code cademy

Hypothesis testing: Testing a Sample Statistic

Converting P-Values

P-values are probabilities. Translating from a probability into a significant or not significant result involves setting a significance threshold between 0 and 1. P-values less than this threshold are considered significant and p-values higher than this threshold are considered not significant.

Significance Threshold

The significance threshold is used to convert a p-value into a yes/no or a true/false result. After running a hypothesis test and obtaining a p-value, we can interpret the outcome based on whether the p-value is higher or lower than the threshold. A p-value lower than the significance threshold is considered significant and would result in the rejection of the null hypothesis. A p-value higher than the significance threshold is considered not significant.

code|cademy

Hypothesis Testing Errors

When using significance thresholds with hypothesis testing, two kinds of errors may occur. A type I error, also known as a false positive, happens when we incorrectly find a significant result. A type II error, also known as a false negative, happens when we incorrectly find a non-significant result:

Null hypothesis:	is true	is false
P-value significant	Type I Error	Correct!
P-value not significant	Correct!	Type II error

Type I Error Rate

A significance threshold is used to convert a p-value into a yes/no or a true/false result. This introduces the possibility of an error: that we conclude something is true based on our test when it is actually not true. A type I error occurs when we calculate a "significant" p-value when we shouldn't have. It turns out that the significance threshold we use for a hypothesis test is equal to our probability of making a type I error.

Multiple Hypothesis Test Error Rate

When working with a single hypothesis test, the type I error rate is equal to the significance threshold and is therefore easy for a researcher to control. However, when running multiple hypothesis tests, the probability of at least one type I error increases beyond the significance threshold for each test. The probability of an error occurring when running multiple hypothesis tests is $1-(1-a)^n$, where a is the significance threshold and n is the number of tests.

codecademy

Binomial Hypothesis Tests

Binomial hypothesis tests compare the number of observed "successes" among a sample of "trials" to an expected population-level probability of success. They are used for a sample of one binary categorical variable. For example, if we want to test whether a coin is fair, we might flip it 100 times and count how many heads we get. Suppose we get 40 heads in 100 flips. Then the number of observed successes would be 40, the number of trials would be 100, and the expected population-level probability of success would be 0.5 (the probability of heads for a fair coin).
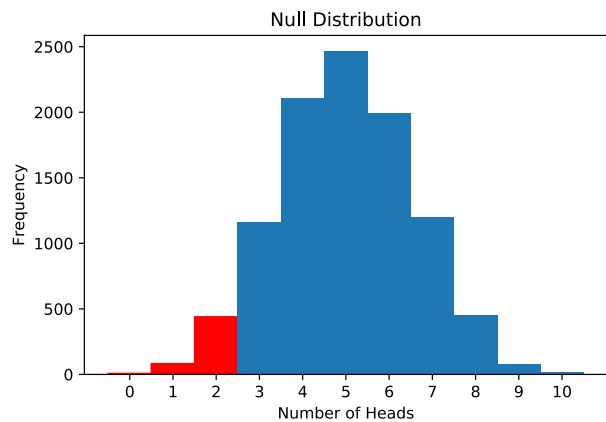
Null and Alternative Hypotheses

Hypothesis tests start with a null and alternative hypothesis; the null hypothesis describes no difference from the expected population value; the alternative describes a particular kind of difference from an expected population value (less than, greater than, or different from). For example if we wanted to perform a hypothesis test examining if there is a significant difference between the temperature on earth in 1990 as compared to the temperature in 2020, we could define the following null and alternative hypotheses:

- Null: The average temperature on earth in 1990 was the same as the average temperature in 2020
- Alternative: The average temperature on earth in 1990 was less than the average temperature in 2020

P-Values

When running a hypothesis test, it is common to report a p-value as the main outcome for the test. A p-value is the probability of observing some range of sample statistics (described by the alternative hypothesis) if the null hypothesis is true. For example, the image shown here illustrates a p-value calculation for a binomial test to determine whether a coin is fair. The p-value is equal to the proportion of the null distribution colored in red. The null and alternative hypotheses for this test are as follows:

- Null: The probability of heads is 0.5
- Alternative: The probability of heads is less than 0.5

Simulating Hypothesis Tests

The example code shown here simulates a binomial hypothesis test with the following null and alternative hypotheses:

- Null: The probability that a visitor to a website makes a purchase is 0.10
- Alternative: The probability that a visitor to a website makes a purchase is less than 0.10.

The p-value is calculated for an observed sample of 500 visitors where 41 of them made a purchase.

```python
import numpy as np
import pandas as pd

null_outcomes = []
observed_value = 41

# simulate the null distribution
for i in range(10000):
    simulated_visitors =
np.random.choice(['y', 'n'], size=500,
p=[0.1, 0.9])
    num_purchased =
np.sum(simulated_visitors == 'y')
    null_outcomes.append(num_purchased)

# calculate the p-value:
null_outcomes = np.array(null_outcomes)
p_value = np.sum(null_outcomes <=
observed_value)/len(null_outcomes)
```

Binomial Tests in Python

The `scipy.stats` library of Python has a function called `binom_test()`, which is used to perform a Binomial Test. `binom_test()` accepts four inputs, the number of observed successes, the number of total trials, an expected probability of success, and the alternative hypothesis which can be 'two-sided', 'greater', and 'less'.

```python
from scipy.stats import binom_test

pval = binom_test(observed_successes,
sample_size,
expected_probability_of_success,
alternative = 'greater')
```

One-Sample T-Tests

One-sample t-tests are used compare a sample mean to an expected population mean. They are used for a sample of one quantitative variable. For example, we could use a one-sample t-test to determine if the average amount of time customers spend browsing a shoe boutique is longer than 10 minutes.

One-Sample T-Tests In Python

A one-sample t-test can be implemented in Python using the `ttest_1samp()` function from `scipy.stats`. The function requires a sample distribution and expected population mean. As shown, the t-statistic and the p-value are returned.

```python
tstat, pval =
ttest_1samp(sample_distribution,
expected_mean)
```

code|cademy

Binomial and T-Test Assumptions

Before running a one-sample t-test, it is important
to check the following assumptions.

- The sample should be independently and
  randomly sampled from the population of
  interest
- The sample should be normally distributed
  or the sample size should be large

⤓ **Print**   ⋘ **Share** ▼