Margin of Error

The margin of error is *the* most widely misunderstood and misleading concept in statistics. The basic idea of it is very simple. Most of the time when we're doing statistics, we're doing statistics based on a sample - that is, the entire population we're interested in is difficult to study; so what we try to do is pick a *representative subset* called a sample. If the subset is *truly* representative, then the statistics you generate using information gathered from the sample will be *the same* as information gathered from the population as a whole.

However, research is never simple. We *never* have perfectly representative samples; in fact, it's *impossible* to select a perfectly representative sample. So we do our best to pick good samples, and we use probability theory to work out a predication of how confident we can be that the statistics from our sample are representative of the entire population. ***That's basically what the margin of error represents: how well we think the selected sample will allow us to predict things about the entire population.***

The way that we compute a margin of error consists of a few factors:

1. The *size* of the sample.
2. Magnitude of *known* problems in the sample.
3. The *extremity* of the statistic.

The larger a sample is, the more likely it is to be representative. The intuition behind that is pretty simple: the more individuals that we include in the sample, the less likely we are to accidentally omit some group or some trend that exists within the population. Using the presidential election as an example, if you polled 100 randomly selected people in Manhattan, you would probably get mostly democrats, and a few republicans. But Manhattan actually has a fairly sizeable group of people who vote for independents. With the smaller sample size, you'd wind up with statistics that overstated the number of democrats in Manhattan, because the independent voters would probably be "hidden" inside of the democratic number. If you sampled 1,000 people, you would be more likely to get a more realistic picture of NYC. You should capture the democrats, republicans, the conservative party, the tea party, etc., all groups you would have missed in the smaller sample.

When we start to look at a statistic, we start with an *expectation*: a very rough sense of what the outcome is likely to be. When we work with a sample, we tend to be *less* confident about how representative that sample is the farther the measured statistic varies from the *expected* value.

Finally, sometimes we know that the mechanism we use for our sample is imperfect - that is, we know that our sample contains an unavoidable bias. In that case, we expand the margin of error to try to represent the reduced certainty caused by the known bias. For example, in elections, we know that in general, there are certain groups of people who simply are less likely to participate in exit polls. An exit poll typically does not generate an unbiased sample because the outcome is partially determined by who is willing to stop and take the poll. Another example is with polls/surveys involving things like sexuality, drugs or religion where social factors may discourage people from admitting to certain things. If you're trying to measure something like "What percentage of people have had extramarital affairs?" you know that many people are not going to tell the truth - so your result will include an expected bias.

Margin of Error

There are several errors in how people generally use the margin of error:

1. The most glaring error is not citing the confidence interval. You cannot know what a margin of error means if the confidence interval isn't specified.
2. Many people, especially journalists, believe that the margin of error includes all possible sources of error. It most emphatically does not. It only specifies the magnitude of error introduced by non-deliberate sampling errors. In a scientific experiment, experimental errors and measurement errors always affect the outcome of the experiment, but the margin of error does not include those factors. It only includes the sampling error. In a poll/survey, the wording of a question and the way in which its asked have a huge impact on the response, and that is not part of the margin of error.

   For example, if you wanted to measure support for concealed weapons carry permit for guns, you could ask people "Do you believe that people should be allowed to carry concealed weapons anywhere they go, including government buildings, schools, and churches?", you'd get one result. If you asked "Do you believe that citizens have a right to carry a weapon to protect themselves and their families from criminals?", you'd get a very different answer - the phrasing of the first question is likely to bias people against saying yes, by deliberately invoking the image of guns in a school or a church. The phrasing of the second question is likely to generate far more "Yes" answers than the first, because it invokes the image of self-protection from rampaging bad-guys.

3. People frequently believe that the margin of error is a measure of the quality of a statistic. That is, that a well-designed poll will have a smaller margin of error than a poorly designed poll. It does not. The margin of error only represents sampling errors. A well designed poll/survey with a sample size of 100 will virtually always have a considerably larger margin of error than a poorly designed poll/survey with a sample size of 1,000. If you want to know the quality of a poll, you need to know more information than just the margin of error. If you want to gauge the relative quality of two different polls, you need to know more than just the margin of error. In either case, you really need to know the sample size, how the sample was collected, and most importantly exactly what the polls/survey are intended to measure. To give another political example, there are a number of different polls taken on a very regular basis of the approval ratings of the president. These polls vary quite drastically. For example, in this week's polls, the number of people who approve of the president range from 30% to 39%, with margins of error in the 3% range.

Source: http://scienceblogs.com/goodmath/2007/01/basics_margin_of_error.php