**Hypothesis Testing**

**Concepts and approach to apply:**

**1. What is alpha in hypothesis testing?**

Tags: #hypothesis testing #probability

Alpha is the level of significance. It is basically the probability of an error that we reject the null hypothesis when it is true. It is basically a small chance that our test result is by chance. 5% and 1% are standard values of alpha taken for most practical problems. A 5% value of alpha would mean that we are 95% confident that our test results are significant and not by chance. Note that the value of alpha is decided prior to the test and is not changed after that.

**2. What are the steps of conducting a hypothesis test?**

Tags: #hypothesis testing #p-value

The following are the steps of conducting a hypothesis test

1. Formulate the null and the alternative hypotheses appropriately for the problem in question.
2. Based on the nature of the data under the hypothesis, select the appropriate test.
3. Decide the significance level prior to conducting the test.
4. Collect the appropriate data for the test.
5. Calculate the p-value by conducting the test.
6. Compare the p-value with alpha and make a conclusion based on the p-value

**3. How to decide whether to use the normal distribution or the t-distribution to calculate the confidence interval?**

Tags: #distribution #confidence interval #standard deviation

The distribution that will be used for the calculation of confidence interval depends on whether the population standard deviation is known or unknown. When the population standard deviation is known, the normal distribution is used. On the other hand, when the population standard deviation is unknown, we estimate it using the sample standard deviation and calculate the confidence interval using the t-distribution.

**4. Which Python function is used to compute a confidence interval for the population mean?**

Tags: #confidence interval #population #sample

The *interval()* function is used to calculate a confidence interval. The *interval()* function calculates the endpoints within which the specified percentage of values of the distribution lie.

For example, from a population, we have drawn a random sample of size 100, whose mean is calculated to be 500 and the population standard deviation is known to be 25. Then the 95% confidence interval for the population mean can be calculated by using the *interval()* function of the [norm](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html) class of*scipy.stats* as shown below:

import numpy as np  
from scipy.stats import norm  
np.round(norm.interval(0.95, loc = 500, scale = 25/np.sqrt(100)), 2)

where *np.round()* is used to round off the output to the specified number of decimal points.

Now, for the same example as above, consider that the population standard deviation is not known to us. In that case, the 95% confidence interval for the population mean can be calculated by using the*interval()* function of the [t](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.t.html) class of *scipy.stats* as shown below:

import numpy as np  
from scipy.stats import t  
np.round(t.interval(0.95, df = 100-1, loc = 500, scale = 25/np.sqrt(100)), 2)

where the *df* argument denotes the degrees of freedom of the t-distribution. It is calculated as *(n-1)*, *n* being the sample size.