Sample Size Determination

The Margin of Error

The width of the interval is determine by the *margin of error*.

The **margin of error**, E, in a $(1-\alpha) \times 100\%$ confidence interval in which σ is known is given by

$$E = z^* \frac{\sigma}{\sqrt{n}}$$

where n is the sample size.

Note: We require that the population from which the sample was drawn be normally distributed or the sample size n be greater than or equal to 30.

The formula for obtaining the margin of error depends on three quantities:

1. Level of confidence CL, also $1-\alpha$.

2. Standard deviation of the population, σ.

3. Sample size, n.

Sample Size Determination

Suppose that we want to know the number of cars that we should sample to estimate the mean speed of all cars traveling outside the subdivision within 2 miles per hour with 95% confidence. If we solve the margin of error formula for n, we obtain a formula for determining sample size:

Determining the Sample Size n

The sample size required to estimate the population mean, μ, with a level of confidence $(1-\alpha)\times 100\%$ with a specified margin of error, E, is given by

$$n = \left(\frac{z^*\sigma}{E}\right)^2$$

where n is rounded up to the nearest whole number.

Example 1. A two-lane highway with a posted speed limit of 45 miles per hour is located just outside a small 40-home subdivision. The residents of the neighborhood are concerned that the speed of cars on the highway is excessive and want an estimate of the population mean speed of the cars on the highway. The population standard deviation is 8 mph. How large a sample is required to estimate the mean speed within 2 miles per hour with 90% confidence?

Example 2. A poll conducted on November 16-19, 2007, by American Research Group, Inc., asked American adults how much they planned to spend on gifts for Christmas. How many subjects are needed to estimate the amount of planned Christmas spending within \$5 with 95% confidence? Initial survey results indicate that $\sigma = \$82.25$.

Example 3. A used-car dealer wishes to estimate the mean number of miles on 4-year old Hummer H2s.

(a) How many vehicles should be in a sample to estimate the mean number of miles within 2,000 miles with 90% confidence, assuming that $\sigma = 19,000$ miles.

(b) How many vehicles should be in a sample to estimate the mean number of miles within 1,000 miles with 90% confidence, assuming that $\sigma = 19,000$ miles.

Example 4. A tennis enthusiast wants to estimate the mean length of men's singles matches held during the Wimbledon tennis tournament.

(a) How many matches should be in a sample to estimate the mean length within 10 minutes with 98% confidence, assuming that $\sigma = 45$ minutes.

(b) How many matches should be in a sample to estimate the mean length within 5 minutes with 98% confidence, assuming that $\sigma = 45$ minutes.

(c) What effect does halving the required accuracy have on the sample size? Why is this the expected result?

Determine the Sample Size Necessary for Estimating a Population Proportion within a Specified Margin of Error

We previously discussed determining the sample size, n, required to estimate the population mean within a certain margin of error for a specified level of confidence. We can follow the same approach to determine sample size when estimating a population proportion.

We noted the margin of error given by $E = z^* \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. We solve for n and obtain $n = \hat{p}(1-\hat{p})\left(\frac{z^*}{E}\right)^2$.

The problem is that this formula depends on $\hat{p}$, and $\hat{p} = \frac{x}{n}$ depends on the sample size n, which is what we are trying to determine in the first place. So, how do we resolve this issue? We have 2 possibilities:

1. We could determine a preliminary value for $\hat{p}$ based on a pilot study or an earlier study, or
2. we could let $\hat{p} = 0.50$. When $\hat{p} = 0.5$, the maximum value of $\hat{p}(1-\hat{p}) = 0.25$ is obtained. Using the maximum value gives the largest possible value of n for a given level of confidence and a given margin of error.

Sample Size Needed for Estimating the Population Proportion p

The sample size required to obtain a $(1-\alpha)\cdot 100\%$ confidence interval for p with a margin of error E is given by

$$n = \hat{p}(1-\hat{p})\left(\frac{z^*}{E}\right)^2 \quad (1)$$

(rounded up to the next integer), where $\hat{p}$ is a prior estimate of p. If a prior estimate of p is unavailable, the sample size required is

$$n = 0.25\left(\frac{z^*}{E}\right)^2 \quad (2)$$

rounded up to the next integer.

Example 5. With gas prices on the rise, an economist wants to know if the proportion of the U.S. population who commutes to work via car-pooling is on the rise. What size sample should be obtained if the economist wants to estimate within 2 percentage points of the true proportion with 90% confidence if

(a) the economist uses the 2006 estimate of 10.7% obtained from the American Community Survey?

(b) the economist does not use any prior estimate?

We can see the effect of not having a prior estimate of p. In this case, the required sample size more than doubled.

Example 6. A researcher wishes to estimate the proportion of households that have broadband Internet access. What size sample should be obtained if she wishes the estimate to be within 0.03 with 99% confidence if

1. she uses a 2009 estimate of 0.635 obtained from the National Telecommunications and Information Administration?

2. she does not use any prior estimates?

Example 7. A researcher for the U.S. Department of the Treasury wishes to estimate the percentage of Americans who support abolishing the penny. What size sample should be obtained if he wishes the estimate to be within 2 percentage points with 98% confidence if

1. he uses a 2006 estimate of 15% obtained from a Coinstar National Currency Poll?

2. he does not use any prior estimates?

