

ESTTA Tracking number: **ESTTA298123**

Filing date: **07/29/2009**

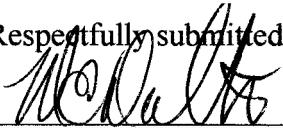
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE  
BEFORE THE TRADEMARK TRIAL AND APPEAL BOARD

Proceeding	91173105
Party	Defendant Michael Dalton
Correspondence Address	Michael Dalton P.O. Box 18137 670 Northland Blvd. Cincinnati, OH 45218-0137 UNITED STATES DALTONME@hotmail.com
Submission	Defendant's Notice of Reliance
Filer's Name	Michael Dalton
Filer's e-mail	DALTONME@hotmail.com
Signature	/ Michael Dalton /
Date	07/29/2009
Attachments	RelianceWikiSampleSizeJJ.pdf ( 3 pages )(42369 bytes ) RelianceWikiSampleSizeJJ.a.pdf ( 5 pages )(2102329 bytes )

TAB JJ



Respectfully submitted,



---

Michael Dalton, Applicant pro se  
PO Box 18137  
670 Northland Blvd.  
Cincinnati, Ohio 45218-0137  
(513)557-2901  
[DALTONME@hotmail.com](mailto:DALTONME@hotmail.com)

# Sample size

From Wikipedia, the free encyclopedia

The **sample size** of a statistical sample is the number of observations that constitute it. It is typically denoted  $n$ , a positive integer (natural number).

Typically, all else being equal, a larger sample size leads to increased precision in estimates of various properties of the population. This can be seen in such statistical rules as the law of large numbers and the central limit theorem. Repeated measurements and replication of independent samples are often required in measurement and experiments to reach a desired precision.

A typical example would be when a statistician wishes to estimate the arithmetic mean of a continuous random variable (for example, the height of a person). Assuming that they have a random sample with independent observations, then if the variability of the population (as measured by the standard deviation  $\sigma$ ) is known, then the standard error of the sample mean is given by the formula:

$$\sigma/\sqrt{n}.$$

It is easy to show that as  $n$  becomes very large, this variability becomes small. This yields to more sensitive hypothesis tests with greater statistical power and smaller confidence intervals.

## Contents

- 1 Further examples
  - 1.1 Central limit theorem
  - 1.2 Estimating proportions
  - 1.3 Extension to other cases
- 2 Required sample sizes for hypothesis tests
- 3 Stratified sample size
- 4 See also
- 5 References
- 6 External links



## Further examples

### Central limit theorem

The central limit theorem is a significant result which depends on sample size. It states that as the size of a sample of independent observations approaches infinity, provided data come from a distribution with finite variance, that the sampling distribution of the sample mean approaches a normal distribution.

### Estimating proportions

A typical statistical aim is to demonstrate with 95% certainty that the true value of a parameter is within a distance  $B$  of the estimate:  $B$  is an error range that decreases with increasing sample size ( $n$ ). The value of  $B$  generated is referred to as the 95% confidence interval.

For example, a simple situation is estimating a proportion in a population. To do so, a statistician will estimate the bounds of a 95% confidence interval for an unknown proportion.

The rule of thumb for (a maximum or 'conservative')  $B$  for a proportion derives from the fact the estimator of a proportion,  $\hat{p} = X/n$ , (where  $X$  is the number of 'positive' observations) has a (scaled) binomial distribution and is also a form of sample mean (from a Bernoulli distribution  $[0,1]$  which has a maximum variance of 0.25 for parameter  $p = 0.5$ ). So, the sample mean  $X/n$  has maximum variance  $0.25/n$ . For sufficiently large  $n$  (usually this means that we need to have observed at least 10 positive and 10 negative responses), this distribution will be closely approximated by a normal distribution with the same mean and variance.

Using this approximation, it can be shown that ~95% of this distribution's probability lies within 2 standard deviations of the mean. Because of this, an interval of the form

$$(\hat{p} - 2\sqrt{0.25/n}, \hat{p} + 2\sqrt{0.25/n}) = (\hat{p} - B, \hat{p} + B)$$

will form a 95% confidence interval for the true proportion.

If we require the sampling error  $\varepsilon$  to be no larger than some bound  $B$ , we can solve the equation

$$\varepsilon \approx B = 2\sqrt{0.25/n} = 1/\sqrt{n}$$

to give us

$$1/\varepsilon^2 \approx 1/B^2 = n.$$

So,  $n = 100 \Leftrightarrow B = 10\%$ ,  $n = 400 \Leftrightarrow B = 5\%$ ,  $n = 1000 \Leftrightarrow B = \sim 3\%$ , and  $n = 10000 \Leftrightarrow B = 1\%$ . One sees these numbers quoted often in news reports of opinion polls and other sample surveys.

### Extension to other cases

In general, if a population mean is estimated using the sample mean from  $n$  observations from a distribution with variance  $\sigma^2$ , then if  $n$  is large enough (typically  $>30$ ) the central limit theorem can be applied to obtain an approximate 95% confidence interval of the form

$$(\bar{x} - B, \bar{x} + B), B = 2\sigma/\sqrt{n}$$

If the sampling error  $\varepsilon$  is required to be no larger than bound  $B$ , as above, then

$$4\sigma^2/\varepsilon^2 \approx 4\sigma^2/B^2 = n$$

Note, if the mean is to be estimated using  $P$  parameters that must first be estimated themselves from the same sample, then to preserve sufficient "degrees of freedom," the sample size should be at least  $n + P$ .

### Required sample sizes for hypothesis tests

A common problem facing statisticians is calculating the sample size required to yield a certain power for a test, given a predetermined Type I error rate  $\alpha$ . A typical example for this is as follows:

Let  $X_i, i = 1, 2, \dots, n$  be independent observations taken from a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . Let us consider two hypotheses, a null hypothesis:

$$H_0: \mu = 0$$

and an alternative hypothesis:

$$H_a: \mu = \mu^*$$

for some 'smallest significant difference'  $\mu^* > 0$ . This is the smallest value for which we care about observing a difference. Now, if we wish to (1) reject  $H_0$  with a probability of at least  $1 - \beta$  when  $H_a$  is true (i.e. a power of  $1 - \beta$ ), and (2) reject  $H_0$  with probability  $\alpha$  when  $H_0$  is true, then we need the following:

If  $z_\alpha$  is the upper  $\alpha$  percentage point of the standard normal distribution, then

$$\Pr(\bar{x} > z_\alpha \sigma / \sqrt{n} | H_0 \text{ true}) = \alpha$$

and so

$$\text{'Reject } H_0 \text{ if our sample average } (\bar{x}) \text{ is more than } z_\alpha \sigma / \sqrt{n}$$

is a decision rule which satisfies (2). (Note, this is a 1-tailed test)

Now we wish for this to happen with a probability at least  $1 - \beta$  when  $H_a$  is true. In this case, our sample average will come from a Normal distribution with mean  $\mu^*$ . Therefore we require

$$\Pr(\bar{x} > z_\alpha \sigma / \sqrt{n} | H_a \text{ true}) \geq 1 - \beta$$

Through careful manipulation, this can be shown to happen when

$$n \geq \left( \frac{\Phi^{-1}(1 - \beta) + z_\alpha}{\mu / \sigma} \right)^2$$

where  $\Phi$  is the normal cumulative distribution function.

## Stratified sample size

With more complicated sampling techniques, such as Stratified sampling, the sample can often be split up into sub-samples. Typically, if there are  $k$  such sub-samples (from  $k$  different strata) then each of

them will have a sample size  $n_i$ ,  $i = 1, 2, \dots, k$ . These  $n_i$  must conform to the rule that  $n_1 + n_2 + \dots + n_k = n$  (i.e. that the total sample size is given by the sum of the sub-sample sizes). Selecting these  $n_i$  optimally can be done in various ways, using (for example) Neyman's optimal allocation.

According to Leslie Kish,<sup>[1]</sup> there are many reasons to do this; that is to take sub-samples from distinct sub-populations or "strata" of the original population: to decrease variances of sample estimates, to use partly non-random methods, or to study strata individually. A useful, partly non-random method would be to sample individuals where easily accessible, but, where not, sample clusters to save travel costs.

In general, for  $H$  strata, a weighted sample mean is

$$\bar{x}_w = \sum_{h=1}^H W_h \bar{x}_h,$$

with

$$\text{Var}(\bar{x}_w) = \sum_{h=1}^H W_h^2 \text{Var}(\bar{x}_h).$$

The weights,  $W(h)$ , frequently, but not always, represent the proportions of the population elements in the strata, and  $W(h) = N(h)/N$ . For a fixed sample size, that is  $n = \text{Sum}\{n(h)\}$ ,

$$\text{Var}(\bar{x}_w) = \sum_{h=1}^H W_h^2 \text{Var}(h) \left( \frac{1}{n_h} - \frac{1}{N_h} \right),$$

which can be made a minimum if the sampling rate within each stratum is made proportional to the standard deviation within each stratum:  $n_h / N_h = kS_h$ .

An "optimum allocation" is reached when the sampling rates within the strata are made directly proportional to the standard deviations within the strata and inversely proportional to the square roots of the costs per element within the strata:

$$\frac{n(h)}{N(h)} = \frac{KS(h)}{\sqrt{C(h)}},$$

or, more generally, when

$$n(h) = \frac{K'W(h)S(h)}{\sqrt{C(h)}}.$$

## See also

- Design of experiments

- Replication (statistics)
- Sampling (statistics)
- Statistical power
- Stratified Sampling
- Engineering response surface example under Stepwise regression.

## References

1. ^ Kish, L. (1965), Survey Sampling, New York: Wiley.

## External links

- NIST: Selecting Sample Sizes
- Raven Analytics: Sample Size Calculations
- ASTM E122-07: Standard Practice for Calculating Sample Size to Estimate, With Specified Precision, the Average for a Characteristic of a Lot or Process

Retrieved from "[http://en.wikipedia.org/wiki/Sample\\_size](http://en.wikipedia.org/wiki/Sample_size)"

Categories: Sampling (statistics)

Hidden categories: [Statistics articles linked to the portal](#) | [Statistics articles with navigational template](#)

---

- This page was last modified on 20 June 2009 at 18:03.
- Text is available under the Creative Commons Attribution/Share-Alike License; additional terms may apply. See Terms of Use for details.  
Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.