Estimation of sample size

One of the most crucial aspects of statistical analysis is determining the optimum sample size. If the sample size is too small, the results will be invalid and the population being investigated will not be appropriately represented. The likelihood that a statistical test will reject the null hypothesis when it is false is known as its power. That is, power is a measure of the likelihood of not making a type II error. The sample size and the effect size are the two most important criteria that influence a study’s power. The power of the significance test increases as the sample size grows. This is because a bigger sample size narrows the test statistic’s distribution. This reduces the standard error of the distribution and the acceptance region, increasing the level of power.

A study should only be carried out if there is a reasonable probability of obtaining relevant information. As a result, determining the optimal sample size for a study is a critical stage in its design.

One sample

For single mean :

This test is used to determine a specific sample size while considering preference proportions and other features of the dataset. We must first define an acceptable margin of error d before determining a suitable sample size. Remember that the wiggle room around the point estimate is the margin of error d. Sometimes half the width of the confidence interval may be used. The wider the C.I., The less reliable the sample statistic is, the less likely it is to provide an accurate approximation of the true value of the population parameter.

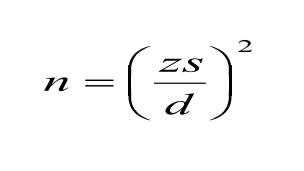
Procedure :

1. Calculate standard deviation.

2. Specify the confidence interval and normal deviation.

3. Find the error rate.

4. Compute the sample size using the formula.



where:

n = sample size

z = confidence interval in standard error units

s = standard error of the mean

d= acceptable magnitude of error

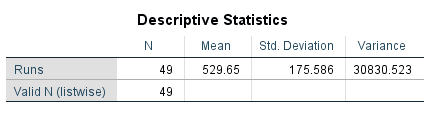
5. Conclusion.

Problem :

we want to estimate the average runs of players , and we want the error of estimation to be less than 20 runs of the true mean, with probability of 95% (e.g.,error rate of 5%).[z=1.96 for 95% confidence level ]

Solution :

Estimate the standard error of the mean s from the data



From the table we came to know standard deviation s= 175.586

And z=1.96 ; d=20 (from data)

Therefore,

n = 1.96\*1.96\*175.586\*175.586/20\*20

n =296.0955

The sample size required to estimate the mean runs with 95% confidence and error of estimation less than 20 runs is 296

For a single proportion :

This test is used to determine a specific sample size while considering preference proportions and other features of the dataset. In this scenario, we try to estimate the sample size required for a particular preference given a specific proportion in the population.

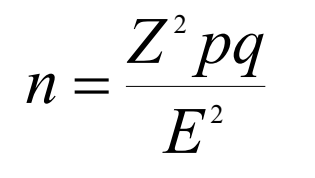
Procedure :

1. Specify the proportion.

2. Specify the confidence interval and normal deviation.

3. Find the error rate.

4. Compute the sample size using the formula:



Where:

n = number of items in samples

Z^2 = square of confidence interval in standard error units

p = estimated proportion of success

q = (1-p) or estimated the proportion of failures

E^2 = square of maximum allowance for error between true

proportion and sample proportion, or Zsp squared.

5. Conclusion.

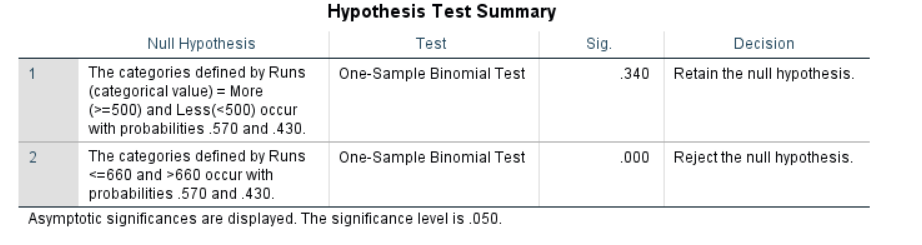
Problem :

The proportion of preference for players who scored greater than 500 runs is 57%. We want to estimate the preference p in a community within 5% with 95% confidence interval.

Solution :

H0 : The proportion of preference for players who scored greater than 500 is 57%

H1 : The proportion of preference for players who scored less than 500 is 57%



From the table,

We can accept the H0.

with 95% of significance level i.e α=0.05

p=57%=0.57 => q=0.43

E=0.05 (given)

When we are having 95%significance z=1.96

So,

N = Z^2 pq/E^2

= 1.96\*1.96\*0.57\*0.43/0.05\*0.05

= 376.6304

The sample size required to estimate the preference of proportion with 95% confidence and error of estimation less than 5.0% is 376.6