Sample size

Student Learning Centre
Semester 2
2006

Lincoln University
Te Whare Wānaka o Aoraki

SAMPLE SIZE

It is usual, when using statistics, to plan data size at the same time as planning data collection. This enables you to have high level of confidence as well as a small margin of error.

Recall that the margin of error in a confidence interval (for a mean) is $Z \times \frac{s}{\sqrt{n}}$

If the degree of accuracy of our estimate for μ (the population mean) has to be less than some given accuracy (*e*), then we obtain the inequality :

$$Z \times \frac{s}{\sqrt{n}} < e$$

This inequality should then be solved for *n* to find the minimum sample size which will yield the required degree of accuracy.

ie $\frac{Z^2 \times s^2}{n} < e^2$ *and by rearranging the formula,*

$$\Rightarrow n \quad > \frac{Z^2 \times s^2}{e^2}$$

The investigator, as well as deciding what sampling difference to accept, must decide the level of confidence.

Hence the formula is $$n > \frac{Z^2 \times s^2}{e^2} = \left(\frac{Z \times s}{e}\right)^2$$ **(either version).**

Here, *n* is sample size, *e* is the sampling error and s^2 is the variance.

To Determine the Sample Size for Calculating the Mean

Example *What sample size should be taken from a population of toothpaste tubes, with weight that has a standard deviation of 3g, to estimate the mean weight to within 0.4g, with 98% confidence?*

$\sigma = 3$ $e = 0.4,$ *98%confide nce* $\Rightarrow Z = 2.326$

$$n > \frac{z^2 \sigma_x^2}{e^e} = \frac{2.326^2 \text{ x } 3^2}{0.4^2} (or = \left(\frac{2.326 \times 3}{0.4}\right)^2)$$

$= 304.3$

ie minimum sample size is 305.

To Determine the Sample Size for Proportion

By adjusting for the variance, sample size for a proportion can be found from:

$$n > \frac{z^2 \times p(1-p)}{e^2}$$

Example *A market research company wishes to estimate the percentage of people in a certain age bracket who read a current affairs magazine. The degree of accuracy required for the sample is 94%. What size sample should be taken to estimate the percentage to within 4%*

$94\% \Rightarrow z = 1.881,\ e = 0.04$

.

Since p and 1-p are unknown, the best assumption to make is that they are both equal to 0.5. This gives the widest confidence interval:

$$n > \frac{z^2 p(1-p)}{e^2} = \frac{1.881^2 \text{ x } 0.5 \text{ x } 0.5}{0.04^2} = 552.8$$

ie minimum sample size would be 553

Estimation and Sample Size Determination for Finite Populations

So far we have considered an undefined population. If however, we know the population size and *n* is at least 5% of the population size N, or if sampling is without replacement, we can reduce this error by using a finite population correction factor: (F.P.C.) once the initial sample size has been calculated.

$$n = \frac{n_0 . N}{n_0 + (N-1)}$$ where n_0 represents the initial sample size

Hence, determining sample size becomes:

1. Calculate an initial sample size:

$$n_0 = \frac{z^2 . \sigma^2}{e^2}$$ (for mean) or $$n_0 = \frac{z^2 . p(1-p)}{e^2}$$ (for proportion)

2. Substitute this value in the formula $n = \frac{n_0 . N}{n_0 + (N-1)}$

Example *A simple random sample (without replacement) is to be taken from 630 students to calculate the mean age of the cars that students drive. The standard deviation of the age of students' cars is known to be about 7 years. What sample size would you need to be within 1.5 years of the true mean value, with 95% confidence?*

Solution: $$n_0 = \frac{z^2 . \sigma^2}{e^2} \Rightarrow n_0 = \frac{1.96^2 \times 7^2}{1.5^2} = 84$$

Since 84 > 5% of 630, we can apply the F.P.C. Therefore, $n = \frac{84 \times 630}{(84 + (630-1))} = 75$

Example *LUSA is interested in the number of students who receive some sort of financial aid. They randomly sample 200 students and find that 118 of them are receiving financial aid. If LUSA wants to calculate a 99% confidence interval estimate for the proportion of students receiving aid to within ± 3%, how many of the 4000 Lincoln students need to be sampled?*

Sample proportion: $p = \frac{118}{200} = 0.59 \Rightarrow (1-p) = 0.41$

Hence, $n_0 = \frac{z^2 . p(1-p)}{d^2} = n_0 = \frac{2.57^2 \times 0.59 \times 0.41)}{0.03^2} = 1776$.

and, $n = \frac{1776 \times 4000}{1776 + 3999} = 1230$.

Notice that when the FPC is applied, the sample size is reduced. This has obvious implications for cost, time, and resources, so should be applied if at all possible.

Note also

- If the variance or standard deviation are not given, use $s^2 = \frac{(max - min)^2}{6^2}$ in the formula for sample size involving a mean.
- If sample proportion not given, use, $p = 0.5$ as a conservative estimate in formula for sample size involving proportion.

Practice

1. A pilot survey was conducted to find the opinions of New Zealanders on the number of members of parliament they wanted for New Zealand. The survey questioned 32 people.

Option	Frequency
60 Members of Parliament	20
90 Members of Parliament	6
120 Members of Parliament	3
150 Members of Parliament	2
No response	1

According to the above table, how many people would need to be surveyed if the proportion of people supporting 60 members of parliament was to be estimated to within ±8 percentage points with 95% confidence?

2. The director of a hospital wants to estimate the mean number of people who are admitted to the emergency room during a 24-hour period. From a random sample of 64 different 24-hour periods, the mean number of admissions was estimated at 19.8 with a variance of 25.

a) If this director wishes to estimate the mean number of admissions to within $\pm$ 1 admission with 99% reliability, what sample size is required?

b) If this director decided that it is too expensive to sample this number of admissions what two changes could be made to reduce the required sample size?

The Lincoln University Students Association (LUSA) is interested in the number of students who receive some sort of financial aid. They randomly sampled 200 students and found that 118 of them are receiving financial aid.

c) If LUSA wants to calculate a 99% confidence interval estimate for the proportion of students receiving aid to within $\pm$ 3%, how many of the 4000 Lincoln students need to be sampled?

3. a) For a statistical test, appropriate sample size, *n*, can be determined using following formula,

$$n = \frac{Z^2\sigma^2}{e^2}$$

where e = acceptable sampling error,
Z = value of standard normal variate for desired level of confidence, and
σ^2 = estimate of population variance.

Describe in what situations we need to have larger samples.

b) A telecommunication company is planning to increase the toll call rates and wants to get an estimate of the mean monthly toll bill of residential connections. It is assumed that toll bills are normally distributed with an estimate mean of $55.00, and minimum and maximum monthly bills are $3.00 and $166.57, respectively.

i. If the telecommunication company used a sample of 60 households, what percentage sampling error with respect to the population mean do they have to deal with at 99% confidence level?

ii. The Marketing manger of the telecommunication company conducted a survey to find out the proportion of customers who are satisfied with their service. How large a random sample should be taken to estimate the true proportion of satisfied customers within $\pm$ 0.1 with 95% confidence?