

9.5: Sample Size Considerations

Learning Objectives

- To learn how to apply formulas for estimating the size samples that will be needed in order to construct a confidence interval for the difference in two population means or proportions that meets given criteria.

As was pointed out at the beginning of Section 7.4, sampling is typically done with definite objectives in mind. For example, a physician might wish to estimate the difference in the average amount of sleep gotten by patients suffering a certain condition with the average amount of sleep got by healthy adults, at 90% confidence and to within half an hour. Since sampling costs time, effort, and money, it would be useful to be able to estimate the smallest size samples that are likely to meet these criteria.

Estimating $\mu_1 - \mu_2$ with Independent Samples

Assuming that large samples will be required, the confidence interval formula for estimating the difference $\mu_1 - \mu_2$ between two population means using independent samples is $(\bar{x}_1 - \bar{x}_2) \pm E$, where

$$E = z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

To say that we wish to estimate the mean to within a certain number of units means that we want the margin of error E to be no larger than that number. The number $z_{\alpha/2}$ is determined by the desired level of confidence.

The numbers s_1 and s_2 are estimates of the standard deviations σ_1 and σ_2 of the two populations. In analogy with what we did in Section 7.4 we will assume that we either know or can reasonably approximate σ_1 and σ_2 .

We cannot solve for both n_1 and n_2 , so we have to make an assumption about their relative sizes. We will specify that they be equal. With these assumptions we obtain the minimum sample sizes needed by solving the equation displayed just above for $n_1 = n_2$.

Minimum Equal Sample Sizes for Estimating the Difference in the Means of Two Populations Using Independent Samples

The estimated minimum equal sample sizes $n_1 = n_2$ needed to estimate the difference $\mu_1 - \mu_2$ in two population means to within E units at $100(1 - \alpha)\%$ confidence is

$$n_1 = n_2 = \frac{(z_{\alpha/2})^2 (\sigma_1^2 + \sigma_2^2)}{E^2} \text{ rounded up}$$

In all the examples and exercises the population standard deviations σ_1 and σ_2 will be given.

✓ Example 9.5.1

A law firm wishes to estimate the difference in the mean delivery time of documents sent between two of its offices by two different courier companies, to within half an hour and with 99.5% confidence. From their records it will randomly sample the same number n of documents as delivered by each courier company. Determine how large n must be if the estimated standard deviations of the delivery times are 0.75 hour for one company and 1.15 hours for the other.

Solution

Confidence level 99.5% means that $\alpha = 1 - 0.995 = 0.005$ so $\alpha/2 = 0.0025$. From the last line of Figure 7.1.6 we obtain $z_{0.0025} = 2.807$.

To say that the estimate is to be “to within half an hour” means that $E = 0.5$. Thus

$$n = \frac{(z_{\alpha/2})^2 (\sigma_1^2 + \sigma_2^2)}{E^2} = \frac{(2.807)^2 (0.75^2 + 1.15^2)}{0.5^2} = 59.40953746$$

which we round up to 60, since it is impossible to take a fractional observation. The law firm must sample 60 document deliveries by each company.

Estimating $\mu_1 - \mu_2$ with Paired Samples

As we mentioned at the end of Section 9.3, if the sample is large (meaning that $n \geq 30$) then in the formula for the confidence interval we may replace $t_{\alpha/2}$ by $z_{\alpha/2}$, so that the confidence interval formula becomes $\bar{d} \pm E$ for

$$E = z_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

The number s_d is an estimate of the standard deviations σ_d of the population of differences. We must assume that we either know or can reasonably approximate σ_d . Thus, assuming that large samples will be required to meet the criteria given, we can solve the displayed equation for n to obtain an estimate of the number of pairs needed in the sample.

Minimum Sample Size for Estimating the Difference in the Means of Two Populations Using Paired Difference Samples

The estimated minimum number of pairs n needed to estimate the difference $\mu_d = \mu_1 - \mu_2$ in two population means to within E units at $100(1 - \alpha)\%$ confidence using paired difference samples is

$$n = \frac{(z_{\alpha/2})^2 \sigma_d^2}{E^2} \text{ rounded up}$$

In all the examples and exercises the population standard deviation of the differences σ_d will be given.

Example 9.5.2

A automotive tire manufacturer wishes to compare the mean lifetime of two tread designs under actual driving conditions. They will mount one of each type of tire on n vehicles (both on the front or both on the back) and measure the difference in remaining tread after 20,000 miles of driving. If the standard deviation of the differences is assumed to be 0.025 inch, find the minimum samples size needed to estimate the difference in mean depth (at 20,000 miles use) to within 0.01 inch at 99.9% confidence.

Solution

Confidence level 99.9% means that $\alpha = 1 - 0.999 = 0.001$ so $\alpha/2 = 0.0005$. From the last line of Figure 7.1.6 we obtain $z_{0.0005} = 3.291$.

To say that the estimate is to be “to within 0.01 inch” means that $E = 0.01$. Thus

$$n = \frac{(z_{\alpha/2})^2 \sigma_d^2}{E^2} = \frac{(3.291)^2 (0.025)^2}{(0.01)^2} = 67.69175625$$

which we round up to 68. The manufacturer must test 68 pairs of tires.

Estimating $p_1 - p_2$

The confidence interval formula for estimating the difference $p_1 - p_2$ between two population proportions is $\hat{p}_1 - \hat{p}_2 \pm E$, where

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

To say that we wish to estimate the mean to within a certain number of units means that we want the margin of error E to be no larger than that number. The number $z_{\alpha/2}$ is determined by the desired level of confidence.

We cannot solve for both n_1 and n_2 , so we have to make an assumption about their relative sizes. We will specify that they be equal. With these assumptions we obtain the minimum sample sizes needed by solving the displayed equation for $n_1 = n_2$.

Minimum Equal Sample Sizes for Estimating the Difference in Two Population Proportions

The estimated minimum equal sample sizes $n_1 = n_2$ needed to estimate the difference $p_1 - p_2$ in two population proportions to within E percentage points at $100(1 - \alpha)\%$ confidence is

$$n_1 = n_2 = \frac{(z_{\alpha/2})^2 (\hat{p}_1(1 - \hat{p}_1) + \hat{p}_2(1 - \hat{p}_2))}{E^2} \text{ rounded up}$$

Here we face the same dilemma that we encountered in the case of a single population proportion: the formula for estimating how large a sample to take contains the numbers \hat{p}_1 and \hat{p}_2 , which we know only after we have taken the sample. There are two ways out of this dilemma. Typically the researcher will have some idea as to the values of the population proportions p_1 and p_2 , hence of what the sample proportions \hat{p}_1 and \hat{p}_2 are likely to be. If so, those estimates can be used in the formula.

The second approach to resolving the dilemma is simply to replace each of \hat{p}_1 and \hat{p}_2 in the formula by 0.5. As in the one-population case, this is the most conservative estimate, since it gives the largest possible estimate of n . If we have an estimate of only one of p_1 and p_2 we can use that estimate for it, and use the conservative estimate 0.5 for the other.

✓ Example 9.5.3

Find the minimum equal sample sizes necessary to construct a 98% confidence interval for the difference $p_1 - p_2$ with a margin of error $E = 0.05$,

1. assuming that no prior knowledge about p_1 or p_2 is available; and
2. assuming that prior studies suggest that $p_1 \approx 0.2$ and $p_2 \approx 0.3$.

Solution

Confidence level 98% means that $\alpha = 1 - 0.98 = 0.02$ so $\alpha/2 = 0.01$. From the last line of Figure 7.1.6 we obtain $z_{0.01} = 2.326$.

1. Since there is no prior knowledge of p_1 or p_2 we make the most conservative estimate that $\hat{p}_1 = 0.5$ and $\hat{p}_2 = 0.5$. Then

$$\begin{aligned} n_1 = n_2 &= \frac{(z_{\alpha/2})^2 (\hat{p}_1(1 - \hat{p}_1) + \hat{p}_2(1 - \hat{p}_2))}{E^2} \\ &= \frac{(2.326)^2 ((0.5)(0.5) + (0.5)(0.5))}{0.05^2} \\ &= 1082.0552 \end{aligned}$$

which we round up to 1,083. We must take a sample of size 1,083 from each population.

2. Since $p_1 \approx 0.2$ we estimate \hat{p}_1 by 0.2, and since $p_2 \approx 0.3$ we estimate \hat{p}_2 by 0.3. Thus we obtain

$$\begin{aligned} n_1 = n_2 &= \frac{(z_{\alpha/2})^2 (\hat{p}_1(1 - \hat{p}_1) + \hat{p}_2(1 - \hat{p}_2))}{E^2} \\ &= \frac{(2.326)^2 ((0.2)(0.8) + (0.3)(0.7))}{0.05^2} \\ &= 800.720848 \end{aligned}$$

which we round up to 801. We must take a sample of size 801 from each population.

Key Takeaway

- If the population standard deviations σ_1 and σ_2 are known or can be estimated, then the minimum equal sizes of independent samples needed to obtain a confidence interval for the difference $\mu_1 - \mu_2$ in two population means with a given maximum error of the estimate E and a given level of confidence can be estimated.
- If the standard deviation σ_d of the population of differences in pairs drawn from two populations is known or can be estimated, then the minimum number of sample pairs needed under paired difference sampling to obtain a confidence interval for the difference $\mu_d = \mu_1 - \mu_2$ in two population means with a given maximum error of the estimate E and a given level of confidence can be estimated.

- The minimum equal sample sizes needed to obtain a confidence interval for the difference in two population proportions with a given maximum error of the estimate and a given level of confidence can always be estimated. If there is prior knowledge of the population proportions p_1 and p_2 then the estimate can be sharpened.

This page titled [9.5: Sample Size Considerations](#) is shared under a [CC BY-NC-SA 3.0](#) license and was authored, remixed, and/or curated by [Anonymous](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.