Hypothesis Testing For A Proportion

Colby Community College

Example 1

The following question comes from a book written by Hans Rosling, Anna Rosling Rönnlund, and Ola Rosling called *Factfulness*.

How many of the world's 1 year old children today have been vaccinated against some disease?

(a) 20% (b) 50% (c) 80%

What is your answer (or guess)?

The correct answer is 80%.

Note

If we take a multiple choice test, then we might like to distinguish between the two possibilities:

- People never learn these particular topics and their responses are simply equivalent to random guessing.
- People have knowledge that helps them do better than random guessing, or perhaps, they have false knowledge that leads them to actually do worse than random guessing.

Definition

In statistics, a **hypothesis** is a claim or statement about some property of a population.

Definition

A **hypothesis test** (or **test of significance**) is a procedure for testing a claim about some property of a population.

Definition

The **null hypothesis** (H_0) represents a skeptical perspective or a claim to be tested.

Definition

The **alternative hypotheses** (H_A) represents an alternative claim under consideration and is often represented by a range of possible parameter values.

Example 2

A US court considers two possible claims about a defendant: they are either innocent or guilty.

Which would be H_0 and which H_A?

The jury considers whether the evidence is so convincing that there is no reasonable doubt regarding the person's guilt.

$$H_0 : \text{Innocence}$$
$$H_A : \text{Guilty}$$

Note

Just because the jurors leave unconvinced of guilt beyond a reasonable doubt, does not mean they believe the defendant is innocent.

We may reject or fail to reject the alternative hypothesis, but we typically never accept the null hypothesis as true.

Example 3

In Example 1 we has three choices for the question:

How many of the world's 1 year old children today have been vaccinated against some disease?

(a) 20% (b) 50% (c) 80%

If someone has pick an answer completely at random, what would the probability of selecting the correct answer be?

$p = \frac{1}{3} = 0.333$

So, if the null hypothesis is that someone was guessing at random, the alternative would be they were not guessing.

$$H_0 : p = 0.333$$
$$H_A : p \neq 0.333$$

Definition

The value we are comparing the parameter to is called the **null value**.

Example 4

It may seem incredibly unlikely that the proportion of people who get the correct answer is *exactly* 33.3%.

If we don't believe the null hypothesis, should we simply reject it?

No. While we may not believe the null hypotheses, we need strong evidence before we reject the null hypothesis and conclude something more interesting.

Even if we don't believe the proportion is *exactly* 33.3%, that doesn't tell us anything useful about if people do better or worse than random guessing.

Note

We will be using the `rosling_responses` data set to evaluate the hypothesis test evaluating whether college-educated adults who get the question about infant vaccination correct is different from 33.3%.

This data set summarizes the answers of 50 college-educated adults. Of these 50 adults, 24% of respondents got the question correct.

Example 5

For this data set, we have $n = 50$ and $\hat{p} = 0.24$. Lets see if it's reasonable to construct a confidence interval.

$$np = 50 \cdot 0.24 = 12 \geq 10 \checkmark$$

$$n(1-p) = 50(1-0.24) = 50 \cdot 0.76 = 38 \geq 10 \checkmark$$

The conditions are met, so let's construct a 95% confidence interval.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.24 \pm 1.96\sqrt{\frac{0.24(1-0.24)}{5}} = 0.24 \pm 0.118381$$

The confidence interval is $(0.122, 0.358)$, which means we are 95% confident that the proportion of college-educated adults to correctly answer the question is between 12.2% and 35.8%.

Since the null hypothesis of $p = 0.333$ falls in this range, we cannot say the null hypotheses is implausible. We fail to reject the null hypothesis.

Just because we conclude that it's plausible that $p = 0.333$ does not mean we actually accept the null hypothesis.

Example 6

There are 2 billion children in the world today aged 0-15, how many children will there be in 2100 according to the United Nations?

(a) 4 billion (b) 3 billion (c) 2 billion

What are the null and alternative hypotheses?

$$H_0 : p = 0.333 \qquad H_A : p \neq 0.333$$

If we take a larger sample of 228 college-educated adults, 34 (14.9%) selected the correct answer: (c) 2 billion

Let's construct a 95% confidence interval.

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.149 \pm 1.96\sqrt{\frac{0.149(0.851)}{228}} = 0.149 \pm 0.04622$$

The confidence interval is $(0.103, 0.195)$, which means we are 95% confident that the proportion of college-educated adults that answered the question correctly is between 10.3% and 19.5% Since $p = 0.333$ is implausible, we reject the null hypothesis.

Note

In the last example, because the confidence interval $(0.103, 0.195)$ is below p=0.333, we can conclude that college-educated adults do worse than random guessing on this question.

Note

This shows a general trend, in that many people are more pessimistic about progress than reality suggests.

Note

It is possible to come to difficult conclusions depending on the confidence level you use.

A 99.9% confidence interval for the last example is $(0.072, 0.227)$, which is wider than the 95% confidence interval.

This means you always need to write what confidence level you used.