs `p=0.5` and `alternative='two.sided'` for the test:

[1] This is a common way to test representativeness: if the data is not consistent with a population with 50% females, the sample is potentially biased.

```
table(mydat$sex)

##
##   Male Female
##    881   1119

binom.test(x=1119, n=2000, p=0.5, alternative='two.sided', conf.level=0.95)

##
##  Exact binomial test
##
## data:  1119 and 2000
## number of successes = 1119, number of trials = 2000, p-value =
## 1.124e-07
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.5374187 0.5814066
```

```
## sample estimates:
## probability of success
##                 0.5595
```

We see that about 56% of the sample is female. The p-value, $1.12 \times 10^{-7} = 0.000000112$, tell us that if we truly took a random sample of 2,000 adults from a population that was 50% female, there is only about a 1 in 10,000,000 chance that we would get such an imbalanced sample[2]. This tells us either (1) the true proportion of females in the population is not 0.5, (2) we have a non-representative sample, or (3) we just happened to see that 1 in 10 million chance in this particular sample. We usually discount the last option since it would be such an unlikely outcome. Our decision between (1) and (2) would depend on our confidence in the representativeness of our sample.

[2] See also that the confidence interval is completely above 0.5.

Had the p-value been greater than 0.05 (the conventional cutoff), we would have simply said that our data is consistent with a representative sample from a population with is 50% female. (Again, this is not proof the sample truly is representative or that the population is exactly 50% female, just that we don't have evidence to claim otherwise.)

Why it works

The hypothesis test begins by assuming $p = p_0$. If this is actually true, then the Central Limit Theorem tells us that $\hat{p}$ is normally distributed[3] with mean p_0 and variance $\frac{p_0(1-p_0)}{n}$. Then we know what values of $\hat{p}$ to expect from a sample.

[3] Using mathematical shorthand, $\hat{p} \sim \mathcal{N}\left(p_0, \frac{p_0(1-p_0)}{n}\right)$

Since about 95% of values are within 2 standard errors of p_0, anything outside the range $p_0 - 2\sqrt{\frac{p_0(1-p_0)}{n}}$ to $p_0 + 2\sqrt{\frac{p_0(1-p_0)}{n}}$ is "inconsistent" and can be taken as evidence that $p \neq p_0$. These values of $\hat{p}$ far from p_0 correspond exactly to values of $\hat{p}$ that will give small p-values.