

BASIC RESEARCH FOR CLINICIANS

Sample Size for Experimental Studies

Padam Singh, MD, Gurgaon, India

ABSTRACT

An essential part of any medical research is to decide how many subjects need to be studied. A formal sample size calculation is a prerequisite to justify that the study with this size is capable of answering the research questions. This article highlights the statistical principles involved in sample size calculation along with formulae used in different situation illustrating with examples. The implications of deviations are also discussed.

Key Words: Sample size, experimental studies.

Introduction

The aims and objectives of a clinical research are used in generating a “research hypothesis.” The calculation of sample size requires that we quantify the research hypothesis. The theory of statistical test of significance centers around setting a “null hypothesis.” This statistical null hypothesis is often the negation of the research hypothesis. In general the null hypothesis indicates that the effect of interest is zero or that two treatments/drugs/devices are equally effective. The “null hypothesis” is then tested against the competitive “alternative hypothesis,” indicating that the effect of interest is not zero.

While drawing inference two types of errors are encountered.

Types of Error in Hypothesis Tests

Types of Error in Hypothesis Tests		
Reality		
Conclusion of significance test	Null hypothesis is true	Null hypothesis is false
Reject null hypothesis	Type I Error	Correct conclusion
Do not reject null hypothesis	Correct conclusion	Type II Error

From: Medanta–The Medicity, Gurgaon, India. (P.S.)

Corresponding Author: Padam Singh
 Medanta–The Medicity,
 Sector 38, Gurgaon, Haryana-122001, INDIA.
 Ph: +91-124-4141414 | Fax: +91-124-4834111
 Email: padamsingh@medanta.org

Type I error is committed when true H₀ is rejected that is false claim of superiority gets accepted. The implication of Type I error in the context of a clinical trial is that the new regimen, although not effective drug, is adopted as prescribed. Thus, ineffective drug is allowed to be marketed. On the other hand, Type II error is said to be committed when false H₀ is accepted that is true claim of superiority is rejected. In a trial on a new regimen, Type II error would imply that the new regimen is not approved when it is actually more effective. Thus the medical profession and the society is deprived of the benefits of this regimen.

Normal distribution

Most measurements follow a pattern of the type shown in Fig. 1. The pattern has peak at the middle and rate of decline on either side has a specific pattern “bell shaped”; this is known as Gaussian or Normal distribution.

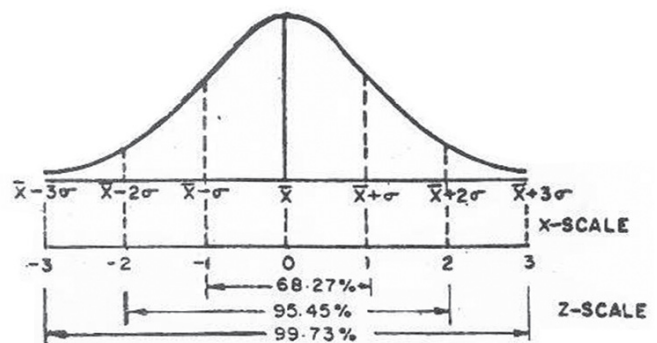


Figure 1. Normal distribution.

Generally, the values of standard normal variate are needed in computation of sample size for desired confidence level and power along with other assumptions. Table showing area under standard normal curve are available to find out the values of standard normal variate corresponding to desired values of α and β . However, some of the most commonly used value of $Z_{1-\alpha/2}$ and $Z_{1-\alpha}$ corresponding to different desired levels of significance are given as under:

Alpha(α)	Two Sided Test $Z_{1-\alpha/2}$	One Sided Test $Z_{1-\alpha}$
0.01	2.576	2.326
0.05	1.960	1.645
0.10	1.645	1.282
0.20	1.282	0.842

Similarly, values of standard normal variate ($Z_{1-\beta}$) corresponding to different values of power ($1-\beta$) are as under:

$1 - \beta$	$Z_{1-\beta}$
0.99	2.326
0.95	1.645
0.90	1.282
0.80	0.842

For other values of α and β , readers are advised to refer to following table:

Sample size for experimental studies

An experimental study is a type of evaluation that seeks to determine whether a program or intervention had the intended causal effect. Out of the various components of experimental studies, we are focusing mainly on studies/designs comparing a treatment group and a control group.

Clinicians are not at ease with formulae and computations. Therefore tables have been prepared for use by them,

Entry is area A under the standard normal curve from $-\infty$ to $z(A)$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319

1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9969	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

which could serve as ready reckoners in determining the sample size for a given situation.

In studies, the outcome variable for evaluation could be binary, such as “cure rate” or “survival rate,” efficacy, etc. Also there are studies aimed at evaluating the changes in the parameters which are of continuous nature such as BP measurements, HbA_{1C} values, etc. The computation of sample size is discussed separately for both types of outcomes.

Binary outcome

In most experimental studies, a new drug is compared for its efficacy with the standard drug as control. The sample size in this situation is determined based on the following information:

- Approximate value of efficacy/cure rate for standard treatment (e.g., 60%) = P₁
- Approximate value of efficacy/cure rate for new drug (e.g., 80%) = P₂
- Effect size (i.e., difference in efficacy of control

and experimental group, e.g., 20%) = (P₂-P₁)

- Level of significance (usually 5%)
- How high should the probability of obtaining significant result be (“power,” e.g., 90%)?

The formula used in the situation is as under:

$$n = \frac{[Z_{1-\alpha/2} \sqrt{2 P (1-P)} + Z_{1-\beta} \cdot \sqrt{\{P_1(1-P_1) + P_2 (1-P_2)\}}]^2}{(P_1-P_2)^2}$$

where P=(P₁+P₂)/2

In the above example, the efficacy/cure rate on standard treatment is 60%. The new treatment would be considered superior if it offers at least 20% more over the standard treatment. It is desired to estimate the minimum number of patients required in each group in this study. The computation of sample size is as under:

Here, P₁ = 0.6, P₂ = 0.8
 P = (0.6 + 0.8)/2=0.7
 α = 0.10, β = 0.20

Therefore;

$$n = \frac{[1.96 \sqrt{2(0.7)(1-0.7)} + 0.842 \sqrt{(0.6)(1-0.6) + (0.8)(1-0.8)}]^2}{(0.6-0.8)^2}$$

$$= 82$$

The sample size to study the above claim with 95% confidence level and 80% power will be 82.

In this very example, other thing remaining the same, if the power of the test is changed from 80% to 90%, the sample size increases from 82 to 105 as could be seen from table below:

Confidence level	Power	P_1	P_2	n
95%	80%	0.6	0.8	82
95%	90%	0.6	0.8	105

Similarly, if confidence level is increased from 95% to 99%, the sample size increases from 82 to 121.

Confidence level	Power	P_1	P_2	n
95%	80%	0.6	0.8	82
99%	80%	0.6	0.8	121

Further, if the effect size ($P_2 - P_1$) is 10% as against 20%, then also the sample size increases from 82 to 356.

Confidence level	Power	P_1	P_2	n
95%	80%	0.6	0.8	82
95%	80%	0.6	0.7	356

In view of above, the sample size calculations should be based on our best guesses of the situation and realistic assumptions because of the following facts:

- Sample size will be higher for higher power of detecting the difference as significant
- Sample size will be higher for higher confidence levels
- Sample size will be higher for detecting an effect size which is of smaller magnitude

Many a times, the investigator uses a lower sample size than desired. In that case, the implications of

the reduction in sample size could be examined statistically. Also, the estimated effect size could be different than assumed, the implication of which could also be analyzed using statistical principles.

Continuous outcome variable

In this case besides confidence level and power, the other information required is as under:

- mean of the outcome variable in group 1 = M_1
- mean of the outcome variable in group 2 = M_2
- SD of the outcome variable in group 1 = σ_1
- SD of the outcome variable in group 2 = σ_2

The formula used in calculation of sample size is given below:

$$n = \frac{(\sigma_1^2 + \sigma_2^2) [Z_{1-\alpha/2} + Z_{1-\beta}]^2}{(M_1 - M_2)^2}$$

Example:

A nutritionist wishes to study the effect of lowering sodium in diet on systolic blood pressure (SBP). During a pilot study, it was observed that the standard deviation of SBP in community with a low sodium diet was 11.3 mmHg while that with high sodium diet it was 13.0 mmHg. If $\alpha = 0.05$ and $\beta = 0.10$, how many persons from each community should be studied if one wants to detect 3.0 mmHg difference in SBP in the two communities as significant?

Solution:

$$n = \frac{(\sigma_1^2 + \sigma_2^2) [Z_{1-\alpha/2} + Z_{1-\beta}]^2}{(M_1 - M_2)^2}$$

$$= \{11.3^2 + 13^2\} (1.96 + 1.28)^2 / 3^2$$

$$= 347$$

Therefore, a minimum of 347 persons from each community would be required to be included in the study.

Equivalence of two interventions

When the standard therapy is invasive, expensive,

or toxic and the experimental therapy is conservative, one may be interested in showing that the experimental therapy is equivalent in efficacy of standard therapy, rather than necessarily superior.

In this situation, the null hypothesis may specify that the difference in success rate of the control therapy (P_c) and standard therapy (P_s) by up to d amount will be considered equivalent.

The sample size for equivalence studies is calculated using the following formula:

$$n = \frac{2P(1-P)[Z_{1-\alpha} + Z_{1-\beta}]^2}{(d)^2}$$

Here, $P_c = P_s = P$

Example:

If the underlying 5-year survival rate in two groups is 80% and the threshold for equivalence is 10%, the sample size required for establishing equivalence with 80% power and 95% confidence level is as under:

$$n = \frac{2*(0.8)(0.2)[1.645+0.84]^2}{(0.1)^2}$$

$n = 198$ subjects per group

Computation of power

Many times it is desirable to compute the power of the study for given sample size and other assumptions. The formula used in this case is given by

$$Z_{1-\beta} = \frac{\sqrt{n}(P_1 - P_2) - Z_{1-\alpha/2}\sqrt{2P(1-P)}}{\sqrt{\{P_1(1 - P_1) + P_2(1 - P_2)\}}}$$

The steps involved are as under:

- Compute $Z_{1-\beta}$ using the above formula for given n , P_1 , P_2 , and α .
- Then making use of the table of normal distribution, find the cumulative probability corresponding to the calculated value of $Z_{1-\beta}$.

Illustration 1

In the example considered earlier, if the resulting sample size is 64 as against 82 planned, the power is computed as under:

The value of $Z_{1-\beta}$ works out as 0.50.

The area of the normal curve upto 0.50 is 0.6915.

Thus the power of the study reduces to 69.15% as against 80% planned.

Illustration 2

In the example considered earlier, if the difference resulting as significant is 10% as against 20% assumed, in that situation the power is computed as under:

The value of $Z_{1-\beta}$ works out as -0.62.

Area of the normal curve upto -0.62 is 0.2716. Thus the power reduces to 27.16%.

Adjustment for Varying Sample Size

Many a times, it may not be possible or even desirable to have same sample size for each group. It may be of interest to know as to what difference would it make on the inference if the sample size varies in two groups as well as what way it could be adjusted for desired results. Table 1 gives appropriate adjustment factors for the number of sample in the experimental group according to the differing ratio of sample in the experimental versus control groups.

If n is the estimated sample size for equal-size groups and the size of the larger group is desired to be c times that of the smaller group, factor (f) is applied to the smaller group which equals $(c + 1)/(2c)$, in this case, the sample size of the smaller group is “ fn ” and that of the larger group as “ cn .”

Adjustments for loss to follow-up

It is generally the case that some of the subjects originally recruited in the study might be lost to follow-up or withdraw. Thus, the calculated sample size should be increased to allow for possible nonresponse or loss to follow-up.

Ratio of larger to smaller group (c)	Adjustment to sample size of smaller group (f)
1	2/2(1)
2	3/4
3	4/6(2/3)
4	5/8
5	6/10(3/5)
6	7/12
7	8/14(4/7)
8	9/16
9	10/18(5/9)
10	11/20

Adjustment factor for x% loss = $100/(100 - x)$

For example, if x = 20% the multiplying factor equals $100/(100 - 20) = 1.25$.

Adjustment for confounding

If there is a likelihood of a confounding variable, to adjust for the effect, the sample size is required to be increased by 10% (1). For several confounding variables that are jointly independent, as a rough guide one could add the extra sample size requirements for each variable

separately.

Adjustment for clustered designs

In cluster randomized trials, in which randomization is applied to clusters of people rather than individuals, the sample size is required to be adjusted for clustering effect. The amount by which the sample size needs to be multiplied is known as design effect (Deff), which depends on the intraclass correlation coefficient (ICC).

Design effect (Deff) = $1 + (n' - 1) \times ICC$

where ICC = intraclass correlation coefficient

n' = average cluster size

For more details about sample size calculation for cluster randomized trials, readers are advised to refer to Donner and Klar (2).

References

1. Smith PG, Day NE. The design of case-controlled studies: the influence of confounding and interaction effects. *Int J Epidemiol.* 1984; 13:356–65.
2. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research.* Arnold: London; 2000.

Further Reading

1. Kirkwood BR, Sterne JAC. *Medical Statistics*, 2nd edition. Oxford: Blackwell Science Ltd; 2003.
2. Pandey RM. *Sample Size Calculation.* Lecture Notes, 2010.