Hypothesis Testing Tool

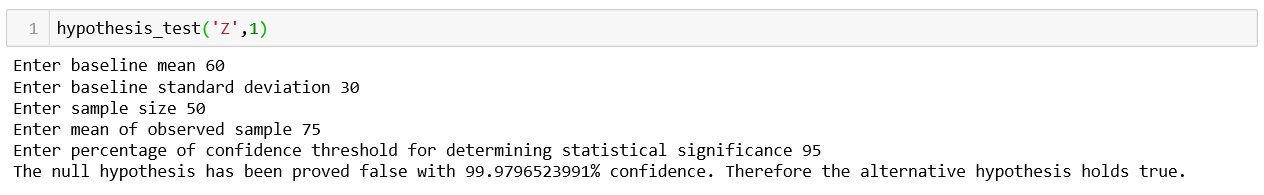
**High Level Summary:**

Hypothesis testing is a crucial activity in Data Science which helps us determine how statistically significant a hypothesis is based on known stats.

**Detailed Summary:**

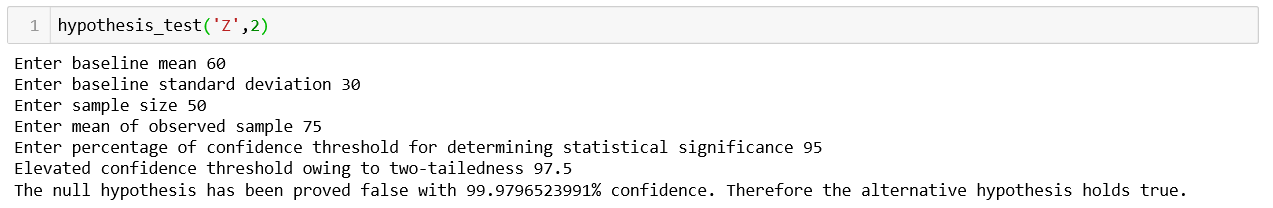
Z test, one tail:

In a mass survey conducted last year, users were asked how much time they spent on a certain website per day. The average was 1 hour and the standard deviation was 30 minutes. In a recent survey where only 50 users participated, we found the average activity time to be 1.25 hours. Can we state with 95% confidence that the activity time has increased in general among users?



Z test, two tails:

In a mass survey conducted last year, users were asked how much time they spent on a certain website per day. The average was 1 hour and the standard deviation was 30 minutes. In a recent survey where only 50 users participated, we found the average activity time to be 1.25 hours. Can we state with 95% confidence that the activity time has changed in general among users?

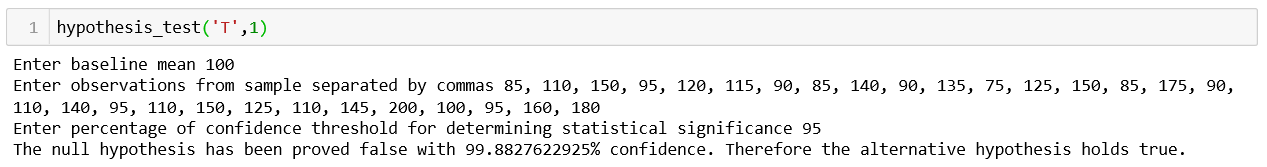


T test, one tail:

The average American spends $100 on gas for commute each month. In a survey involving 30 random users of a certain website, we got the following responses when asked how much they spent on gas to commute to work.

85, 110, 150, 95, 120, 115, 90, 85, 140, 90, 135, 75, 125, 150, 85, 175, 90, 110, 140, 95, 110, 150, 125, 110, 145, 200, 100, 95, 160, 180

It’s evident that within this sample, users of this certain website spend more than average on gas for commute each month (with a mean of around 121). But can we say with 95% confidence that these users in general spend more than average on commute each month?

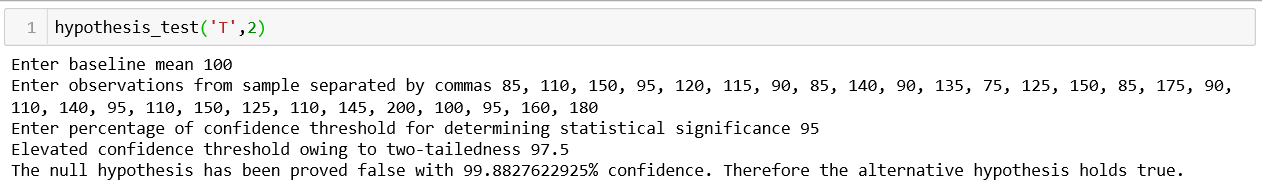


T test, two tails:

The average American spends $100 on gas for commute each month. In a survey involving 30 random users of a website, we got the following responses when asked how much they spent on gas to commute to work.

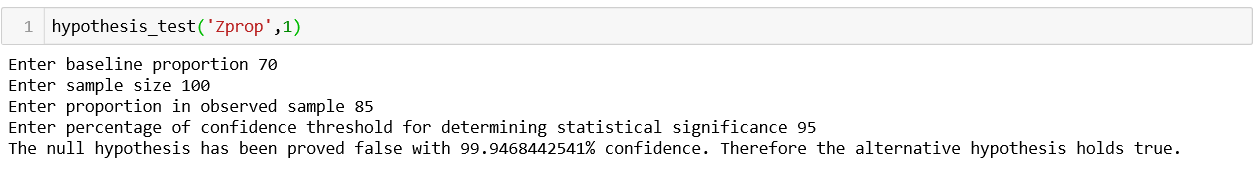
85, 110, 150, 95, 120, 115, 90, 85, 140, 90, 135, 75, 125, 150, 85, 175, 90, 110, 140, 95, 110, 150, 125, 110, 145, 200, 100, 95, 160, 180

It’s evident that within this sample, these users spend more than average on gas for commute each month (with a mean of around 121). But can we say with 95% confidence that users of this website in general spend more than average or less than average on commute each month?



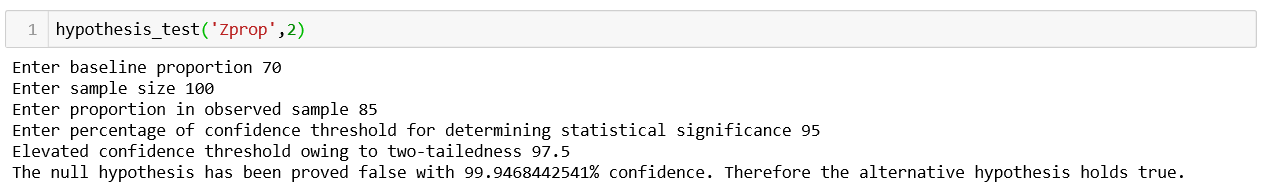
Z Proportion test, one tail:

In a mass survey conducted last year, users of a website were asked if the website helped them get ahead in their careers. 70% responded ‘Yes’ and 30% responded ‘No’. In a recent survey where only 100 users participated, 85% answered affirmatively and the rest responded negatively. Can we state with 95% confidence that the website has paved way for more users to get ahead in their careers?



Z Proportion test, two tails:

In a mass survey conducted last year, users of a website were asked if the website helped them get get ahead in their careers. 70% responded ‘Yes’ and 30% responded ‘No’. In a recent survey where only 100 users participated, 85% answered affirmatively and the rest responded negatively. Can we state with 95% confidence that the website has paved way for more users to get ahead in their careers?



**NOTE:**

1. When doing a Z test, remember the following:
   1. The members of the sample should be chosen at random.
   2. The members of the sample should be independent of each other.
   3. The population’s standard deviation should be known.
   4. The sample should contain at least 30 members.
2. Perform T test only when the baseline standard deviation is unknown AND the sample size is less than 30.
3. A one-tailed test is appropriate if you only want to determine if there is a difference between groups (for example, populations from the past and the present) in a specific direction and you are completely uninterested in the possibility that the opposite outcome could be true. When in doubt, use a two tailed test.