**Power Analysis**

By Guobo Chen

Power calculation and sample size determination are routines in statistical analysis. Upon the data used in analysis, the power calculation can be for continuous data, often following a normal distribution, and discrete data, often following a binomial distribution. For both types of data, one-tail and two-tail hypothesis tests can be conducted and their corresponding sample size can be derived given controlled type I () and type II error rate (). This note is a summary for some commonly used power calculation in application.

**For Normal distribution**

Given the type I error rate of and type II error rate of , the required sample size will give statistical power of given controlled type I error rate of . The corresponding quantities for and is and . Given , and ; given , , and .

This equation gives the generic form for sample size determination for a statistical test, as exampled with z score below

**Scenario I: one-sample test for the mean of a normal-distributed sample**

The observation is distributed as . The hypothesis test is to examine whether the sample’s mean and the target . Upon it is one-tail or two-tail test, the sample size can be determined as below. indicates the cumulative-distribution function (cdf) for a standard normal distribution that – follows .

Power= (Eq 1)



in which and are sampling variances of the means for the null and an alternative populations, respectively

As is assumed to be equal to , and , we have

(Eq 2)

Similarly, when it is for two-tailed test that , we only need to replace the cut-off with , and the power is:



If ,

(Eq 3)

Because and its probability is negligible under ;

if , (Eq 4)

because is negligible small.

Under both scenarios, Eq 3 and Eq 4 can be written as

And the sample size can be, to have balanced type I and type II error rates, expressed as

(Eq 5)

The result is as summarized below.

|  |  |  |
| --- | --- | --- |
| Sample | Test | Sample size |
| Single sample | One-tail test |  |
|  | Two-tail test |  |

In which the sampling variance can be directly estimated from the sample as . It is assumed that the sampling variance of the null distribution is identical to the observed sample.

**Scenario II: Two-sample test for means**

The two samples are distributed as and , respectively. The interest of the study is to compare and test their significance, and it follows . If these two samples have the equal sampling variance, the test statistic is

and , and the pooled estimate for the sampling variance is . and . The test statistic has degrees of freedom.

However, if their sampling variances are unequal (can be tested via *F* test), then the corresponding *t*-test is

The degrees of freedom of the *t* test is

The result is summarized as below

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Two samples | One-tail test |  |
|  | Two-tail test |  |
|  |  |  |

**Test for Equivalence**

The object for equivalence is to test the hypothesis below

When the null hypothesis is rejected,

As

It is approximated to

And .

**Power calculation for**

Alternatively, the test statics often is chosen to be distribution, and the power analysis can be conducted correspondingly. If , then . is the non-centrality parameter. If , in which .

The corresponding power can be evaluated by

x=seq(0,20,length=100)

alpha=qchisq(0.95,1)

curve(dchisq(x,df=1), xlim=c(0, 40), ylab="density", lwd=3)

abline(v=alpha, col="gray", lty=2)

NCP=c(1, 5, 10, 20)

for(i in 1:length(NCP)) {

curve(dchisq(x,df=1, ncp=NCP[i]), add = T, col=NCP[i])

pchisq(alpha, df=1, ncp=1, lower.tail = F)

print(pchisq(alpha, 1, ncp=NCP[i], lower.tail = F))

}



**Binomial samples**

**Scenario I: One sample test**

is from Bernoulli distribution with probability . For observed observations, is estimated as , and .

Because it is asymptotically a normal distribution, its sample size determination is similar to the “one-sample” test above for both one-tail and two-tail tests.

**Scenario II: Two-sample test**

in the first sample is distributed as and in the sample second sample is distributed as . Given the null hypothesis , we have have the mean of 0 and the sampling variance under the null is , in which can be estimated as

For one-tail test, the sample size can be determined by

and .

Similarly, for the two-tail test, the sample size can be determined by

and .

**Linear regression**

It is known that

and the sampling variance of the regression coefficient is

Its corresponding sample size for the two-tailed test is

If both and are standardized, the equation can be written as

**Genetic prediction**

For the model above, the predicted value for the testing sample is

The covariance between and the predicted is

follows with NCP , it’s corresponding statistical power is .

**Genetic prediction with overlapping samples**

The covariance between and the predicted is

And the corresponding , approaching 1 when is very large.